



# **Oligopolio, istituzioni e performance delle imprese**

# **Oligopoly, Institutions and Firms' Performance**

A cura di Luciano Fanti

PISA  
UNIVERSITY  
PRESS



Oligopolio, istituzioni e performance delle imprese = Oligopoly, institutions and firms' performance / a cura di Luciano Fanti. - Pisa : Pisa university press, 2017. (Saggi e studi)

338.82 (WD)

I. Fanti, Luciano 1. Economia aziendale 2. Oligopoli

CIP a cura del Sistema bibliotecario dell'Università di Pisa

La pubblicazione di questo volume è stata finanziata nell'ambito del Progetto di Ricerca di Ateneo 2016 dell'Università di Pisa dal titolo "Oligopolio, istituzioni e performance delle imprese"

MEMBRO DEL COORDINAMENTO  
UNIVERSITY PRESS ITALIANE

**UPI**  
UNIVERSITY  
PRESS ITALIANE

© Copyright 2017 by Pisa University Press srl  
Società con socio unico Università di Pisa  
Capitale Sociale € 20.000,00 iv. - Partita IVA 02047370503  
Sede legale: Lungarno Pacinotti 43/44 - 56126 Pisa  
Tel. + 39 050 2212056 Fax + 39 050 2212945  
press@unipi.it  
www.pisauniversitypress.it

ISBN 978-88-6741-897-8

progetto grafico: Andrea Rosellini  
impaginazione: Ornella Ambrosio

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da CLEARedi - Centro Licenze e Autorizzazione per le Riproduzioni Editoriali - Corso di Porta Romana, 108 - 20122 Milano - Tel. (+39) 02 89280804 - E-mail: info@clearedi.org - Sito web: www.clearedi.org

# Indice

Introduzione <i>Luciano Fanti</i>	7
Endogenous sequence of contracts in managerial unionised duopolies. Welfare analysis and product differentiation <i>Luciano Fanti, Nicola Meccheri</i>	23
On the effects of firing costs on employment and welfare in a duopoly market with entry <i>Simone D'Alessandro, Nicola Meccheri, Davide M. Coluccia</i>	43
Timing and the selection of the bargaining agenda <i>Luciano Fanti, Domenico Buccella</i>	57
Paolo Sylos-Labini vindicated <i>Simone D'Alessandro, Neri Salvadori, Rodolfo Signorino</i>	85
Regulation and deregulation policies in an economy with imperfect competition and firms entry <i>Lorenzo Corsini</i>	119
Price competition in an international mixed market <i>Alessandra Chirco, Marcella Scrimatore</i>	141

La dimensione temporale del processo produttivo in Nicholas Georgescu-Roegen e Gordon C. Winston <i>Giuseppe Vittucci Marzetti, Mario Morroni</i>	153
Free from what? Competition, regulation and antitrust in American economics 1870-1914 <i>Nicola Giocoli</i>	173
Specific investments and residual control rights: a reassessment <i>Massimiliano Vatiere</i>	223
Un database storico su imprese e settori dell'economia italiana, 1979-2015: nota metodologica <i>Alberto Bianchi, Carlo Brambilla, Giuseppe Conti, Fabio Lavista, Valeria Pinchera</i>	235
Settore bancario e frontiere di efficienza <i>Carlo Brambilla, Raffaello Seri</i>	241
Società e gruppi nei settori tessile-abbigliamento dal 1979 al 2009 <i>Giuseppe Conti, Valeria Pinchera</i>	261
Struttura, comportamento, performance dell'industria farmaceutica italiana dal 1979 al 2008 <i>Alberto Bianchi</i>	289
Analisi di struttura-comportamento-performance: l'industria pesante <i>Fabio Lavista, Carlo Brambilla</i>	309
Longevità e modelli di business dei social network online: un'analisi empirica <i>Andrea Mangani</i>	329

Firm growth, export behavior, and the age-size profile: evidence from the Italian manufacturing sector <i>Marco Grazzi, Daniele Moschella</i>	355
Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica <i>Annamaria Nifo, Sabrina Ruberto, Gaetano Vecchione</i>	381
Profili degli autori	409



# Introduzione

LUCIANO FANTI

Negli ultimi decenni le strutture di mercato e le performance delle imprese hanno presentato evoluzioni drastiche, accompagnate anche da mutamenti dei contesti istituzionali e delle visioni economico-politiche.<sup>1</sup>

Queste evoluzioni sono state accompagnate dallo sviluppo di nuovi studi accademici ma hanno acceso anche ampi dibattiti sulla stampa e fra i “policy-makers”.

La produzione scientifica sull'argomento qui trattato si è concentrata soprattutto sui filoni aperti dalla nuova letteratura di Industrial Organization, all'interno della quale i temi connessi all'oligopolio e all'impresa come istituzione sono stati rivisitati ampiamente, mentre anche l'indagine storica si è avvalsa di nuovi strumenti metodologici e statistici, fra cui la creazione e l'uso di nuovi database. Di questi sviluppi si dà una parziale ma ampia testimonianza nei capitoli di questo volume.

I saggi qui contenuti costituiscono il primo parziale risultato di una ricerca nata dall'interesse degli autori verso aspetti diversi di questo problema; essi non hanno l'obiettivo di un confronto sistematico sulla recente evoluzione delle teorie dell'oligopolio e delle istituzioni né della storia delle imprese e dei loro risultati, ma propongono analisi su singoli temi che sono apparsi talora emblematici e significativi delle più recenti trasformazioni. La ricerca è stata condotta all'interno del Progetto di Ricerca di Ateneo 2016 dell'Università di Pisa dal titolo omonimo “Oligopolio, istituzioni e performance delle imprese”, che

<sup>1</sup> Un breve riferimento, specie per l'Europa, va alla evoluzione del dibattito fra le varie visioni economico-politiche, da quella “welfarista” di ispirazione “socialdemocratica” a quella “neoliberista” fino alla visione “ordoliberalista” nelle sue varie declinazioni e varianti quale l’“economia sociale di mercato”, che probabilmente, nell'opinione dello scrivente, tende a rappresentare meglio i recenti cambiamenti nelle strutture di mercato e nelle istituzioni europee.

ha consentito a studiosi di queste tematiche talvolta affrontate separatamente (teoria economica formalizzata, analisi istituzionale, analisi empirica, indagine storico-economica) un proficuo confronto e interscambio. Inoltre un convegno tenuto presso il Dipartimento di Economia e Management dell'università di Pisa nel gennaio 2017 ha permesso di presentare e discutere differenti aspetti delle strutture di mercato nelle loro implicazioni, teoriche, empiriche, istituzionali, storiche e metodologiche.

Scopo del progetto di ricerca era l'indagine sulle molteplici interrelazioni tra il funzionamento di mercati non perfettamente competitivi dei prodotti e del lavoro, la performance delle imprese e, più in generale, il benessere sociale, sia da una prospettiva teorica che da una applicata e storico-istituzionale.

Sebbene le teorie dell'oligopolio siano state sostanzialmente formulate fin dal XIX secolo, una ripresa intensa del tema avviene alla fine degli anni ottanta, in concomitanza con l'affermarsi della teoria della "Industrial Organization"<sup>2</sup>. Come noto, nel caso dell'oligopolio poiché il comportamento della singola impresa influenza quello delle altre (diversamente dal monopolio o dalla concorrenza monopolistica) non è possibile determinare sulla base della tradizionale curva di domanda alcuna soluzione determinata<sup>3</sup>. Pertanto la diffusione della teoria dei giochi è venuta a fornire uno strumento metodologico per la soluzione del problema della indeterminatezza della teoria dell'oligopolio<sup>4</sup>. In particolare, la moderna teoria rivisita anche gli importanti progressi teorici compiuti negli anni successivi alla seconda guerra mondiale, i cui autori più noti sono Bain, Sylos Labini e Modigliani. Gli argomenti che guidano la rivisitazione sono ben esemplificati da Dixit<sup>5</sup> che in un suo riesame delle teorie

<sup>2</sup> Il manuale "fondativo" della teoria può essere considerato quello di Tirole, J. (1988), *The Theory of Industrial Organization*, Cambridge, MA: MIT Press.

<sup>3</sup> "La critica moderna ha infatti mostrato che tutte le soluzioni indicate dagli economisti precedenti, ed in particolare quelle di Cournot e di Edgeworth, si basano su ipotesi speciali che non hanno alcun titolo per essere considerate come una descrizione esatta ed esauriente della realtà" (Napoleoni, C., 1976, *Il pensiero economico del '900*, Torino, Einaudi, 76).

<sup>4</sup> Tuttavia non c'è unanime consenso sul "realismo" delle soluzioni consentite dalla teoria di giochi, come si evidenzia, per esempio, dall'opinione di Napoleoni (1976, 148): "La teoria dei giochi ha comportato sviluppi matematici imponenti....per quanto riguarda l'economia, peraltro, la rilevanza della teoria sembra molto limitata, giacché i comportamenti che hanno luogo in tutte le situazioni di mercato di tipo oligopolistico sono generalmente lontanissimi dai comportamenti propri dei giocatori...".

<sup>5</sup> Dixit, A. (1982), "Recent developments in Oligopoly Theory", *The American Economic Review*, Vol. 72, N. 2, Papers and Proceedings of the Ninety-Fourth Annual Meeting of the AEA, May, 12-17.



oligopolistiche si concentra sul tema dell'entrata: "My chosen topic, like that of Franco Modigliani nearly twenty-five years ago, is entry... The new work can be characterised by its recognition of the interdependent sequential decisions that the issue of entry involves, and formulation of logically consistent set of assumptions and equilibrium concepts to replace earlier ad hoc methods." (Dixit, 1982, p. 12).

Più in generale, le caratteristiche della moderna teoria dell'oligopolio sono ben descritte nelle parole di Shapiro (1989, p. 332): "The various modern theories of oligopoly behavior are essentially a set of different games that have been analyzed; these games do not represent competing theories, but rather models relevant in different industries or circumstances... These modern theories include capacity investment, cost-reducing R&D investment, advertising, patent licensing, product selection, long-term contracts with customers, and so forth."<sup>6</sup>

Fra gli sviluppi successivi alla fotografia scattata da Shapiro, sono particolarmente degne di menzione le tematiche legate alla diffusa presenza della "delega manageriale" e della "sindacalizzazione" nei moderni oligopoli. Rispetto a quest'ultima, infatti, la presenza di sindacati e l'esistenza di strutture di mercato di tipo oligopolistico sembrano essere fortemente correlati, come suggerito da Booth (1995, p. 95)<sup>7</sup>: "[i]t appears to be an empirical regularity that imperfections in the labor market are correlated with imperfections in the product market". Rispetto alla prima tematica, la moderna visione degli oligopoli ha consentito una ulteriore spiegazione della separazione fra ownership and management, osservata fin dagli anni trenta, in termini della interazione strategica fra imprese oligopolistiche, come chiaramente intuito da Fershtman e Judd (1987): "when we discuss oligopolistic markets, the individual owner-manager relationships must be examined within the context of rivalrous owner-manager pairings. More generally, whenever the profit accruing to a principal-agent pair depends on decisions that other rational agents make, the potential interactions must be considered." (p. 927).<sup>8</sup>

Infatti, nonostante la numerosità delle industrie oligopolistiche con presenza di sindacati, le analisi approfondite sono state (fino alla fine degli anni '80) in larga parte trascurate sia nell'ambito dell'economia del lavoro che in quello della economia industriale, mentre il tema degli "oligopoli sindacalizzati" merita di

<sup>6</sup> Shapiro, C. (1989), "Theories of Oligopoly Behavior", in R. Schmalensee and R. D. Willig (eds), *Handbook of Industrial Organization*, Vol. 1, Amsterdam: North-Holland, 329-414.

<sup>7</sup> Booth, A.L. (1995), *The Economics of the Trade Union*, Cambridge, UK: Cambridge University Press.

<sup>8</sup> Fershtman, C., Judd K.L. (1987), "Equilibrium incentives in oligopoly", *American Economic Review*, 77, 927-940.

essere approfondito, per esempio studiando l'influenza della presenza dei sindacati (e altri fattori istituzionali) sulla competizione strategica tra imprese, che si attua anche attraverso la delega manageriale, la scelta della capacità produttiva e le decisioni di entrata nel mercato e di fusioni fra imprese.

La letteratura sugli "oligopoli sindacalizzati" ha mirato a studiare principalmente (p.e., Horn e Wolinsky, 1988)<sup>9</sup> gli effetti strategici "incrociati" dei sindacati sui mercati del lavoro e del prodotto, con competizione nelle quantità o nei prezzi, e con differenti strutture sia dell'organizzazione sindacale che della contrattazione del salario. L'interazione della "sindacalizzazione" con altri fattori economici e istituzionali, per esempio con la delega manageriale, l'entrata di nuove imprese e la presenza di legislazioni regolamentative dei mercati e dell'occupazione, rappresenta un filone di ricerca estremamente attuale. Contributi a questo filone della letteratura appaiono nei capitoli 1, 2, 3 e 5.

Un altro tema di ricorrente importanza nella letteratura sugli oligopoli concerne la "Bertrand competition", che è riconoscibile sia negli economisti Classici che in Sylos Labini (1957), dove l'entrata strategica è stata generalmente studiata in presenza di un solo "incumbent". Rispetto a questo caso classico di oligopolio due argomenti sono di particolare interesse: (i) il triopolio nel modello di Edgeworth; (ii) l'entrata strategica (vedi anche le parole di Dixit sopra menzionate). Su quest'ultimo argomento, il capitolo 4 si propone di studiare il caso in cui ci sono più "incumbents", proponendo una nuova versione del modello di oligopolio di Sylos Labini.

Alla crescente diffusione di strutture di mercato non competitive, si sono accompagnate fin dai suoi albori sia l'analisi economica delle sue conseguenze (la "perdita di benessere" divulgata dai manuali di microeconomia) che l'analisi di "Law and Economics" con le ricadute sulle politiche legislative in materia.

Data la possibilità di "restringere" la concorrenza attraverso pratiche più o meno "collusive", da tempo, specialmente negli USA, si è sviluppata una letteratura teorica a supporto della legislazione anti-collusione e della relativa applicazione a casi concreti. Pertanto, l'obiettivo di "ricostruire" il contributo delle teorie economiche alla storia recente dell' "American antitrust law", in particolare l'espansione delle teorie giuridiche ed economiche "conservatrici" della Scuola di Chicago, le conseguenze negative che questa espansione ha avuto sulla legislazione antitrust e il fallimento dell'approccio Post-Chicago, è apparso meritevole per la ricerca.

<sup>9</sup> Horn, H., Wolinski A. (1988), "Bilateral Monopolies and Incentives for Merger", *RAND Journal of Economics*, 19, 3, 408-419.

Infatti tale analisi può essere rilevante per la policy, da un lato anche nell'ambito dell'attuale dibattito politico-istituzionale sulla struttura sindacale e/o la regolazione del mercato del lavoro, e dall'altro lato nell'ambito di una più fine comprensione della struttura di mercati che competono sui prezzi come dell'agenda sulla legislazione antitrust. Rispetto a quest'ultima, sebbene l'economia della legislazione antitrust in America abbia una lunga storia e alcuni storici (Hovenkamp, 1991)<sup>10</sup> abbiano esaminato il ruolo delle teorie economiche nel disegno e nell'applicazione di tale legislazione, la loro ricostruzione può non apparire soddisfacente. Ad una rivisitazione di quella ricostruzione contribuisce il capitolo 8.

Infine, la dinamica storica dell'industria europea ha presentato nelle recenti decadi mutazioni di vasta portata e di non immediata interpretazione, di cui forse il comune denominatore può essere considerato proprio il rafforzarsi della struttura oligopolistica dei mercati sebbene in forme mutevoli e talvolta nuove<sup>11</sup>. Volendo assumere, ma solo indicativamente, la data del Trattato di Maastricht come inizio del processo, possiamo osservare la trasformazione dell'industria europea nel senso di rafforzamento oligopolistico, con alcune specifiche caratteristiche: i) una struttura sempre più sovranazionale; ii) una delocalizzazione dell'apparato produttivo verso l'est-europeo e fuori dell'Europa; iii) una politica di forte contenimento salariale; iv) la formazione di nuove strutture di mercato, che possono indicativamente distinguersi in a) un agglomerarsi di sistemi di imprese attorno a un'azienda leader che controlla la parte finale del processo produttivo con livelli di sub-fornitura via via meno specializzati e di minor valore aggiunto (catena a livelli), oppure b) la formazione di reti di imprese produttive e di servizi, entrambi ad alta specializzazione, che lavorano per molte imprese leader (struttura a rete); v) il nucleo centrale della produzione manifatturiera europea nell'area tedesca (caratterizzata dalla linea neo-mercantilista e dai conseguenti surplus commerciali), seguita a di-

<sup>10</sup> Hovenkamp, H. (1991), *Enterprise and American Law, 1836-1937*, Cambridge, MA: Harvard University Press, x, 443.

<sup>11</sup> Naturalmente le strutture oligopolistiche possono presentare alcune caratteristiche differenzianti (per esempio in termini delle modalità di "governance", di finanziamento e di relazioni industriali) anche sulla base della loro evoluzione storica, territoriale e organizzativa che meriterebbero ulteriori approfondimenti analitici. Per semplificare, si sottolinea qui la differenza fra il capitalismo germanico-giapponese, in cui predomina l'efficienza produttiva al posto di quella allocativa tipica invece del capitalismo anglosassone; nel primo appare un capitalismo di tipo manageriale orientato alla produzione e non dominato strettamente dagli interessi degli azionisti (Berselli, E., 2010, *Leconomia giusta*, Torino, Einaudi, 60).

stanza da quella italiana, con una creazione di catene di sub-fornitura nei paesi est-europei; vi) modi di integrazione di queste strutture di imprese, che fanno intravedere nuove forme di governance oligopolistica rispetto a molteplici elementi non solo produttivi ma anche finanziari e più genericamente socio-politici.<sup>12</sup> Non è forse un caso che, all'interno di queste tendenze, l'Italia, come paese industriale spesso caratterizzato dal "piccolo è bello" e dai "distretti", si trovi in una situazione particolarmente critica in cui si sta richiedendo una trasformazione verso dimensioni più larghe attraverso l'integrazione delle filiere produttive non tanto sul piano geografico, ma anche finanziario ed economico. Infatti, gli ultimi 30 anni di storia dell'industria in Italia sono stati caratterizzati sia da modificazioni strutturali e cicliche che da nuovi e mutevoli paradigmi interpretativi. Per comprendere ciò che è accaduto e che sta attualmente accadendo all'industria italiana, un importante obiettivo di ricerca è stato individuato nel completamento di un database per gli studi storico-empirici sull'industria (tratto dalle pubblicazioni annuali di Mediobanca dal 1979 a oggi)<sup>13</sup> della cui metodologia di costruzione si dà conto nel capitolo 10 e di una stesura di una storia dell'industria italiana, delle sue componenti principali e delle sue dinamiche di performance. Quindi con specifico riferimento al caso italiano e alla sua recente evoluzione, appaiono qui specifici contributi sull'industria tessile-abbigliamento (capitolo 12), sull'industria farmaceutica (capitolo 13), sull'industria pesante (capitolo 14), su quella dei servizi finanziari (capitoli 11 e 17) e sull'industrie orientate all'export (capitolo 16).

Più in generale, la naturale domanda che motiva la ricerca (sebbene essa sia necessariamente specifica nei temi e specializzata tecnicamente) riguarda come si è modificato il ruolo degli oligopoli all'interno dell'economia e della società nell'esperienza degli ultimi decenni. Un suggerimento sulla tendenza al rafforzamento della struttura oligopolistica dell'economia può essere ricavato

<sup>12</sup> Su queste tendenze dell'industria europea, peraltro inserite nel più generale contesto economico, sociale e politico dell'Europa, si veda, per esempio, le riflessioni critiche in Simonazzi A., Ginzburg A. and Nocella G. (2013), "Economic relations between Germany and southern Europe", *Cambridge Journal of Economics*, 37, 653-675, e Bellofiore R., Garibaldo F., e Mortagua M., A credit-money and structural perspective on the European crisis: why exiting the euro is the answer to the wrong question, (paper presented at the international workshop on Full Employment in Europe: With or Without the Euro? In Grenoble, May 15 -16, 2014).

<sup>13</sup> Per la storia dell'industria in Italia molti studi si sono già avvalsi dei dati Mediobanca ma non esistevano ancora studi sistematici, né settoriali di tali dati per un arco temporale sufficientemente lungo da poter poi consentire un'analisi approfondita della storia dell'industria italiana nei recenti decenni.

dai seguenti dati (che non hanno solo una valenza aneddotica): con riferimento agli Usa per il recente ventennio a partire dal 1996, si è assistito da una parte ad una riduzione delle imprese quotate in Borsa (3556 nel 2016 contro 7322 nel 1996), ma dall'altra parte ad un aumento della loro valorizzazione (approssimativamente il 135% del Pil nel 2016 contro il 105% nel 1996). Un minore numero, una più estesa durata e una maggiore dimensione (in valore) delle imprese rappresentano quindi un sintomo evidente del rafforzamento della struttura oligopolistica dell'economia. Di pari passo, dovrebbe emergere, da un punto di vista delle istituzioni, la tesi della necessità di una maggiore attenzione agli accresciuti poteri di mercato degli oligopoli,<sup>14</sup> per esempio da parte dell' "antitrust law"<sup>15</sup>.

Se il fenomeno del rafforzamento della struttura oligopolistica dei mercati è probabilmente indubbio, differenti sono le angolazioni dalle quali l'analisi economica osserva tale fenomeno. Da un lato, vi è una visione diffusa nella "mainstream economics" secondo la quale i) una struttura oligopolistica del mercato può anche produrre risultati competitivi, per cui ogni regolazione può essere ridondante e spesso controproducente, dato che in un oligopolio è difficile stabilire se le imprese stanno sfruttando il loro potere di mercato o stanno competendo in modo "concorrenziale" (cioè è difficile distinguere tra movimenti di prezzi coordinati e movimenti di prezzi concorrenziali in un mercato oligopolistico)<sup>16</sup>, e ii) persino i nuovi settori della elettronica e delle telecomunicazioni, in cui emergono imprese "giganti" i cui fatturati superano persino quelli di molti Stati, sono proba-

<sup>14</sup> Si può anche far notare come, ad un livello macroeconomico, questo rafforzamento della struttura oligopolistica dei mercati sia stato accompagnato da un aumento della disuguaglianza dei redditi.

<sup>15</sup> Per esempio, una evidente relazione causale diretta fra istituzioni legislative e strutture oligopolistiche emerge dalla ondata di fusioni e acquisizioni ("merger wave") negli Usa negli anni '60, caratterizzata dalla tipologia di "conglomerate merger": "*we note... a rise in "pure" conglomerate mergers from the 1960s... The wave of the 1960s was preponderantly conglomerate, reflecting a much more stringent antitrust policy against horizontal mergers.*" (Shy, O., 1995, *Industrial Organization: Theory and applications*, The MIT Press, Cambridge, Ma, 174).

<sup>16</sup> La posizione che suggerisce cautela nel regolare i mercati oligopolistici in Europa è ben esemplificata dalle parole di Bishop e Walker (2010), i quali, nel loro esame della relazione fra teoria economica e sua applicazione alla "European competition law" attraverso l'analisi delle decisioni delle Commissioni e dei giudizi delle Corti, affermano: "...most market outcomes that are consistent with firms engaging in tacit co-ordination are also consistent with those firms competing aggressively. For example, the parallel movement of prices over time is entirely consistent with firms competing aggressively against one another. One should therefore be extremely cautious in moving from a theoretical possibility to a policy prescription in which normal market conduct becomes subject to regulatory intervention." (Bishop, S., Walker M., 2010, *The Economics of EC Competition Laws*, Sweet and Maxwell, 3rd. Ed., 340)..

bilmente caratterizzati dalla concorrenza à la Bertrand, per cui anche tali strutture altamente oligopolistiche possono risultare socialmente ottimali. Inoltre, in tale visione viene altresì obiettato che la regolamentazione, sebbene possa giovare al consumatore nel breve periodo (efficienza statica), può essere dannosa nel lungo periodo in termini di minori investimenti e minore innovazione e quindi riduzione dell'efficienza dinamica complessiva (come suggerito per esempio dall'analisi della interazione fra concorrenza e innovazione da Aghion et al., 2005)<sup>17</sup>.

Da un altro lato, secondo talune opinioni più critiche, l'importanza assunta dagli oligopoli nelle recenti decadi è un fatto stilizzato dalle conseguenze di lungo periodo che necessita di essere compreso a fondo anche per i suoi possibili esiti negativi, per esempio nello spirito dell'analisi di Baran e Sweezy<sup>18</sup> che si focalizzava sull'analogo fenomeno di diffusione della monopolizzazione a partire dal secolo precedente: si sta infatti diffondendo una economia caratterizzata da oligopoli generalizzati (nel senso di una loro diffusione anche in settori lontani dal "core business" dove detengono esplicitamente il potere di mercato, basti pensare al fenomeno delle "conglomerate"), globalizzati (nel senso dato dalla tendenza - politicamente organizzata anche da organismi sopranazionali - alla liberalizzazione dei mercati a livello mondiale)<sup>19</sup> e finanziarizzati (nel senso di una partecipazione delle grandi imprese oligopoliste nei settori produttivi ai grandi oligopoli finanziari).<sup>20</sup>

<sup>17</sup> Aghion P., Bloom N., Blundell R., Griffith R., Howitt P. (2005), "Competition and Innovation: an Inverted-U Relationship", *Quarterly Journal of Economics*, 120 (2): 701-728.

<sup>18</sup> Baran, P.A., Sweezy P. (1966), *Monopoly Capital*. New York and London: Monthly Review Press, Ix+402.

<sup>19</sup> Una relazione causale in direzione inversa fra istituzioni e strutture oligopolistiche emerge recentemente dalla pressione che i mercati oligopolistici (e altre lobbies) esercitano tramite la globalizzazione verso una convergenza (e nel contempo un indebolimento - superamento) delle istituzioni nazionali, sebbene non vi sia consenso su tale effetto ( per esempio Hall A. P., and Soskice D., 2003, *Varietà dei capitalismi: alcuni tratti fondamentali*, *Stato e Mercato*, 3, 341, concludono che "national institutional frameworks need not converge in the face of pressures from globalization because firms can exploit and will often offer political support for the distinctive institutional advantages present in their political economy").

<sup>20</sup> Un esempio, peraltro non molto studiato teoricamente, della "finanziarizzazione" delle imprese industriali oligopolistiche ha avuto inizialmente origine dalla formazione di divisioni finanziarie interne, tradizionalmente votate alla concessione di prestiti per favorire l'acquisto dei propri prodotti come l'auto; a partire da queste "tali divisioni sono passate ad ogni sorta di operazione bancaria....da cui deriva un reddito paragonabile a quello della produzione industriale, per di più ottenuto con una quantità di personale assai inferiore....Fu per questo che venne allora coniata da alcune autrici la definizione delle case automobilistiche quali 'banche che occasionalmente fabbricano automobili' " (Gallino, L., 2015, *Il denaro, il debito e la doppia crisi*, Einaudi, Torino, p. 40).

Parallelamente all'accrescimento del ruolo e del potere degli oligopoli e delle conseguenti ricadute teoriche ed operative sulle istituzioni economiche e sociali, si sono venuti affermando nelle configurazioni e nelle dinamiche dei mercati nuovi fenomeni quali l'apparizione e la rapida espansione delle "network industries" e la diffusa privatizzazione delle imprese pubbliche. Riguardo alle prime (Shy, 2001)<sup>21</sup>, possiamo osservare che esse, per la loro stessa natura, sono industrie imperfettamente competitive e pongono quindi interessanti problemi sia alla regolazione pubblica sia alla teoria del consumatore per i relativi aspetti di forte interazione sociale nelle scelte di consumo. Riguardo al fenomeno della privatizzazione delle industrie, il problema probabilmente più interessante che si pone alla ricerca è perfettamente evidenziato da Stiglitz (2006, p. 81):<sup>22</sup> *"Negli anni ottanta c'è stato un forte movimento di privatizzazione delle imprese pubbliche, anche in settori in cui c'era un monopolio naturale...Ciò ha sollevato naturalmente la domanda: in quali casi è possibile essere sicuri che la privatizzazione giovi al benessere economico?"*. Rispetto ad entrambi i fenomeni sopra citati, il libro fornisce un contributo rispettivamente nei capitoli 15 e nei capitoli 6 e 14.

Come è noto, sin dai contributi di Adam Smith, le istituzioni economiche e sociali hanno esercitato una importante influenza sulla struttura dei mercati e sui pattern storici delle performances delle imprese.

Pertanto anche l'impresa vista come istituzione non può non subire i mutamenti connessi con i fenomeni economici più recenti, modificandosi rispetto alla tipica grande "corporation" americana che ha ispirato, per esempio, i lavori fondativi di Coase nel 1937 e di Chandler nel 1962. Ponendo per esempio al centro del problema economico della produzione il coordinamento, sia interno che esterno all'impresa, di plurime fasi e ponendo altresì esplicitamente l'attenzione sul perché si rende necessaria una attività organizzativa laddove il mercato invece avrebbe dovuto assicurare un coordinamento perfetto, possono emergere specifiche problematiche come il diverso profilo temporale di utilizzo degli input o come, in presenza di specificità delle risorse, il mercato venga sostituito da contratti di lungo periodo che però possono anche essere incompleti. Su queste ultime problematiche il libro presenta contributi nei capitoli 7 e 9.

<sup>21</sup> Shy, O. (2001), *The Economics of Network Industries*, Cambridge, UK: Cambridge University Press. Le industrie "network" includono, tanto per citare solo alcuni prodotti, "telephony, email, Internet, computer hardware and software, music and video, banking and airline services", e sono caratterizzate, per esempio, a differenza delle industrie tradizionali, da complementarità, compatibilità, standards, esternalità nel consumo, "switching costs and lock-in", ed elevate economie di scala produttive.

<sup>22</sup> Stiglitz, J. (2006), *Economia e informazione*, Danews Editrice, Roma.

Entrando più in dettaglio, descriviamo gli argomenti specifici trattati, gli obiettivi di ricerca e i risultati raggiunti nei vari capitoli contenuti nel volume.

Fanti e Meccheri, osservando che nei moderni oligopoli: 1) gestione e proprietà sono separati, per cui i proprietari stipulano contratti manageriali (basati, ad esempio, su un mix di profitti e vendite) al fine di ottenere un vantaggio strategico, 2) i lavoratori sono sindacalizzati, e 3) i contratti manageriali e salariali possono avere differente durata, indagano la sequenza di mosse (relativa ai contratti manageriali e salariali) nel gioco di delegazione strategica tra i proprietari. Essi dimostrano che quando i sindacati sono sufficientemente (ma non estremamente) orientati all'occupazione (e i pagamenti laterali tra i proprietari non sono fattibili), la sequenza endogena di contratti massimizza sia il surplus del consumatore che il benessere sociale (sebbene i profitti non siano massimizzati). Inoltre, l'introduzione della differenziazione del prodotto nell'analisi conduce ad un insieme più ricco di equilibri, confermando in alcuni casi la preferenza contro-intuitiva dei proprietari delle imprese per un ruolo di "follower" rispetto ai propri sindacati nel determinare la sequenza dei contratti manageriali e salariali.

D'Alessandro, Meccheri e Coluccia analizzano il ruolo dei "firing costs" – assunti come esempio della presenza di una legislazione di protezione dell'occupazione, che finora la letteratura sia teorica che empirica aveva trovato non sempre efficace – su occupazione e benessere in un gioco a due stadi, con un "incumbent" e un "entrant", mostrando che l'occupazione e il benessere aumentano effettivamente con i "firing costs", a condizione che questi si trovino entro un certo intervallo. Tali risultati possono essere spiegati dal fatto che in un ambiente dinamico l'incumbent sfrutta la presenza dei "firing costs" per "vincolarsi" credibilmente a un livello elevato di produzione nella seconda fase. Quindi la conclusione è che produzione, occupazione e il benessere complessivo sono massimizzati per un livello strettamente positivo dei "firing costs".

Fanti e Buccella riesaminano il tema dell'agenda di negoziazione tra imprese e sindacati in un duopolo di Cournot, studiando l'agenda ("Right to Manage" o "Efficient Bargaining") che può emergere in equilibrio sotto tre differenti strutture del 'timing' - implicite nel caso del gioco di contrattazione di Efficient Bargaining - e mostrano che, applicando un corretto "game-theoretic approach" all'argomento della agenda di contrattazione, potrebbe rivelarsi la mancanza di robustezza di alcuni risultati tradizionali basati sul confronto di agende esogenamente date. Un nuovo risultato è che la Efficient Bargaining 'sequenziale' emerge come agenda in equilibrio per una gamma del potere sindacale più ampia di quella 'simultanea', ma è Pareto inferiore, mentre l'agenda RTM emerge per un potere sindacale sufficientemente grande ed è Pareto- superiore.



Nel contributo di D'Alessandro, Salvadori e Signorino, l'interpretazione di Modigliani nel 1958 di "Oligopolio e Progresso Tecnico" di Sylos Labini (1957) - e il successivo dibattito sul "Sylos Postulate" e sull'accumulo di capacità produttiva da parte degli "incumbents" come deterrente strategico delle entrate - vengono accuratamente discussi nella prima parte del lavoro. Nella seconda parte gli autori sviluppano un modello in cui postulano tre fasi: nella prima gli "incumbents" scelgono la capacità produttiva; nella seconda un potenziale concorrente sceglie se entrare o no e, in caso di ingresso, sceglie il livello della sua capacità produttiva; nella fase finale gli "incumbents" e il concorrente (eventualmente) competono sui prezzi (à la Bertrand), esplorando in particolare il ruolo delle strategie miste in equilibrio. Con tale modello gli autori raggiungono conclusioni utili "*to 'vindicate' Sylos Labini's theory of oligopoly and set it free from the Procrustean bed of Cournot competition.*"

Il lavoro di Corsini analizza un modello dinamico di entrata in concorrenza imperfetta con l'obiettivo di fornire implicazioni politiche in termini di regolamentazione e deregolamentazione dei mercati, nella misura in cui queste ultime influenzano le imperfezioni di mercato, quindi la profittabilità e conseguentemente l'entrata di imprese. Viene determinato "endogenamente" il numero delle imprese presenti sul mercato nel lungo periodo e quindi vengono esplorati gli effetti della concorrenza imperfetta in tale periodo, mostrando, in particolare, che nel lungo termine il potere contrattuale dei sindacati non ha alcun effetto sul livello dei profitti. Infine, viene mostrato che non esiste alcuna complementarità nelle politiche di deregolamentazione di questi mercati in un orizzonte temporale, mentre nascono alcune complementarità intertemporali.

Chirco e Scrimatore, rilevando la crescente importanza della "transizione" economica, della concorrenza internazionale e della globalizzazione nel contesto economico attuale, sviluppano un modello di oligopolio misto con un'impresa statale e varie imprese estere private che competono nei prezzi e offrono beni differenziati. Assumendo sia movimenti simultanei che movimenti sequenziali con leadership pubblica, gli autori - concentrandosi principalmente sul grado di sostituibilità dei prodotti e la struttura del mercato - valutano gli effetti benefici della privatizzazione e delle acquisizioni transfrontaliere, indipendentemente da qualsiasi differenza di costo esogena o endogena tra imprese pubbliche e private, e rilevano come la presenza di concorrenti internazionali sul mercato interno induce una convergenza complessiva al prezzo efficiente.

Vittucci e Morroni evidenziano nel loro contributo come gli aspetti temporali dei processi produttivi - nonostante l'attività di produzione sia intrinsecamente "time-consuming" - siano trascurati nella maggior parte delle analisi

economiche. Tuttavia gli autori sottolineano che vi sono due eccezioni di rilievo: infatti due studiosi come Georgescu-Roegen e Winston hanno proposto due diverse rappresentazioni analitiche del processo produttivo che consentono di analizzarlo come un processo che si svolge nel tempo. Conseguentemente, il contributo analizza gli aspetti salienti e le implicazioni principali delle opere di questi due studiosi, mettendone in risalto gli aspetti originali, i risultati principali e i caratteri di complementarità, specialmente nella sezione conclusiva che presenta la teoria della produzione di Winston (1982) come il naturale complemento del modello fondi-flussi di Georgescu-Roegen.

Il saggio di Giocoli analizza la giurisprudenza costituzionale della “Gilded Age” e della “Progressive Era” (il periodo della storia americana che va dagli anni ‘70 dell’800 alla fine della prima guerra mondiale), sostenendo la tesi che, indipendentemente dalla traiettoria giurisprudenziale che ha portato storiche sentenze (p.e. la Corte *Lochner*) ad abbracciare la libertà di contratto, la nozione classica della libertà di contratto - insieme al suo classico alter ego, la libertà di commercio - sia un filo rosso che collega gli aspetti legale, politico ed economico di quei periodi rispetto alla libertà dei mercati, ai limiti della concorrenza e all’auspicabilità della regolamentazione. La principale conclusione è che - dato che nei settori della regolamentazione e dell’anti-trust ancora viviamo sotto molti aspetti all’ombra della legislazione sviluppata all’epoca - la dicotomia del XIX secolo “libertà contrattuale contro libertà di commercio” possa influenzare tuttora la tematica legale-politico-economica.

Vatiero affronta nel suo saggio il tema dei “contratti incompleti” sottolineando come base di partenza le distinzioni esistenti fra l’approccio di Coase - concentrato sulla relazione bilaterale e, invece, piuttosto disinteressato alla presenza di “outside options” - e quello ‘à la Hart’, in cui il problema noto come “hold-up” emerge proprio dalla disponibilità di “outside options”, e come, all’interno del secondo punto di vista, si possa ulteriormente distinguere fra chi ritiene, come Hart medesimo, che l’assegnazione dei diritti di controllo residuo all’investitore favorisca gli investimenti e chi invece ritiene, come Williamson, che investimenti specifici, riducendo la disponibilità di “outside options”, tendano a mitigare l’ “hold-up problem”. L’autore sostiene che, contrariamente a quanto talvolta affermato, i due punti di vista sopradetti siano intrinsecamente differenti, e che se gli investimenti specifici possono ridurre, secondo il meccanismo proposto da Williamson, il numero delle alternative disponibili, allora l’ “hold-up problem” può avere soluzione anche senza l’assegnazione dei diritti di controllo residuo.

Bianchi, Brambilla, Conti, La Vista e Pinchera presentano con una nota metodologica un database storico su imprese e settori dell’economia italiana

nel periodo 1979-2015, sul quale i successivi contributi degli stessi autori sono basati.

Il saggio di Conti e Pinchera analizza la performance dei settori “leggeri” italiani nel trentennio 1979-2009, al fine di verificare, almeno per tali settori, alcune ipotesi storiografiche alternative sul declino dell’industria italiana o sulla sua vitalità. La prima tesi imputa il declino a settori “labour intensive” e a tecnologie “mature” come quelli del tessile e abbigliamento, mentre la tesi contrapposta ritiene invece che una parte delle manifatturiere “leggere” abbia contribuito alla crescita e all’incremento delle quote di export. Il lavoro innova i metodi solitamente utilizzati per sostenere l’una o l’altra tesi, basandosi su un’analisi dei dati, in particolare i bilanci di 11 società grandi e medio-grandi, messi a confronto con gli andamenti di settore. Dall’analisi emergono situazioni molto variegata, caratterizzate da ampie divergenze tendenziali e dinamiche concorrenziali che hanno fortemente risentito di fenomeni di riorganizzazioni di gruppo. Gruppi e “conglomerate” hanno permesso una riallocazione e diversificazione delle risorse produttive, e di quelle finanziarie infra e extra-gruppo, realizzando una maggiore flessibilità organizzativa con risultati nel complesso eterogenei.

Bianchi studia il settore farmaceutico italiano, rilevando negli anni Novanta una discontinuità dovuta alla diffusione di strategie di *management* delle grandi multinazionali estere, centrate sui processi di concentrazione industriale e, inoltre, evidenziando a partire dal nuovo millennio, un mutamento nelle condizioni di redditività delle imprese farmaceutiche influenzate da una regolazione pubblica sempre più stringente. Infatti dagli elevati livelli di margini economici registrati negli anni Ottanta, il settore ha iniziato a sperimentare, soprattutto dagli inizi del nuovo millennio, un graduale peggioramento della propria *performance* economica a causa di una crescente regolazione, in termini di controllo dei prezzi delle specialità medicinali e di contenimento della spesa farmaceutico-sanitaria. L’insufficiente dinamica dei prezzi amministrati ha fatto così venire meno nel nostro paese la pressione competitiva proveniente dal mercato dei prodotti a brevetto scaduto che, invece, ha conosciuto una forte espansione in altri paesi industrializzati.

Brambilla e Seri studiano le trasformazioni avvenute a partire dagli anni Ottanta nell’architettura del sistema bancario italiano che avevano l’obiettivo di preparare, in termini di maggiore concorrenza e maggiore efficienza, tale sistema al processo di integrazione europea e alle nuove sfide dei mercati globalizzati. Utilizzando i bilanci per un ampio numero di banche che coprono quasi tutto il sistema, tra il 1979 e il 2000, e usando la stima non parametrica del “technology set” attraverso il DEA (Analisi dei dati), gli autori determinano

gli effetti di nuovi regolamenti e quadri istituzionali sull'efficienza del sistema bancario e delle singole banche, evidenziando in particolare, da un lato, una forte polarizzazione dimensionale, con una crescita della dimensione media degli intermediari maggiori e una quasi sparizione delle banche di dimensione media e medio piccola, e, dall'altro lato, la performance molto negativa delle banche medie, sia perché sempre posizionate molto lontane dalla frontiera produttiva, sia perché nel complesso i livelli di efficienza tendono a ridursi nel corso del periodo.

Brambilla e LaVista presentano i primi risultati di uno studio sull'industria metallurgica italiana, utilizzando un nuovo data set dei bilanci delle grandi aziende tra 1970 e 2000, con cui analizzano i cambiamenti nella struttura del mercato, nel comportamento e nelle prestazioni di numerose imprese, in un periodo in cui l'industria ha subito un profondo processo di ristrutturazione in tutto il mondo. In particolare, in Italia è stato attuato un processo di privatizzazione da parte del governo italiano, avviato in concomitanza con la crisi monetaria del 1992. L'analisi evidenzia le ambiguità del processo di privatizzazione (caratterizzato per esempio dalla scarsa solidità finanziaria degli acquirenti, che si è tradotta in mancati nuovi investimenti), in termini di miglioramento della redditività delle imprese e di vantaggi per il consumatore finale, in linea con le evidenze messe in luce nell'analisi retrospettiva recentemente eseguita dalla Corte dei Conti (2010).

Il saggio di Mangani è focalizzato sul settore dei "social networking sites" (SNS), con l'obiettivo di individuare la relazione fra le caratteristiche peculiari dei SNS e la loro probabilità di sopravvivenza attraverso la costruzione di un appropriato database dei social network apparsi on-line a partire dal 1995. L'analisi empirica rivela che quando i SNS sono i) rivolti a un pubblico specializzato e che rimangono indipendenti, ii) più "giovani" (rispetto a quelli più "vecchi" e con maggiore esperienza), hanno una maggiore probabilità di sopravvivenza nel mercato. Allora il risultato ii) può essere coerente con una visione schumpeteriana della concorrenza in cui le nuove imprese innovano per convincere i consumatori a sostenere 'switching costs' per abbandonare il vecchio network ed entrare in uno nuovo, mentre il risultato i) potrebbe alleviare, nella misura in cui favorisce la frammentazione del mercato e riduce il "potere" delle piattaforme social generaliste e dominanti (come Facebook), le preoccupazioni di natura antitrust riguardanti il settore dei SNS.

Il lavoro di Grazzi e Moschella considera i modelli di crescita relativi delle imprese in dipendenza della interazione fra età, dimensione e status di esportazione, individuando aziende "nuove" che dall'inizio della loro attività sono impegnate nel commercio internazionale, caratterizzate da una maggiore effi-

cienza e dalla disponibilità ad assumere opportunità di crescita. Viene quindi mostrato che queste piccole “giovani” imprese esportatrici mostrano un atteggiamento più pronunciato verso la crescita sia sul mercato nazionale che sul mercato internazionale. Inoltre, nonostante la loro esperienza molto limitata, tali “giovani” imprese sembrano ben equipaggiate per affrontare shocks come un apprezzamento del tasso di cambio reale. Le evidenziate performance di sviluppo delle “giovani” imprese esportatrici supportano l'ipotesi che l'atteggiamento verso le opportunità di crescita sia una determinante, talvolta trascurata, della crescita delle imprese.

Nifo, Ruberto e Vecchione nel loro lavoro partono dalla considerazione che recentemente la letteratura economica ha, da un lato, dato grande attenzione all'ampio ricorso al multi-affidamento bancario, motivato da fattori sia microeconomici che macroeconomici, mentre dall'altro lato ha rilevato l'importanza del ruolo delle istituzioni locali e del contesto geografico sulle scelte delle piccole imprese. Rispetto a questo ultimo punto, il saggio si pone l'obiettivo di stimare il peso della qualità delle istituzioni provinciali nel ricorso al multi-affidamento bancario delle piccole e medie imprese manifatturiere italiane. Viene mostrato, conformemente con altri studi, che le istituzioni provinciali sono una determinante fondamentale dei differenziali osservati nel numero di 'banking relationships' per impresa. In particolare le imprese del Mezzogiorno - più “povere istituzionalmente” - registrano sempre un numero di relazioni bancarie relativamente più alto rispetto a quelle del Centro-Nord.

In conclusione, vorrei ricordare che sebbene i capitoli di questo volume siano soltanto piccoli contributi “tecnici” e accademici a temi di più ampio respiro che, a partire dalle strutture del mercato, dalle istituzioni e dai risultati economici delle imprese, coinvolgono l'evoluzione - guidata da forze potenti, spesso permeate da "spiriti animali" e "apprendisti stregoni" - delle società nel loro complesso, una massima keynesiana continua ad motivare tuttora la fatica della ricerca con il valore morale del suo messaggio: *“I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas”*.<sup>23</sup>

<sup>23</sup> Keynes, J.M. (1964), *The General Theory of Employment, Interest and Money*, New York: Harcourt Brace Jovanovich, Publishers, p. 383.



# **Endogenous sequence of contracts in managerial unionised duopolies. Welfare analysis and product differentiation\***

LUCIANO FANTI AND NICOLA MECCHERI

## **1. Introduction**

This paper stems from three stylised facts: 1) the separation of ownership from management (i.e., control) in modern corporations (e.g., Berle and Means, 1932); 2) the widespread presence of unionisation in modern oligopolies (Booth, 1995); and 3) the different length, in general, of managerial and workers' wage contracts.<sup>1</sup> With regard to the first point, a recent branch of

\* We would like to thank seminar participants at the Workshop "Oligopoly, Institutions and Firm Performance" (Pisa, 26-27 January 2017) financed by the University of Pisa (Project: PRA 2016\_2) and at the XV Workshop SIEPI (Palermo, 9-10 February 2017) for useful comments and suggestions. The usual disclaimers apply.

<sup>1</sup> For instance, Liu and Xuan (2016) in a sample of CEO contracts (characterised by a minimal length be at least two years and fixed-term with an exact ending date) covering 130 firms and 138 CEOs included in the Standard & Poor's (S&P) 500 index from 2001 to 2010 and publicly available through corporate filings with the U.S. Securities and Exchange Commission (SEC) an average contract length of 3.2 years. Xu (2013) evidences that, for 3,954 employment fixed-term contracts for 2,964 CEOs by 2,901 firms entered between 1992 and 2008 collected from SEC filing exhibits and, when possible, from The Corporate Library, contract length decreases steadily over time, from 3.77 in 1995 to 2.78 in 2008. This trend seems to hold not only for the U.S. but also for the United Kingdom,

the Industrial Organization (IO) literature has argued that not only managers may deviate from the target of the pure profit maximisation (for example, pursuing sales maximisation, as already discussed by Baumol, 1958), but also that owners, in oligopolistic contexts, may manipulate the decisions of their own managers by incentive contracts (based, for instance, on a mix of profits and sales) in order to attain a strategic advantage. Main contributions in this direction, denoting the so-called managerial delegation literature,<sup>2</sup> are early originated by Vickers (1985), Fershtman (1985), Fershtman and Judd (1987) and Sklivas (1987).<sup>3</sup>

As regards as the second stylised fact, a strand of the IO literature concentrated on unionized oligopolies started from, among others, Horn and Wolinsky (1988), Davidson (1988) and Dowrick (1989), which have emphasised the role of labour unions in determining workers' wages and affecting final outcomes in oligopolistic industries. Although these issues can exist and be

where CEO contract periods have decreased even more dramatically: while the average contract in 1992/93 lasted three years (Conyon, 1994), most UK companies have moved to one year contracts by 2011 (BIS, 2011). On the other hand, also the length of wage contracts has been highly variable in the last decades: for instance, in Sweden (European Commission, 2006) length of new wage agreements increased from one- and two-year contracts in the 1950s to two-year contracts in 1960 to 1966, culminating in the three-year contracts of 1966-1969 and 1971-1973. Two-year agreements were concluded for 1975-1976, 1978-1979, 1981-1982 and 1984-1985 while the agreements shortly after the two oil shocks, 1974 and 1980 were of one year's duration, and finally a two-year contract in 1993 and then three-year contract in 1995 and thereafter.

<sup>2</sup> Broadly speaking, managerial delegation models may be considered as a subcategory of principal-agent models: however, while the latter focused, generally, on inducing the agent to exert a high level of effort by designing the agent's compensation package (e.g., Asseburg and Hofmann, 2010), the former consider that the agent's pay also affects, for instance in an oligopolistic context, competing firms. As a consequence, a strategic element to contract design emerges, and principals may manipulate decisions taken in rival firms through agent's incentive contracts.

<sup>3</sup> Different forms of the incentive schemes have been analysed in the literature. For example, in Vickers (1985) managers are given a contract to maximize profits, but with an additional incentive for sales, while Fershtman and Judd (1987) and Sklivas (1987) use revenue as an additional incentive scheme. Jansen et al. (2007) and Ritz (2008) consider a weighted sum of profits and market share as managerial incentive scheme. Salas-Fumas (1992) and Miller and Pazgal (2002) analyse the relative performance incentive scheme, which is a weighted sum of the firm's own profit and the rival's profit. Finally, Overvest and Veldman (2008) introduce the managerial incentive for process innovation, which is a weighted sum of profits and cost-reducing R&D investment.



analysed independently, a recent and increasing literature, originated by Szymanski (1994) and Bughin (1995), argued that they are coexisting features of real imperfectly competitive product markets, hence it investigates how the interaction between labour unions and the choice of managerial delegation contracts by firms affects product market outcomes. Following Szymanski (1994) and Bughin (1995), other managerial-unionized oligopoly models have been developed by Liao (2010, 2014), Chatterjee and Saha (2017), Fanti and Meccheri (2013, 2015) and Meccheri and Fanti (2014, 2017a). A broader review of this literature reveals that various issues, such as strategic delegation<sup>4</sup> in wage bargaining and optimal managerial delegation contracts under alternative unionization regimes, have been deeply analysed. In particular, Szymanski (1994) (resp. Bughin, 1995) considers a two-stage game in which at the first stage, wages are bargained between managers (resp. owners) and unions and, at the second stage, output decisions are made by managers (resp. cooperatively by a profit-maximising owner and a sales-maximising manager). Subsequently, Chatterjee and Saha (2011) introduce in the above-mentioned context bilateral delegation, where workers offer incentives to their union leader and shareholders to their manager, and bargaining takes place between the union leader(s) and the manager(s). Liao (2010) examines an owner's choice of whether to hire or not a manager to bargain wages and employment with a union in an 'efficient bargaining' context and Liao (2014), where the two firms are assumed to be managerial firms and only wages are bargained in a 'right-to-manage' context, investigates an owner's choice between single-task (i.e., output determination) delegation and multiple-task (i.e., output determination and wage bargaining) delegation. Fanti and Meccheri (2013) and Mec-

<sup>4</sup> Note that another branch of the literature focuses only on strategic union delegation (instead of strategic managerial delegation), that is whether and how union members strategically delegate authority to a negotiator: for example, Jones (1989) early argued that delegation of authority to a negotiator permits the union members to achieve advantageous commitment, while Conlin and Furusawa (2000), developing a game-theoretic model that endogenizes the items included in the bargaining agenda, show that in certain equilibria, the surplus-maximizing union selects wage-maximizing delegates (such as senior union members) to negotiate the contract. Goerke and Hefeker (2000) investigate under what conditions delegation of wage bargaining power is beneficial or harmful for union members and their potential leaders, while Mauleon and Vannetelbosch (2006) assume (in a firms' merger context) that surplus-maximizing unions have the option to delegate the wage bargaining to wage-maximizing delegates (such as senior union members), showing that, the option of strategic delegation may harm both the unions and the firms.

cheri and Fanti (2014) study how managerial delegation in a duopoly product market interact with wage decisions taken by a monopoly union in the labour market, focusing on the “sales delegation” with homogeneous product under Cournot and on both “sales delegation” and “relative profit delegation” with differentiated products under both Cournot and Bertrand, respectively, while Fanti and Meccheri (2015) and Meccheri and Fanti (2017a) analyse the endogenous choice between quantity or price contracts in a two-tier vertical structure where downstream firms’ owners delegate strategic decisions to managers and marginal production costs (i.e., wages) are bargained between managers and upstream suppliers (i.e., unions) and in which the managers or, alternatively, owners choose the type of contracts.

All above-mentioned contributions, considering multiple-stage delegation games with unions, assume that the sequence of moves is exogenously given: for instance, at the third stage managers compete oligopolistically in the product market, at the second stage wages for non-managerial workers are determined and at the first stage firms’ owners choose the managerial delegation contract.<sup>5</sup> However, Meccheri and Fanti (2017b) argue that, while when wages are bargained between managers and unions (e.g., Szymanski, 1994) such a sequence is a natural assumption, when wages are bargained by firms’ owners (e.g., Liao 2010, 2014) or under monopoly unions (e.g., Fanti and Meccheri 2013; Meccheri and Fanti, 2014) there is no (natural) reason to assume that managerial contracts are chosen by firms’ owners before wages have been set by unions. Hence, it is also argued that the stage at which firms’ owners decide the incentive contract of their managers should be – in the spirit of the game-theoretic approach – fully endogenised. Indeed, in a multiple-stage oligopoly game with managerial delegation and unions, Meccheri and Fanti (2017b) investigate in a Cournot duopoly with homogeneous product which sequence of moves (relative to managerial contracts and wages determination) endogenously arises by the strategic delegation game between owners. In particular, they assume that owners can freely decide to arrange the managerial contract before or after the (non-managerial) wage determination stage,<sup>6</sup> showing the occurrence of multiple (sub-game perfect) equilibria: for a given range of the unions’ preference parameter, in which unions are sufficiently but not too much employment oriented, the (less in-

<sup>5</sup> Note that the sequence of moves is assumed to be exogenous also in Fanti and Meccheri (2013) although they compared the equilibrium results obtained alternating the decisions about managerial contracts and wages between stage one and stage two.

<sup>6</sup> We defer to Meccheri and Fanti (2017b) for a broader discussion about this point.

tuitive) situation in which owners “move” after their unions in determining contracts can arise as a possible (sub-game perfect) equilibrium and it is also risk-dominant. This means that, quite counter-intuitively, firms’ owners may prefer not to play a leader position in determining managerial contracts with respect to their own unions, which set wage contracts.

In the present work, we extend Meccheri and Fanti (2017b) in a twofold way: 1) by developing the welfare analysis of their equilibrium results, and 2) by introducing product heterogeneity/differentiation (with products substitutes as well as complements) into the analysis, and re-analysing the endogenous determination of the sub-game perfect equilibria (SPNEs) relative to the choice of the incentive contract stage.

The welfare analysis in the case of exogenously given regimes shows that there is always conflict of interest between firms’ owners and consumers (and society), while unions (workers) may agree with consumers (and society) on the equilibrium in which both owners choose delegation contracts after (resp. before) union wage determination only when they are wage-oriented (resp. strongly employment-oriented). Instead, when we consider the SPNEs endogenously arising from this “timing” game between owners in determining the sequence of contracts, it is shown that, when unions are sufficiently (but not extremely) employment-oriented (and side-payments between owners are not feasible), the sequence of contracts presumably arising (because risk-dominant) from the owners’ choice maximises consumer surplus and overall welfare (even if profits are not maximised).

When we consider product differentiation, we get that with imperfect substitutes, similarly to what we obtain with homogeneous goods, there could be a unique equilibrium, in which both owners choose managerial contracts before unions set wages (*First-First* or *FF* equilibrium), as well as multiple symmetric equilibria, in which both owners choose manager contracts before or after (*Second-Second* or *SS* equilibrium) union wage setting. Although in all the areas the *FF* equilibrium is Pareto-dominant, a sufficiently large parametric set for which the *SS* equilibrium is risk-dominant does exist. When products are complements, a richer plethora of equilibria may occur: i) a unique *SS* equilibrium, which is also Pareto-dominant; ii) two asymmetric *SF* and *FS* equilibria, in which one owner chooses the manager contract before and the other owner after their respective union wage setting, which are Pareto-dominated by the *SS* equilibrium; and iii) a unique *FF* equilibrium, which, according to the structural parameters’ values, can be Pareto-dominant or Pareto-dominated only by the *SS* equilibrium or, instead, also by the asymmetric equilibrium outcome.

To sum up, this work complements previous work of ours (Meccheri and Fanti, 2017b), showing that the realistic case in which owners can freely decide the “timing” of their move, e.g., stipulating the managerial contract before or after the workers’ wage determination stage, has far-reaching welfare implications. Moreover, introducing product differentiation into the analysis leads to a richer set of endogenous equilibria, confirming in some cases the counter-intuitive preference of firms’ owners for a follower role with respect to their own unions in determining the sequence of managerial and wage contracts.

The remaining part of the work proceeds as follows. Section 2 recalls the set-up and the main results obtained by Meccheri and Fanti (2017b), which represents the basic framework of this article also. Section 3 explores the welfare implications of the results recalled in the previous section. Section 4 extends the result of Section 2 by introducing product differentiation into the analysis. Finally, Section 5 concludes, while further details are reported in the final Appendix.

## 2. Basic set-up

This paper builds on Meccheri and Fanti (2017b), which adopts a managerial delegation basic framework (Jansen *et al.*, 2007, 2009; van Witteloostuijn *et al.*, 2007) where we introduce the presence of unions in determining workers’ wages. In particular, we consider a normalised duopolistic Cournot market for a single homogeneous product, with inverse demand given by:

$$(1) \quad p(Q) = 1 - Q,$$

where  $p$  denotes price and  $Q$  is the sum of the firms’ output levels,  $Q = q_i + q_j$ , for  $i, j = 1, 2$  and  $i \neq j$ .<sup>7</sup>

We assume that both firms produce according to a standard production function with constant returns to labour  $q_i = L_i$ , where  $L_i$  represents the level of employment of firm  $i$ . Firm  $i$  faces a marginal (and average) cost  $0 \leq w_i < 1$  for any unit of output produced, where  $w_i$  is the per-worker wage. Therefore, firm  $i$ ’s cost function is linear and described by:

$$(2) \quad C_i(q_i) = w_i L_i = w_i q_i.$$

<sup>7</sup> Note that the standard inverse demand  $p' = a - bQ'$  can be obtained from this normalized model simply by fixing  $p = p'/a$  and  $Q = (b/a)Q'$ .

We assume that wages are monopolistically chosen by firm-specific unions, which have preferences weighted on wage and employment:

$$(3) \quad V_i = \omega_i^\theta L_i^{1-\theta} = \omega_i^\theta q_i^{1-\theta},$$

where  $\theta \in (0,1)$  is the relative weight placed by unions on wages with respect to employment.

The owners of both firms hire a manager and delegate the output decision to this manager. Each manager receives a fixed salary plus a bonus element, which is related to a weighted combination between firms' profits and sales: if firm  $i$ 's profits  $\pi_i$  are positive – otherwise there is no bonus – manager  $i$  receives a bonus that is *proportional* to  $u_i = \pi_i + b_i q_i$ , where the weight  $b_i$  is chosen by owner  $i$  to maximize profits and can be either positive or negative according to whether the owner provides incentives or disincentives to the manager's choice of output (sales).<sup>8</sup>

At the final stage (the market game), each manager sets the level of output for her/his firms to maximize  $u_i = \pi_i + b_i q_i = (1 - \omega_i - Q + b_i) q_i$ . However, before firms (managers) engage in competition, workers' wages should be set by unions and managerial contracts have to be defined by firms' owners. In contrast with the received literature, in Meccheri and Fanti (2017b) we do not assume *a priori* which stage (wage setting or managerial contract determination) first applies. Instead, we are interested in endogenously deriving the owners' choice of whether to define their managers' contracts before or after unions' wage setting. To do this, we calculate and compare the (equilibrium) profits for firms' owners in relation to any pair of possible strategies: i) both owner  $i$  and owner  $j$  choose their managerial contracts before unions' wage setting (*First-First* case); ii) both owner  $i$  and owner  $j$  choose their managerial contracts after unions' wage setting (*Second-Second* case); iii) owner  $i$  ( $j$ ) chooses before wage setting while owner  $j$  ( $i$ ) after wage setting (*Asymmetric strategies* case).

In particular, in Meccheri and Fanti (2017b) we show that the corresponding equilibrium output, wages and profits are:<sup>9</sup>

<sup>8</sup> We also follow the standard assumption made by managerial delegation theory that the fixed component (salary) of the manager's pay is chosen by the firm's owner such that the manager exactly receives his/her opportunity cost, which is normalized to zero.

<sup>9</sup> We defer to Meccheri and Fanti (2017b) for greater details and comments (see also Fanti and Meccheri 2013).

First-First case:

$$(4) \quad q^{FF} = \frac{(1-\theta)(4-\theta)}{(5-2\theta)(2+\theta)}; w^{FF} = \frac{3\theta(4-\theta)}{2(5-2\theta)(2+\theta)}; \pi^{FF} = \frac{(1-\theta)(4-\theta)(4+10\theta-5\theta^2)}{2(5-2\theta)^2(2+\theta)^2}.$$

Second-Second case:

$$(5) \quad q^{SS} = \frac{6(1-\theta)}{5(3-2\theta)}; w^{SS} = \frac{\theta}{3-2\theta}; \pi^{SS} = \frac{18(1-\theta)^2}{25(3-2\theta)^2}.$$

Asymmetric strategies case:

$$(6) \quad q^{FS} = \frac{3(4-\theta-\theta^2-2\theta^3)}{A}; w^{FS} = \frac{9\theta(4+3\theta+2\theta^2)}{2A};$$

$$\pi^{FS} = \frac{3(4-\theta-\theta^2-2\theta^3)(8+34\theta+25\theta^2+14\theta^3)}{2A^2};$$

$$(7) \quad q^{SF} = \frac{2(4+27\theta-39\theta^2+8\theta^3)}{A}; w^{SF} = \frac{\theta(8+65\theta+8\theta^2)}{A}; \pi^{SF} = \frac{2(4+27\theta-39\theta^2+8\theta^3)^2}{A^2},$$

where  $A=24+86\theta-55\theta^2+26\theta^3$  and, in the *Asymmetric strategies case*, the superscript FS (SF) identifies the equilibrium values for the firm that moves before (after) the union when the competitor behaves in the opposite way.

The following table represents the strategic form of such a “moves game” between owners. Specifically, F represents the strategy of moving first (i.e. choosing the delegation contract before wage setting), while S that of moving second (i.e., choosing the delegation contract after wage setting):

owner i \ owner j	F	S
F	$\pi^{FF}, \pi^{FF}$	$\pi^{FS}, \pi^{SF}$
S	$\pi^{SF}, \pi^{FS}$	$\pi^{SS}, \pi^{SS}$

and, given the equilibrium profits reported in (4), (5), (6) and (7), the following Lemma summarises the main propositions in Meccheri and Fanti (2017b), to which we defer for formal proofs, relative to the endogenous choice by owners of managerial contracts (with respect to union wage setting).

**Lemma 1 [Meccheri and Fanti 2017b, Props. 2 and 3].**

When unions are neither extremely oriented towards employment, nor sufficiently oriented towards wages (that is, when  $0.018 < \theta < 0.47$ ), there are two (Subgame Perfect) Nash equilibria, in which owners play symmetric strategies, i.e. First-First (FF) and Second-Second (SS) with the equilibrium FF, in which both owners move before their respective unions, Pareto-dominates the equilibrium SS, in which both owners move after their respective unions. When, instead, unions are either extremely employment-oriented or distinctly wages-oriented, the (Pareto-dominant) equilibrium FF only applies.

There exists a range for  $\theta$ , which is given by  $0.0973 < \theta < 0.2956$ , according to which the strategic situation SS risk-dominates FF. Otherwise, the Pareto-dominant equilibrium FF is also risk-dominant with respect to SS.

**3. Welfare**

In this paper, we want to explore the welfare effects of the choice by owners to choose managerial contracts before or after union wage setting. First, let us define the following key variables related to social welfare:

$$\begin{aligned} \Pi^{FF} &= 2\pi^{FF}; \quad \Pi^{SS} = 2\pi^{SS}; \quad \Pi^A = \pi^{FS} + \pi^{SF}; \\ CS^{FF} &= 2q^{FF^2}; \quad CS^{SS} = 2q^{SS^2}; \quad CS^A = \frac{(q^{FS} + q^{SF})^2}{2}; \\ WQ^{FF} &= 2w^{FF}q^{FF}; \quad WQ^{SS} = 2w^{SS}q^{SS}; \quad WQ^A = w^{FS}q^{FS} + w^{SF}q^{SF}; \\ SW^{FF} &= \Pi^{FF} + CS^{FF} + WQ^{FF}; \quad SW^{SS} = \Pi^{SS} + CS^{SS} + WQ^{SS}; \quad SW^A = \Pi^A + CS^A + WQ^A, \end{aligned}$$

where  $\Pi$  is the total profit obtained by firms, CS the consumer surplus, WQ the total wage bill and SW the overall welfare, while the superscript identify the different cases above analysed (First-First Case, Second-Second Case and Asymmetric Case, respectively).<sup>10</sup>

<sup>10</sup> Relative to the welfare components we consider, two points are worth noting. First, we refer to the sum of the two firms' profits. This implies that side-payments between owners could be required (in the asymmetric case) to coordinate their choices towards the maximising total profit situation. Second, we use the total wage bill instead of union utility in the welfare function. In this latter choice, we follow many others in the literature (e.g., Brander and Spencer, 1988; Mezzetti and Dinopoulos, 1991; Zhao, 2001) and this can be explained by the fact that unions' members are also final good consumers. Alternatively, since in our case the wage bill also corresponds to the union's rent (reservation wage is normalised to zero), the latter can be considered as a part of the producer surplus (Bughin and Vannini, 1995).

From the results outlined in Section 2, we can derive the following rankings in relation to those variables:

$$\begin{cases} \Pi^{FF} > \Pi^A > \Pi^{FF} & \text{if } \theta > 0.156 \\ \Pi^{FF} > \Pi^{SS} > \Pi^A & \text{if } \theta \leq 0.156 \end{cases} \quad \begin{cases} CS^{SS} > CS^A > CS^{FF} & \text{if } \theta > 0.0662 \\ CS^A > CS^{SS} > CS^{FF} & \text{if } \theta \leq 0.0662 \end{cases}$$

$$\begin{cases} WQ^{SS} > WQ^A > WQ^{FF} & \text{if } \theta \geq 0.5265 \\ WQ^A > WQ^{SS} > WQ^{FF} & \text{if } 0.2823 < \theta < 0.5265 \\ WQ^{FF} > WQ^A > WQ^{SS} & \text{if } \theta \leq 0.115 \end{cases} \quad \begin{cases} SW^{SS} > SW^A > SW^{FF} & \text{if } \theta > 0.0662 \\ SW^A > SW^{SS} > SW^{FF} & \text{if } \theta \leq 0.0662 \end{cases}$$

which are also graphically shown by Figure 1 below.

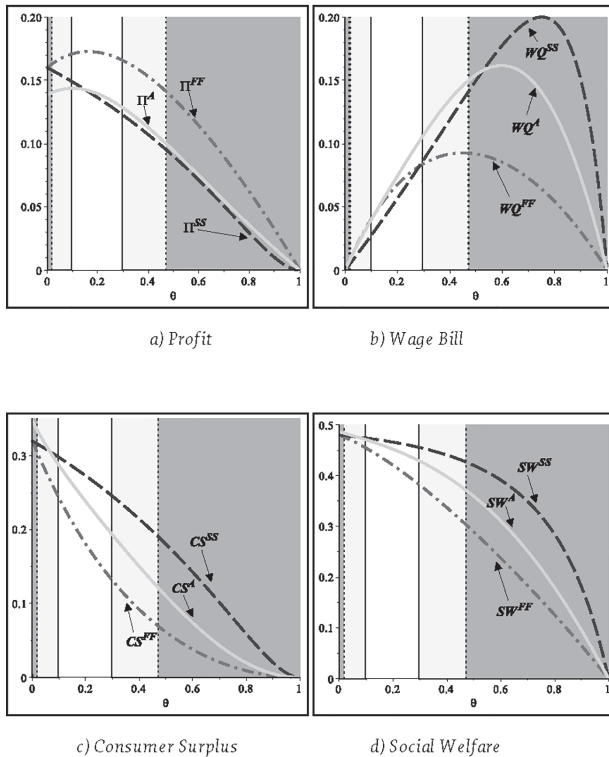


Figure 1. Welfare comparisons

Note: dotted-dashed line: FF-case; dashed line: SS-case; solid line: A-case. Grey areas: FF equilibrium; shaded areas: FF and SS equilibria (FF risk-dominant); white area: FF and SS equilibria (SS risk-dominant)



Figure 1 compares the behaviours, with respect to the union wage-orientation parameter  $\theta$ , of the alternative welfare components and of social welfare as a whole, applying in the different cases (dashed-dotted line in the *First-First Case*, dashed line in the *Second-Second Case* and solid line in the *Asymmetric Strategies Case*). Moreover, in the figure, the values of  $\theta$  are also specified (according to the different figure's shades), for which the alternative equilibrium configurations arise in the sequence of contracts, as summarized by Lemma 1 above. The following Lemma summarizes the main results on social welfare that arise from Figure 1.

## **Lemma 2**

*Relative to welfare comparisons under different possible sequence of contracts, the following results hold true:*

- *the sum of firms' profits is always higher when both owners choose managerial contracts before union wage setting (i.e., in the FF-Case);*
- *consumers and overall society mostly prefer the SS-Case, unless unions are strongly employment-oriented, in which case they prefer the Asymmetric-Case, that is, with an asymmetric length of wage and managerial contracts between firms;*
- *workers may prefer all possible cases, depending on their degree of orientation towards wages vis-à-vis employment: the SS-Case, when they are wage-oriented; the Asymmetric-Case, when they are moderately employment-oriented; and the FF-Case, when they are strongly employment-oriented.*

Moreover, from Lemma 2, we may straightforwardly derive the following corollaries, which distinguish between the general case (or the “exogenous regime”), i.e., independently from what endogenously arises as equilibrium in the game between owners, and what instead actually arises, according to the parameter  $\theta$ , in the endogenous equilibrium of that game (the “endogenous regime”).

## **Corollary 1 [exogenous regime]**

*From Lemma 2, independently from what actually arises in the endogenous game between owners in choosing the sequence of contracts, we generally get that:*

- *a conflict of interest between firms' owners and consumers (and society) always exists;*
- *workers and consumers (and society) both prefer that owners choose managerial contracts after union wage setting only when unions are wage-oriented;*
- *workers and firms' owners both prefer that owners choose managerial contracts before union wage setting only when unions are strongly employment-oriented.*

## Corollary 2 [endogenous regime]

*From Lemma 2, according to what actually arises in the endogenous game between owners in choosing the sequence of contracts, we get that, even if a conflict of interest between firms' owners and consumers (and society) always exists in the sequence of contracts, the endogenous choice presumably made (because risk-dominant) by owners when unions are sufficiently (but not extremely) employed-oriented (i.e., they both choose managerial contracts after wage setting) maximises consumer surplus and social welfare (even if profits are penalised).*

The above results, stated by Lemma 2 and Corollaries 1 and 2, deserve some more comments. First, the conflict of interest between firms' owners and consumers (as well as society), which is not eliminable in an exogenous context because total profits are always higher in the FF-case while consumer surplus and overall welfare are larger in the SS-case,<sup>11</sup> can be solved when a proper "contractual timing" game between firms' owners applies, provided that unions are sufficiently employment-oriented. Moreover, firms' owners and unions agree on (i.e., they prefer) the FF-case when the latter are strongly employment-oriented, and this holds true both in exogenous and endogenous cases (in the latter case the likely of "agreement" between them is only slightly reduced). Finally, it is also worth remarking that it does not exist a sequence of contracts which is Pareto-superior, that is, simultaneously preferred by all the agents (i.e., workers, owners, consumers and society as a whole). These findings offer an interesting empirical implication: when unions are strongly (resp. moderately) employment-oriented, it should more often be evidenced that the length of managerial contracts is longer (resp. shorter) than that of wage contracts of non-managerial workers.

## 4. Product differentiation

In this section, we extend the previous analysis to the case in which firms still compete in a Cournot framework but with differentiated products. Now we consider that each firm  $i$  is faced with the following (inverse) demand function:

$$(8) \quad p_i(q_i, q_j) = 1 - q_i - \gamma q_j,$$

<sup>11</sup> For the sake of precision, when unions are extremely employment-oriented, consumer surplus and social welfare are maximised in the Asymmetric-case. However, this (extreme) situation assumes minor importance.

where  $\gamma \in (-1, 1)$  represents the degree of product differentiation. Specifically when  $-1 < \gamma < 0$  products are complements (the higher  $|\gamma|$ , the higher the degree of complementarity), for  $0 < \gamma < 1$  products are substitutes (the higher  $\gamma$ , the higher the degree of substitutability), while for  $\gamma = 0$  they are independent.

In this context of differentiated products, standard calculations to obtain the sub-game perfect equilibrium values of bonus weight, wage, output and profit for the different cases lead to the following results (where the subscript  $D$  recalls that they refer to the product differentiation case), which parallel those of the homogenous product basic framework of Section 2 (Eqs. (4), (5), (6) and (7)):<sup>12</sup>

*First-First case with product differentiation:*

$$b_D^{FF} = -\frac{\gamma^4\theta^2 - 4\gamma^2(2\theta^2 - \theta + 1) + 16\theta}{2\gamma^4\theta^2 + 2\gamma^3\theta(\theta - 1) - 4\gamma^2(2\theta^2 + \theta + 1) - 8\gamma(\theta - 1) + 16(\theta + 1)};$$

$$w_D^{FF} = \frac{\theta[\gamma^4\theta - 4\gamma^2(\theta + 1) + 16]}{2\gamma^4\theta^2 + 2\gamma^3\theta(\theta - 1) - 4\gamma^2(2\theta^2 + \theta + 1) - 8\gamma(\theta - 1) + 16(\theta + 1)};$$

$$q_D^{FF} = \frac{(4 - \gamma^2\theta)(1 - \theta)}{2\gamma^4\theta^2 + 2\gamma^3\theta(\theta - 1) - 4\gamma^2(2\theta^2 + \theta + 1) - 8\gamma(\theta - 1) + 16(\theta + 1)};$$

$$\pi_D^{FF} = \frac{[\gamma^4\theta^2 - 2\gamma^2\theta(3\theta - 1) - 4\gamma^2 + 8(\theta + 1)](4 - \gamma^2\theta)(1 - \theta)}{2[\gamma^4\theta^2 + \gamma^3\theta(\theta - 1) - 2\gamma^2(2\theta^2 + \theta + 1) - 4\gamma(\theta - 1) + 8(\theta + 1)]^2}.$$

*Second-Second case with product differentiation:*

$$b_D^{SS} = \frac{(\theta - 1)\gamma^2(\gamma^2 - 4)}{(\gamma^2 + 2\gamma\theta - 4)(\gamma^2 - 2\gamma - 4)}; \quad w_D^{SS} = \frac{\theta(\gamma^2 + 2\gamma - 4)}{\gamma^2 + 2\gamma\theta - 4};$$

$$q_D^{SS} = \frac{2(\theta - 1)(\gamma^2 - 4)}{(\gamma^2 + 2\gamma\theta - 4)(\gamma^2 - 2\gamma - 4)};$$

$$\pi_D^{SS} = \frac{2(\theta - 1)^2(\gamma^2 - 2)(\gamma - 2)^2(\gamma + 2)^2}{(\gamma^2 + 2\gamma\theta - 4)^2(\gamma^2 - 2\gamma - 4)^2}.$$

<sup>12</sup> Standard intermediate calculations are omitted here for economy of space and are available from the authors upon request.

Asymmetric strategies case with product differentiation:

$$b_D^{FS} = -\frac{(\gamma^3\theta + 8\gamma^2 - 8\gamma\theta + 4\gamma - 32\theta)B}{2\Psi}; w_D^{FS} = \frac{\theta(8 + \gamma)(\gamma^2 - 4)B}{2\Psi};$$

$$q^{FS} = \frac{(\theta - 1)(8 + \gamma)B}{\Psi};$$

$$\pi_D^{FS} = \frac{(\theta - 1)(8 + \gamma)(\gamma^3\theta + 8\gamma^2 - 6\gamma\theta + 2\gamma - 16\theta - 16)B^2}{2\Psi^2};$$

and

$$b_D^{SF} = \frac{(\theta - 1)C}{3\Psi}; w_D^{SF} = \frac{\theta(8 + \gamma\theta)(\gamma^2 - 4)D}{3\Psi};$$

$$q_D^{SF} = \frac{2(\theta - 1)(4 - \gamma^2\theta)D}{\Psi};$$

$$\pi_D^{SF} = \frac{2(\theta - 1)^2(4 - \gamma^2\theta)D \times E}{3\Psi^2},$$

where  $B, C, D, E$  and  $\Psi$  are all polynomials ( $B, C, D$  and  $E$  always negative, while  $\Psi$  always positive) of very high degree, which are reported for convenience in the final appendix.

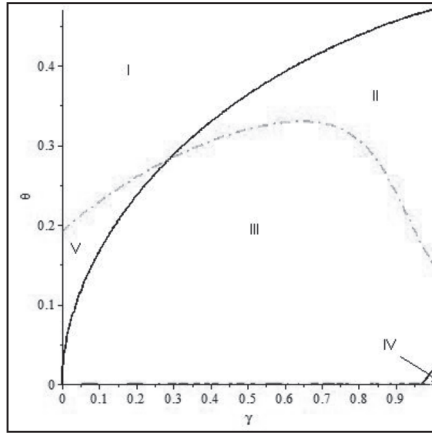
Now let us define the following profit differentials:

$$\Delta_1 = \pi^{FF} - \pi^{SF}; \Delta_2 = \pi^{SS} - \pi^{FS}; \Delta_3 = \pi^{FF} - \pi^{SS};$$

$$\Delta_4 = 2\pi^{FF} - (\pi^{FS} + \pi^{SF}); \Delta_5 = 2\pi^{SS} - (\pi^{FS} + \pi^{SF});$$

$$\Delta_6 = \pi^{FF} + \pi^{FS} - (\pi^{SS} + \pi^{SF}).$$

For the case of substitutes ( $\gamma \in (0, 1)$ ), Figure 2 below shows the signs of those differentials (which permit to establish the SPNEs of the game between owners on the sequence of contracts) according to the structural parameters  $\gamma$  and  $\theta$ .

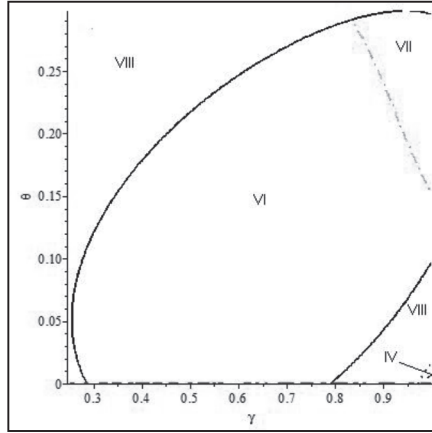


**Figure 2. Profit differentials signs and SPNEs. Substitutes**

Note:  $\Delta_2 = 0$ : solid line;  $\Delta_5 = 0$ : dashed-dotted line. Area I:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 < 0$ ; Area II:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 < 0$ ; Area III:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area IV:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area V:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ .

According to Figure 2, in the areas I, II, IV and V the unique equilibrium is *FF*, while in the area III multiple symmetric equilibria *FF* and *SS* occur. The equilibrium *FF* is always Pareto-dominant from the firms' owners viewpoint. Moreover, in the area III (resp. II) the possible equilibrium *SS* achieves a pay-off larger (resp. smaller) than that obtained by owners in the asymmetric equilibrium (of course provided that proper side-payments between them may be implemented). The latter fact may be relevant when the risk-dominance criterion is applied to select between multiple equilibria.

In Figure 3, it is shown how the area III of Figure 2 may be partitioned in three sub-areas – areas VI, VII, VIII: the former two represent the parametric set for which the equilibrium *SS* risk-dominates the *FF* one; instead, the area VIII relates to a situation where the equilibrium *FF* risk-dominates *SS* equilibrium. Moreover, while in the area VI the risk-dominant equilibrium *SS* is Pareto-dominated by the equilibrium *FF*, in the area VII the risk-dominant equilibrium *SS* is not only Pareto-dominated by the equilibrium *FF* (which still remains the Pareto-dominant one), but it would be inefficient also in comparison with the asymmetric equilibrium; in other words in the area VII the risk-dominant non-cooperative equilibrium *SS* results to be the least efficient one for firms.



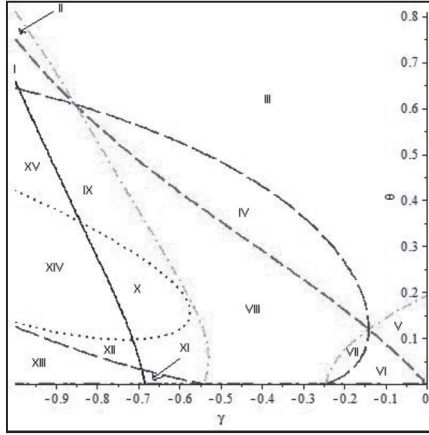
**Figure 3. Profit differentials signs, multiple equilibria and risk-dominance. Substitutes**

Note:  $\Delta_2 = 0$ : solid line;  $\Delta_5 = 0$ : dashed-dotted line. Area VI:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0, \Delta_6 < 0$ ; Area VII:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 < 0, \Delta_6 < 0$ ; Area VIII:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0, \Delta_6 > 0$ .

Figure 4, instead, refers to the case of complement products. According to the figure, and in particular to the signs of the profit differentials in different regions, we can infer that: in regions I, II, III, IV, V, VI, VII, VIII, IX and XI there exists a unique SPNE, in which both owners choose delegation contracts before union wage setting, i.e., a *FF* equilibrium; in area XIV there is a unique *SS* equilibrium, which is also Pareto-dominant; in areas XII, XIII and XV there are multiple symmetric equilibria *FF* and *SS*; and finally, in area X also two *asymmetric* equilibria *SF* and *FS*, do exist, which are Pareto-dominated by *SS*.

While in the areas II, III and V the endogenous equilibrium *FF* is Pareto-dominant, in regions I, VI, VII, IX and XI it is Pareto-dominated by *SS*, hence the former is Pareto-inefficient for firms' owners determining a prisoner's dilemma situation. Moreover, in the areas IV and VIII, the *FF* endogenous equilibrium is pay-off dominated by the asymmetric configuration outcome, while in the area X the asymmetric equilibrium achieves a lower pay-off than the *SS* configuration.

To sum up, only in the areas II, III, V and XIV the endogenous equilibrium (*FF* for the first three ones, *SS* for the latter one) is also Pareto-efficient for firms' owners.



**Figure 4. Profit differentials signs and SPNEs. Complements**

Note: Area I:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area II:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area III:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area IV:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 < 0, \Delta_5 < 0$ ; Area V:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 > 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area VI:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area VII:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area VIII:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area IX:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area X:  $\Delta_1 < 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area XI:  $\Delta_1 > 0, \Delta_2 < 0, \Delta_3 < 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area XII:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area XIII:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 > 0, \Delta_5 > 0$ ; Area XIV:  $\Delta_1 < 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ ; Area XV:  $\Delta_1 > 0, \Delta_2 > 0, \Delta_3 < 0, \Delta_4 < 0, \Delta_5 > 0$ .

## 5. Conclusion

In the present work, we have extended a previous work of ours (Meccheri and Fanti, 2017b) by developing the welfare analysis of the equilibrium results obtained therein and by introducing product heterogeneity/differentiation (with products substitutes as well as complements) into the analysis.

The welfare analysis has shown that when unions are sufficiently (but not extremely) employment-oriented (and side-payments between owners are not feasible), the sequence of contracts (presumably) arising from the owners' choice maximises consumer surplus and overall welfare (even if profits are not maximised). Hence, making the sequence of contracts chosen by the firms' owners endogenous can permit to bring out the preferred result for the society as a whole, which would never apply otherwise. Moreover, introducing product differentiation into the analysis has led to a richer set of endogenous equilibria, confirming in some cases the counter-intuitive preference of firms' owners for a follower role with respect to their own unions in determining the sequence of managerial and wage contracts.

## Appendix

In this Appendix we explicitly report the polynomials B, C, D, E and  $\Psi$  recalled in Section 4.

$$B \equiv 6\gamma^3\theta^2 - 5\gamma^3\theta - 8\gamma^2\theta^2 + 16\gamma^2\theta - 4\gamma\theta^2 - 20\gamma\theta + 20\gamma - 32$$

$$C \equiv \left\{ \begin{array}{l} 3\gamma^6\theta^2 + 4\gamma^5\theta^2 + 9\gamma^5\theta - 35\gamma^4\theta^2 - 70\gamma^4\theta - 48\gamma^3\theta^2 + 32\gamma^3\theta + 260\gamma^2\theta^2 - 36\gamma^3 + 168\gamma^2\theta \\ + 32\gamma\theta^2 - 264\gamma^2 - 80\gamma\theta + 48\gamma - 896\theta - 384 \end{array} \right\}$$

$$D \equiv 3\gamma^3\theta + 10\gamma^2\theta + 14\gamma - 29\gamma\theta + 17\gamma - 56\theta - 40$$

$$E \equiv \left\{ \begin{array}{l} -3\gamma^6\theta^2 - 22\gamma^5\theta^2 - 9\gamma^5\theta - 25\gamma^4\theta^2 - 14\gamma^4\theta + 222\gamma^3\theta^2 - 32\gamma^3\theta + 76\gamma^2\theta^2 \\ + 36\gamma^3 + 312\gamma^2\theta - 32\gamma\theta^2 + 72\gamma^2 - 616\gamma\theta + 360\gamma - 448\theta - 576 \end{array} \right\}$$

$$\Psi \equiv \left\{ \begin{array}{l} \gamma^6\theta^2 - 4\gamma^5\theta^3 + 12\gamma^5\theta^2 - 2\gamma^4\theta^3 + 8\gamma^5\theta - 36\gamma^4\theta^2 + 32\gamma^3\theta^3 + 94\gamma^4\theta - 128\gamma^3\theta^2 + 144\gamma^2\theta^3 \\ - 280\gamma^2\theta^2 + 64\gamma\theta^3 - 32\gamma^3 - 96\gamma^2\theta - 64\gamma\theta^2 - 264\gamma^2 + 256\gamma\theta + 512\theta + 512 \end{array} \right\}$$

## References

- Asseburg H., Hofmann C. (2010), "Relative performance evaluation and contract externalities", *OR Spectrum* 32, 1-20.
- Baumol W. (1958), "On the theory of oligopoly", *Economica* 25, 187-198.
- Berle A.A., Means G.C. (1932), *The Modern Corporation and Private Property*, Harcourt, Brace and World, New York.
- BIS (2011), *Executive remuneration*, Discussion paper, Department of Business, Innovation and skills, UK Government.
- Booth A.L. (1995), *The Economics of the Trade Union*, Cambridge University Press, Cambridge.
- Brander J.A., Spencer B.J. (1988), "Unionized oligopoly and international trade policy", *Journal of International Economics* 24, 217-234.
- Bughin J. (1995), "Unions and strategic managerial incentives", *Economics Letters* 47, 95-100.
- Bughin J., Vannini S. (1995), "Strategic direct investment under unionized oligopoly", *International Journal of Industrial Organization* 13, 127-145.
- Chatterjee I., Saha B. (2017), "Bilateral delegation in duopoly wage and employment bargaining", *Managerial and Decision Economics* 38, 607-621.
- Conlin M., Furusawa T. (2000), "Strategic delegation and delay in negotiations over the bargaining agenda", *Journal of Labor Economics* 18, 55-73.
- Conyon M. (1994), "Tenure and contracts: the experience of UK CEOs", *Personnel Review* 23, 25-33.
- Davidson C. (1988), "Multiunit bargaining in oligopolistic industries", *Journal of Labor Economics* 6, 397-422.
- Dowrick S. (1989), "Union-oligopoly bargaining", *Economic Journal* 99, 1123-1142.



## Endogenous sequence of contracts in managerial unionised duopolies

- European Commission (2006), "Policy rule evaluation by contract-makers: 100 years of wage contract length in Sweden", *Economic papers* n. 270.
- Fanti L., Meccheri N. (2013), "Managerial delegation under alternative unionization structures", *Labour* 27, 38-57.
- Fanti L., Meccheri N. (2015), "On the Cournot-Bertrand profit differential and the structure of unionisation in a managerial duopoly" *Australian Economic Papers* 54, 266-287.
- Fershtman C. (1985), "Managerial incentives as a strategic variable in duopolistic environment", *International Journal of Industrial Organization* 3, 245-253.
- Fershtman C., Judd K. (1987), "Equilibrium incentives in oligopoly", *American Economic Review* 77, 927-940.
- Goerke L., Hefeker C. (2000), "Delegation and wage determination in trade unions", *Labour* 14, 393-416.
- Horn H., Wolinsky A. (1988), "Bilateral monopolies and incentives for merger", *RAND Journal of Economics* 19, 408-419.
- Jansen T., van Lier A., van Witteloostuijn A. (2007), "A note on strategic delegation: The market share case", *International Journal of Industrial Organization* 25, 531-539.
- Jansen T., van Lier A., van Witteloostuijn A. (2009), "On the impact of managerial bonus systems on firm profit and market competition: The cases of pure profit, sales, market share and relative profits compared", *Managerial and Decision Economics* 30, 141-153.
- Jones S.R.G. (1989), "The role of negotiators in union-firm bargaining", *Canadian Journal of Economics* 22, 630-642.
- Liao P.C. (2010), "Strategic delegation under unionised duopoly: Who will bargain with unions?", *Australian Economic Papers* 49, 276-288.
- Liao P.C. (2014), "Strategic delegation of multiple tasks", *Australian Economic Papers* 53, 77-96.
- Liu P., Xuan Y. (2016), "The contract year phenomenon in the corner office: an analysis of firm behavior during CEO contract renewals", Mimeo, downloaded at <http://www.business.illinois.edu/yhxuan/LiuXuanContractYearPhenomenon.pdf>
- Meccheri N., Fanti L. (2014), "Managerial delegation contracts under centralized unionization", *Managerial and Decision Economics* 35, 51-66.
- Meccheri N., Fanti L. (2017a), "A note on endogenous competition mode with managerial-unionized firms", *Metroeconomica* (forthcoming - published online).
- Meccheri N., Fanti L. (2017b), "Should managerial delegation contracts be made before or after union wage setting? A game-theoretic analysis", *Managerial and Decision Economics* (forthcoming - published online).
- Mezzetti C., Dinopoulos E. (1991), "Domestic unionization and import competition", *Journal of International Economics* 31, 79-100.
- Miller N., Pazgal A. (2002), "Relative performance as a strategic commitment mechanism", *Managerial and Decision Economics* 23, 51-68.
- Overvest B.M., Veldman J. (2008), "Managerial incentives for process innovation", *Managerial and Decision Economics* 29, 539-545.
- Salas-Fumás V. (1992), "Relative performance evaluation of management", *International Journal of Industrial Organization* 10, 473-489.
- Sklivas S. (1987), "The strategic choice of managerial incentives", *RAND Journal of Economics* 18, 452-458.

- Szymanski S. (1994), "Strategic delegation with endogenous costs. A duopoly with wage bargaining", *International Journal of Industrial Organization* 12, 105-116.
- van Witteloostuijn A., Jansen T., van Lier A. (2007), "Bargaining over managerial contracts in delegation games: Managerial power, contract disclosure and cartel behaviour", *Managerial and Decision Economics* 28, 897-904.
- Vickers J. (1985), "Delegation and the theory of the firm", *Economic Journal* 95, 138-147.
- Zhao L. (2001), "Unionization, vertical markets, and the outsourcing of multinationals", *Journal of International Economics* 55, 187-202.

# On the effects of firing costs on employment and welfare in a duopoly market with entry\*

SIMONE D'ALESSANDRO, NICOLA MECCHERI AND DAVIDE M. COLUCCIA

## 1. Introduction

The aim of this paper is to investigate whether the conventional detrimental effect that employment legislation would have on employment and overall welfare holds true in a strategic competitive environment, where an incumbent chooses its optimal output given the potential entry of a second firm.<sup>1</sup> As largely discussed in the literature, forms of employment protection legislation (EPL) vary across countries. Nevertheless, all such systems imply that firms have to bear some costs associated to firing workers.

While there appears to be no clear-cut empirical relationship between (average) unemployment rate and the degree of EPL legislation, from a theoretical perspective it can be argued that firing costs may have ambiguous effects on employment (Cahuc and Postel-Vinay, 2002). On the one hand, they are in fact beneficial for already employed workers, who enjoy a safer income

\* We would like to thank seminar participants at the Workshop “Oligopoly, Institutions and Firm Performance” (Pisa, 26-27 January 2017) financed by the University of Pisa (Project: PRA 2016\_2) for useful comments and suggestions. The usual disclaimers apply.

<sup>1</sup> First quantitative measures on EPL in different countries have been provided by Bertola (1990) and Grubb and Wells (1993). See also Layard and Nickell (1999) and OECD (1999, 2004).

source. On the other, however, excessive firing costs may lower the propensity of a firm to employ, given the rigidities that result from a frictional labour market. Thus, there have been various concerns about the possibility of increasing the cost of laying off workers, in the light of such adverse effect.

It is even more difficult to ascertain the effects of an increase in firing costs on employment in a competitive environment, in which decisions on entry and output depend on those undertaken by the potential competitor. While employment is likely to benefit from stronger competition, as industry output overall increases, there may be some causes of concern for the existing firms. To accommodate new entrants, they would in fact have to reduce their scale of operation, thus bearing additional costs depending on the level of firing ones. It is therefore of interest to understand the role such costs play in this context.

In this paper, we study a two-stage duopoly game, with an incumbent and an entrant. At the first stage, the incumbent sets its output level for the period, while at the second the entrant decides whether to enter the market or not. Entry decision leads to competition on quantity (à la Cournot) between firms, so that the incumbent can decide whether to revise its production choices or leave them unchanged with respect to those of the first period. However, output reduction is costly due to the presence of EPL, which translates into firing costs: these influence entry decision, since the incumbent may be unwilling to downsize its production leading to deleterious effects on potential entrant's profits, as well as both firms' decisions in both periods. We show that employment and welfare actually increase with firing costs provided that these lie within a certain interval, over which firms' decisions are no longer related to the level of the EPL.

The intuition underlying our result is pretty much straightforward and can be exemplified as follows. In a dynamic environment, firing costs act as a commitment for the incumbent firm, which binds itself to produce a given quantity of output as long as the loss in profits coming from a production downsizing does not offset the costs implied by that downsizing, which are in fact due to EPL. Through such a commitment output, and hence employment, is thus stabilized, and it can be further argued that the final level of employment can, under some conditions, increase with firing costs. Moreover, overall welfare is maximised for a strictly positive level of firing costs.

This paper thus contributes to the literature that originates from the seminal works by Lazear (1990) and Bentolila and Bertola (1990) which investigates the relationship between firing costs and employment or, more broadly, social welfare. Lazear (1990) shows that firing costs have no effect on employment in a perfect-

ly competitive labour market with flexible wages. Instead, Bentolila and Bertola (1990) consider the effects of firing costs on employment in the presence of rigid wages and uncertainty on (labour) productivity shocks. They compare a flexible economy, with no firing costs, and a rigid one, and show that while average long-run employment is left unaffected by such labour market frictions, short-run employment is higher in the rigid system, provided that the realized productivity shocks are small. The opposite holds when the latter are high instead.

More recently, EPL has been proved to be a “second best” solution in economies with risk-averse workers, insurance and capital markets imperfections (e.g. Pissarides, 2001). However, none of the above-mentioned works considers the role of strategic interaction among firms in imperfectly competitive markets in determining the relation between firing costs and employment.

In Lommerud and Straume (2012) the role of employment protection in the form of firing costs is analysed and compared against a different labour market institution, i.e. “flexicurity”, in a monopolistic product market in the presence of a labour union. In such a framework, strategic interaction between the firm’s output decision in the product market and the union’s wage decision in the labour market is crucial and the role of firing costs in affecting those decisions is relevant. Lommerud and Straume’s (2012) main focus is nonetheless on the effect of alternative labour market institutions on trade unions’ incentives to oppose or endorse labour-saving technology and on firms’ incentives to invest in such technology. Increased flexicurity – interpreted as less employment protection and a higher reservation wage for workers – unambiguously increases firms’ incentives for technology adoption. Moreover, a higher reservation wage generally makes unions more willing to accept technological change, while less employment protection has the opposite effect, as this increases the downside (job losses) of labour-saving technology. However, the possibility for new firms entering the market is not contemplated, so that its effects on employment in the presence of firing costs (which represent the objective of this paper) are not investigated.

To the best of our knowledge, the paper which is most closed to ours is Majumdar and Saha (1998). They examine the implications of job security in the context of entry and wage bargaining. Assuming that production requires industry-specific skills, which can only be acquired through training, and that jobs are secure by law, the entrant will actually enter the market if and only if it can attract workers by offering a wage higher than the unionised wage paid by the incumbent. In this context, the authors show that job security may lead to an unexpected outcome, in which duopolistic competition conveys a lower welfare than the one attained in monopoly.

However, in Majumdar and Saha (1998) job security is more strictly defined as the incumbent's absolute impossibility to lay off workers, which is equivalent to assume that firing costs are infinite. Instead, we admit that firing costs, or the degree of EPL, can vary. This allows us to analyse in detail how firms modify their employment decisions with respect to different levels of firing costs, hence clarifying which type of – possibly non-monotonic – relationship actually holds between firing costs and employment.

The remainder of the paper is organized as follows. In the second section, we formally present our model. We split the discussion between the second and the first period and provide the main results. In the third section, we sketch some considerations on the effect of EPL on total employment, thus presenting the main results of the work. We end up with some conclusions and possible future research developments, while more technical details are relegated in a final appendix.

## 2. Model

We consider a simple dynamic framework with two stages (production periods) and two firms: firm 1 as the incumbent and firm 2 as the entrant. Firms face the following linear demand curve:

$$p = a - Q \tag{1}$$

where  $p$  denotes the price,  $Q = q_i + q_j$  total quantity, with  $i, j \in (1, 2)$ ,  $i \neq j$ , and  $a > 0$  a positive parameter. Firms operate under constant returns to scale technology, using labour as the only input. We can thus normalize firm  $i$ 's output in terms of the labour units it employs. Therefore, we have  $q_i = l_i$  where  $l_i$  is the firm  $i$ 's level of employment. Analogously, firms' total production costs are given by  $C_i = w l_i = w q_i$ , where  $w$  is the wage. We assume that  $w$  is uniform across firms and periods<sup>2</sup>.

We study the following game. At stage 1 ( $t = 1$ ), the incumbent decides its level of output for the first period. At stage 2 ( $t = 2$ ), the entrant decides whether to enter the market. If it decides to enter the market, it bears a fixed cost given by  $F$ . Then, the two firms compete à la Cournot, i.e. on quantities. Firm 1 will thus in general revise its production plan with respect to those whereof the first stage. However, employment protection, as it has already been discussed, makes it costly for the incumbent to downsize employment. Fol-

<sup>2</sup> This can be interpreted as if the firms face a perfectly inelastic supply of labor schedule.

lowing Lommerud and Straume (2012), we model the degree of employment protection through a parameter  $\phi > 0$  representing unitary firing costs that the incumbent must pay for each worker it wishes to dismiss with respect to the production level at the first stage. We solve the game through backward induction. To simplify the notation, we define  $k_1$  the quantity produced by firm 1 at the first stage and  $q_1, q_2$  the quantity of the two firms at stage 2. Furthermore, without loss of generality, we define  $\alpha \equiv a - w$ .

Clearly, if firm 2 does not enter the market and firm 1 is the sole operator, it is trivial to check that firm 1 acts as a monopolist in both production periods. This yields to the following equilibrium values for outputs and profits (which are constant across periods):

$$k_1 = q_1 = \frac{\alpha}{2}; \quad \pi_1 = \frac{\alpha^2}{4}.$$

Let us consider instead the case in which firm 2 enters the market. Then, at stage 2, firms' profits are given by:

$$\pi_1 = \begin{cases} (\alpha - q_1 - q_2)q_1 - \phi(k_1 - q_1) & \text{if } q_1 \leq k_1, \\ (\alpha - q_1 - q_2)q_1 & \text{if } q_1 > k_1, \end{cases} \quad (2)$$

$$\pi_2 = (\alpha - q_1 - q_2)q_2 - F. \quad (3)$$

Firms choose output to maximize (2) and (3), respectively. Given the quantity produced by firm 1, the reaction function of firm 2 is given by:

$$q_2(q_1) = \frac{\alpha - q_1}{2}. \quad (4)$$

The reaction function for firm 1 is obtained given the quantity produced by firm 1 in the first stage and given the quantity produced by firm 2. This yields to<sup>3</sup>:

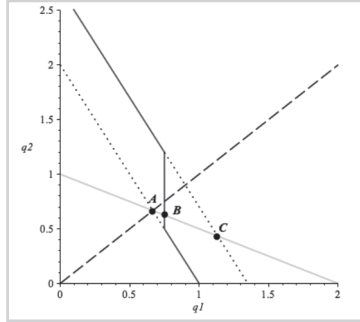
$$q_1(q_2) = \begin{cases} \frac{\alpha - q_2 + \phi}{2} & \text{if } q_1 < k_1 \\ k_1 & \text{if } q_1 = k_1, \\ \frac{\alpha - q_2}{2} & \text{if } q_1 > k_1 \end{cases} \quad (5)$$

The reaction functions whereof (4) and (5) are shown in Figure 1. At the second stage, instead, the equilibrium depends on the firm 1's level of output at the first stage. Indeed, by solving the system of equations (4) and (5), we get the equilibrium outputs for both firms 1 and 2:

<sup>3</sup> Notice that this is similar to Dixit (1980).

$$q_1^* = \begin{cases} \frac{\alpha + 2\phi}{3} & \text{if } k_1 > \frac{\alpha + 2\phi}{3} \\ k_1 & \text{if } \frac{\alpha}{3} \leq k_1 \leq \frac{\alpha + 2\phi}{3} ; \\ \frac{\alpha}{3} & \text{if } k_1 < \frac{\alpha}{3} \end{cases} \quad (6)$$

$$q_2^* = \begin{cases} \frac{\alpha - \phi}{3} & \text{if } k_1 > \frac{\alpha + 2\phi}{3} \\ \frac{\alpha - k_1}{2} & \text{if } \frac{\alpha}{3} \leq k_1 \leq \frac{\alpha + 2\phi}{3} . \\ \frac{\alpha}{3} & \text{if } k_1 < \frac{\alpha}{3} \end{cases} \quad (7)$$



**Figure 1. The reaction functions at the second stages. The three equilibria - A, B, C - represent the possible equilibrium outcome depending on the value of  $k_1$ . In this case the realized equilibrium is  $B = \left\{ k_1, \frac{\alpha - k_1}{2} \right\}$ . Value of parameters:  $\alpha = 2, \phi = 0.7, k_1 = 0.75$**

Hence, taking (1), (3), (6) and (7) into account, firm 2's profits are given by:

$$\pi_2^*(k_1) = \begin{cases} \frac{(\alpha - \phi)^2}{9} & \text{if } k_1 > \frac{\alpha + 2\phi}{3} \\ \frac{(\alpha - k_1)^2}{4} & \text{if } \frac{\alpha}{3} \leq k_1 \leq \frac{\alpha + 2\phi}{3} . \\ \frac{\alpha^2}{9} & \text{if } k_1 < \frac{\alpha}{3} \end{cases} \quad (9)$$

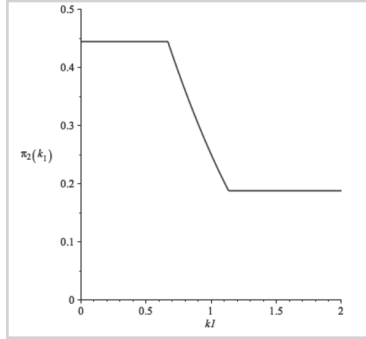
Figure 2 shows the equilibrium profits of firm 2 (without fixed costs) as a function of  $k_1$ . Trivially, if  $F > \frac{\alpha^2}{9}$ , that is, the entry cost is larger than profits in a symmetric duopoly, firm 2 never enters the market. On the contrary, if  $F < \frac{(\alpha - \phi)^2}{9}$ , then firm 2 always enters the market, provided that  $\phi < \alpha - 3\sqrt{F}$ . Between these two values of  $F$ , there is a level of production at the first stage  $\hat{k}$ , such that if  $k_1 \leq \hat{k}$  ( $k_1 > \hat{k}$ ), firm 2 enters (does not enter) the market. In what follows we assume that  $F < \frac{(\alpha - \phi)^2}{9}$ .

<sup>4</sup> This assumption allows us to avoid the analysis of entry deterrence strategy. Although we think that this issue is very interesting, we focus on this simpler case to more clearly isolate the effect of firing costs in duopoly equilibrium.



Similarly, firm 1's profits at the second stage are:

$$\pi_1'(k_1) = \begin{cases} \frac{(\alpha+2\phi)(3+\alpha-\phi)}{9} - \phi \left( k_1 - \frac{\alpha+2\phi}{3} \right) & \text{if } k_1 > \frac{\alpha+2\phi}{3} \\ \frac{(\alpha-k_1)k_1}{2} & \text{if } \frac{\alpha}{3} \leq k_1 \leq \frac{\alpha+2\phi}{3} \\ \frac{\alpha^2}{9} & \text{if } k_1 < \frac{\alpha}{3} \end{cases} \quad (10)$$



**Figure 2.** Gross profit at equilibrium of firm 2 as a function of the production level of firm 1 at the first stage ( $k_1$ ). Values of parameters:  $\alpha = 2$ ,  $\phi = 0.7$

At stage 1, the incumbent chooses output (employment) by maximizing the following profit function:

$$\Pi_1 = v_1(k_1) + \delta \pi_1'(k_1) \quad (11)$$

where  $v_1(k_1) \equiv (\alpha - k_1)k_1$  is the profit at the first stage (as a monopolist), and  $\delta$  is the discount factor. In the first stage, firm 1 thus chooses  $k_1$  to maximize (11) also taking into account the consequences that such choice yields with respect to stage 2. Total profit function is given by three segment of parabola that are determined by (10). This function is continuous and cannot take the maximum for  $k_1 < \frac{\alpha}{3}$ . We summarize these considerations as follows.

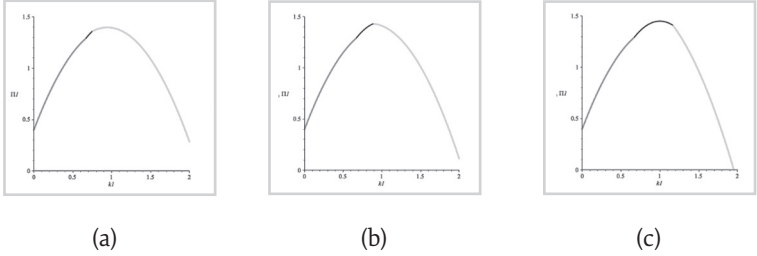
### Remark 1

- For  $k_1 < \frac{\alpha}{3}$ , total profits are given by the function  $(\alpha - k_1)k_1 + \frac{\alpha^2}{9}$ . The maximum is reached on  $k_1 = \frac{\alpha}{2}$ , and the profit is monotonically increasing until this value. However, since  $\frac{\alpha}{2} > \frac{\alpha}{3}$ , the only candidate is  $k_1 = \frac{\alpha}{3}$ .
- For  $\frac{\alpha}{3} \leq k_1 \leq \frac{\alpha+2\phi}{3}$ , total profits are given by  $(\alpha - k_1)k_1 + \frac{(\alpha - k_1)k_1}{2}$ . Again, this function takes its maximum value for  $k_1 = \frac{\alpha}{2}$  and it is increasing for

$\frac{\alpha}{3} \leq k_1 < \frac{\alpha}{2}$ . Thus  $k_1 = \frac{\alpha}{3}$  cannot be a maximum, while candidates are  $k_1 = \frac{\alpha}{2}$  (if  $\frac{\alpha}{2} < \frac{\alpha+2\phi}{3}$ ) and  $k_1 = \frac{\alpha+2\phi}{3}$  (if  $\frac{\alpha}{2} \geq \frac{\alpha+2\phi}{3}$ ).

For  $k_1 > \frac{\alpha+2\phi}{3}$  total profits are given by  $(\alpha - k_1)k_1 + \frac{(\alpha+2\phi)(3+\alpha-\phi)}{9} - \phi \left( k_1 - \frac{\alpha+2\phi}{3} \right)$ . This function takes its maximum value for  $k_1 = \frac{\alpha - \delta\phi}{2}$ . Hence, there are two candidates:  $k_1 = \frac{\alpha+2\phi}{3}$  (if  $\frac{\alpha - \delta\phi}{2} \leq \frac{\alpha+2\phi}{3}$ ) and  $k_1 = \frac{\alpha - \delta\phi}{2}$  (if  $\frac{\alpha - \delta\phi}{2} > \frac{\alpha+2\phi}{3}$ ).

Accordingly, there are three candidates for the maximum:  $k_1^A = \frac{\alpha}{2}$ ,  $k_1^B = \frac{\alpha - \delta\phi}{2}$  and  $k_1^C = \frac{\alpha+2\phi}{3}$ , where  $k_1^A$  and  $k_1^B$  are the maximum of the second and third segment of parabolas of the total profit function, and  $k_1^C$  is the interception between those two segments of the two parabolas. Figure 3 show the possible three cases that can emerge depending on the value of firing cost  $\phi$ .



**Figure 3. Firm 1's profit functions. Values of parameters:**  
 $\alpha = 2, \delta = 0,9$ , **(a):**  $\phi = 0,125$ ; **(b):**  $\phi = 0,\bar{6}$ ; **(c):**  $\phi = 0,75$

### Proposition 1

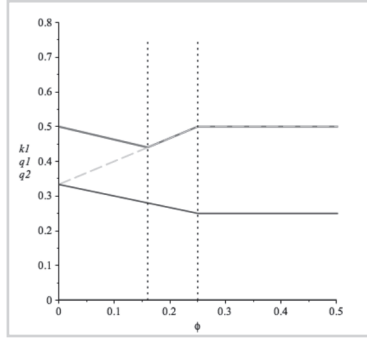
If firing costs are low, that is  $0 < \phi < \frac{\alpha}{4+3\delta}$ , then optimal first and second period output is given by  $k_1^* = \frac{\alpha - \delta\phi}{2}, q_1^* = \frac{\alpha+2\phi}{3}, q_2^* = \frac{\alpha - \phi}{3}$ . For values of the firing costs lying in between the interval defined by  $\frac{\alpha}{4+3\delta} \leq \phi \leq \frac{\alpha}{4}$ , firm 1 does not modify its optimal output between the stages, which is given by  $k_1^* = \frac{\alpha+2\phi}{3} = q_1^*$ , while  $q_2^* = \frac{\alpha - \phi}{3}$ . For high firing costs  $\frac{\alpha}{4} < \phi$ , output is chosen regardless to such costs, yielding  $k_1^* = \frac{\alpha}{2} = q_1^*, q_2^* = \frac{\alpha}{4}$ .

### Proof:

See Appendix 1. ■

While the formal proof of Proposition 1 is provided in the final appendix, Figure 4 shows a graphical analysis of the results stated in the proposition.

On the effects of firing costs on employment and welfare



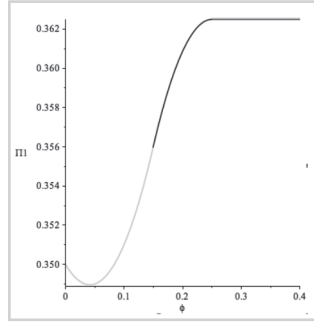
**Figure 4. Equilibrium values of  $k_1^*$ ,  $q_1^*$ ,  $q_2^*$  as a function of  $\phi$ . The two vertical dashed thresholds are  $\phi = \frac{\alpha}{4+3\delta}$  and  $\phi = \frac{\alpha}{4}$ . Values of parameters:  $\alpha = 1$ ,  $\delta = 0.75$**

Proposition 1 points out that the effects implied by the presence of firing costs are diametrically different depending on the period we consider. Indeed, stronger employment protection makes it more costly for the firm to operate with a large workforce in the first period, given the possibility of second-period downsizing in the light of enhanced competitive pressure. Thus, larger firing costs yield lower labour demand in the first period (Figure 3.a)<sup>5</sup>. However, this result only holds for relatively low levels of firing costs. Indeed, if  $\phi$  increases above the threshold  $\frac{\alpha}{4+3\delta}$ , firm 1 can credibly sustain the same production level in both the periods (Figure 3.b). This quantity increases with  $\phi$  until the Stackelberg equilibrium is reached (Figure 3.c).

In the second period, instead, the level of production of firm 1 increases with  $\phi$  until  $\frac{\alpha}{2}$ . This increase is always sufficient to offset the decrease in production at the first stage. Indeed one can check that firm 1 wholly produces  $k_1^* + q_1^* = \frac{\alpha - \delta\phi}{2} + \frac{\alpha + 2\phi}{3} = \frac{5\alpha + (4 - 3\delta)\phi}{6}$ , which is always increasing in  $\phi$ . Thus, total quantity produced by firm 1 is increasing for any  $0 < \phi \leq \frac{\alpha}{4}$ , and constant for  $\phi > \frac{\alpha}{4}$ . Moreover, firm 1's equilibrium profit can be easily obtained substituting  $k_1^*$  and  $q_1^*$  in the profit function. We get that profit are decreasing for  $0 < \phi < \frac{\alpha}{9\delta + 16}$ , increasing for  $\frac{\alpha}{9\delta + 16} \leq \phi \leq \frac{\alpha}{4}$ , and constant for  $\phi > \frac{\alpha}{4}$ , as shown in figure 5.

<sup>5</sup> This parallels the result obtained by Lommerud and Straume (2012) in a monopoly model, where the possibility of second-period downsizing relates to the introduction of a new labor-saving technology. As stated by Lommerud and Straume (2012, p. 187), "This illustrates - in a very simple framework - the standard concern about the dynamic employment effects of employment protection legislation: if the cost of laying off workers is increased, this will make firms less willing to hire workers in the first place".

Relative instead to firm 2, as shown in Figure 4, the quantity produced is decreasing in the firing costs for  $0 < \phi \leq \frac{\alpha}{4}$  and constant for  $\phi > \frac{\alpha}{4}$  and, given equation (9), also its profit follows the same path.



**Figure 5. Firm 1's total profit as a function of firing costs  $\phi$ .**  
**Values of parameters:  $\alpha = 1, \delta = 0.9$**

### 3. Employment and social welfare

In this section, we investigate the effects of firing costs on total employment. These are not clear-cut in our framework. As stated by Proposition 1, and given the assumption that  $q_1 = 1$ , unless  $\phi > \frac{\alpha}{4}$  (for which both firms' employment is not related to firing costs), employment dynamics are not monotonic with respect to the level of employment protection. The following proposition clarifies the effect of firing costs on total employment.

#### Proposition 2

The behaviour of second-period total employment with respect to firing costs parallels that of the incumbent's second-period output (employment): it is increasing for  $\phi \leq \frac{\alpha}{4}$  and unrelated to firing costs for  $\phi > \frac{\alpha}{4}$ . Instead, overall employment (i.e. second-period total employment plus incumbent's first-period employment) is:

- decreasing in  $\phi$  for low levels of firing costs, i.e.  $0 \leq \phi < \frac{\alpha}{4}$ , if and only if  $\delta > \frac{2}{3}$ ;
- unrelated to  $\phi$  for low levels of firing costs if  $\delta \leq \frac{2}{3}$  and for high levels of firing costs, i.e.  $\phi \geq \frac{\alpha}{4}$ ;
- increasing in  $\phi$  for low level of firing costs if  $\delta < \frac{2}{3}$ , and for medium levels of firing costs, i.e.  $\frac{\alpha}{4+3\delta} \leq \phi < \frac{\alpha}{4}$ .

**Proof:**

Proposition 1 shows that for  $0 < \phi < \frac{\alpha}{4+3\delta}$ , we have that  $k_1^* = \frac{\alpha - \delta\phi}{2}$ ,  $q_1^* = \frac{\alpha + 2\phi}{3}$ ,  $q_2^* = \frac{\alpha - \phi}{3}$ . Thus, we get that total employment  $L = k_1^* + q_1^* + q_2^* = \frac{7\alpha + (2-3\delta)\phi}{6}$ . It is easy to show that  $\frac{\partial L}{\partial \phi} < 0 \Leftrightarrow \delta > \frac{2}{3}$ . Following Proposition 1, for  $\phi > \frac{\alpha}{4+3\delta}$ , the results are trivial. ■

**Corollary**

From Proposition 2, it is easy to infer that the maximum level of total employment is obtained for  $\phi > \frac{\alpha}{4}$ , where  $L = \frac{5}{4}\alpha$ .

Relative to social welfare, we define it as the sum of consumer surplus and profits for the two stages, given a discount rate. That is:

$$SW = \frac{k_1^*}{2} + \delta \frac{(q_1^* + q_2^*)^2}{2} + \Pi_1 + \delta \Pi_2 \tag{11}$$

Taking equilibrium values for  $k_1^*$ ,  $q_1^*$  and  $q_2^*$ , as defined by Proposition 1, and the corresponding equilibrium profits, the following proposition applies for social welfare:

**Proposition 3**

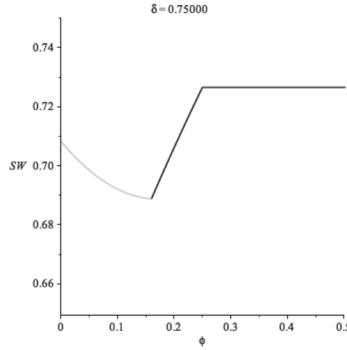
The social welfare is:

- decreasing in the firing costs for  $0 < \phi < \frac{\alpha}{4+3\delta}$ ;
- increasing in the firing costs for  $\frac{\alpha}{4+3\delta} \leq \phi < \frac{\alpha}{4}$ ;
- constant for  $\phi \geq \frac{\alpha}{4}$ .

Moreover, the maximum level of social welfare sustained as a Subgame Perfect Nash Equilibrium is obtained for  $\phi = \frac{\alpha}{4}$ .

Figure 6 provides a graphical proof of Proposition 3 and, in particular, that social welfare is maximised for a strictly positive level of firing costs (namely,  $\phi = \frac{\alpha}{4}$ )<sup>6</sup>.

<sup>6</sup> This result can be checked by noting that the differential  $SW|_{\phi = \alpha/4} - SW|_{\phi = 0} = 7/144\alpha^2 > 0$ .



**Figure 6. Social welfare as a function of firing costs. Values of parameters:  $\alpha=1, \delta=0.75$**

## 4. Conclusion

In this paper, we aimed at providing a first step into the analysis on the role of firing costs in affecting employment and welfare in strategic competitive environments. In particular, we have analysed a duopoly model where an incumbent chooses its optimal output given the potential entry of a second firm. In such a context, the role of firing costs is not trivial since the incumbent may be unwilling to downsize its production and this can affect the potential entrant's decision. Indeed, we have shown that incumbent employment actually increases with firing costs provided that these are above a certain lower bound. Moreover, the total output and overall welfare in the market are strictly increasing in firing costs if the latter lie within a certain interval, over which firms' decisions are no longer related to the level of firing costs. This results can be explained by the fact that, in a dynamic environment, the incumbent exploits the presence of firing costs to credibly commit high level of production in the second stage, hence firing costs act as a commitment for the incumbent firm that contributes to stabilize output, and hence employment, which also leads to a result in which overall welfare is maximised for a strictly positive level of firing costs.

Since this paper has represented a first attempt to analyse the role of firing costs in a strategic environment, further research directed to extend our basic model and assess the robustness of our results can be carried out along possible different lines. For instance, analysing the effects of introducing strategic entry and endogenous wage determination under alternative unioniza-

tion regimes, or comparing the effects of firing costs against different labour market institutional features (e.g. “flexicurity”) deserve to be realized.

### Appendix: proof of Proposition 1

From Remark 1 we get that total profits are maximized for  $k_1^* = \frac{\alpha - \delta\phi}{2}$  when  $\frac{\alpha - \delta\phi}{2} > \frac{\alpha + 2\phi}{3}$ . Solving the inequality with respect to  $\phi$  we get:

$$\frac{\alpha - \delta\phi}{2} > \frac{\alpha + 2\phi}{3} \Leftrightarrow \phi < \frac{\alpha}{4 + 3\delta}.$$

Hence, for  $0 < \phi < \frac{\alpha}{4 + 3\delta}$ ,  $k_1^* = \frac{\alpha - \delta\phi}{2}$  and, taking (6) and (7) into account,  $q_1^* = \frac{\alpha + 2\phi}{3}$ ,  $q_2^* = \frac{\alpha - \phi}{3}$ . Instead, from Remark 1, we know that  $k_1^* = \frac{\alpha + 2\phi}{3}$  when  $\frac{\alpha}{2} \geq \frac{\alpha + 2\phi}{3}$  and  $\frac{\alpha - \delta\phi}{2} \leq \frac{\alpha + 2\phi}{3}$ . While, as shown above, the second inequality holds true for  $\phi \geq \frac{\alpha}{4 + 3\delta}$ , relative to the first inequality we get:

$$\frac{\alpha}{2} \geq \frac{\alpha + 2\phi}{3} \Leftrightarrow \phi \leq \frac{\alpha}{4}.$$

Hence, when  $\frac{\alpha}{4 + 3\delta} \leq \phi \leq \frac{\alpha}{4}$ ,  $k_1^* = \frac{\alpha + 2\phi}{3}$  and, taking (6) into account,  $k_1^* = q_1^* = \frac{\alpha + 2\phi}{3}$  while, by substituting in (7),  $q_2^* = \frac{\alpha - \phi}{3}$ .

Finally, when firing costs are high, that is,  $\phi > \frac{\alpha}{4}$ , taking Remark 1 and above analysis into account, we get that  $k_1^* = \frac{\alpha}{2}$ . Moreover, since with  $\phi > \frac{\alpha}{4}$ , we have  $\frac{\alpha}{3} < \frac{\alpha}{2} < \frac{\alpha + 2\phi}{3}$ , according to (6) we obtain  $q_1^* = k_1^* = \frac{\alpha}{2}$  and, from (7),  $q_2^* = \frac{\alpha}{4}$ . ■

## References

- Bentolila, S. and Bertola, G. (1990). “Firing costs and labour demand: how bad is euro-sclerosis?” *The Review of Economic Studies*, 57 (3), 381-402.
- Cahuc, P. and Postel-Vinay, F. (2002). “Temporary jobs, employment protection and labor market performance”, *Labour Economics*, 9 (1), 63-91.
- Dixit, A. (1980). “The role of investment in entry-deterrence”, *The Economic Journal*, 90 (357), 95-106.
- Grubb, D. and Wells, W. (1993). “Employment regulation and patterns of work in EC countries”, *OECD Economic Studies*, n.21, 1-58.
- Lazear, E. P. (1990). “Job security provisions and employment”, *The Quarterly Journal of Economics*, 105 (3), 699-726.
- Lommerund, K. E. and Straume, O. R. (2012). “Employment Protection Versus Flexicurity: On the Technology Adoption in Unionised Firms”, *The Scandinavian Journal of Economics*, 114 (2), 177-199.
- Majumdar, S. and Saha, B. (1998). “Job security, wage bargaining and duopoly outcomes”, *The Journal of International Trade & Economic Development*, 7 (4), 389-403.
- Nickell, S. and Layard, R. (1999). *Labor market institutions and economic performance*, in *Handbook of Labor Economics*, vol. 3, North Holland-Elsevier, Amsterdam.

- OECD (1997). *Economic performance and the structure of collective bargaining*, in *OECD Employment Outlook 1997*, OECD, Paris.
- OECD (2004). *Employment Outlook 2004*, OECD, Paris.
- Pissarides, C. (2001). "Employment Protection", *Labour Economics*, 8 (2), 131-159.



# Timing and the selection of the bargaining agenda

LUCIANO FANTI AND DOMENICO BUCCELLA

## 1. Introduction

The subjects of the scope of bargaining and the choice of the negotiation agenda are pivotal in labour-management relations and negotiations and, therefore, for the functioning of labour market institutions in the most advanced economies, notably in Europe. In the present work, we focus the analysis on the two most frequently discussed models in the economic literature, and empirically observed in real word bargaining agendas: the Right-to-Manage (RTM) and the Efficient Bargaining (EB) models.

In the RTM model (Nickell and Andrews 1983; Dobson 1997; Naylor 2002; Correa-López and Naylor, 2004), unionised labour and firms bargain over wages; nevertheless, once wages are fixed, the firms have the right to decide the employment level. On the other hand, in the EB model, the firms and unionised labour negotiate over wages and employment either simultaneously (e.g. McDonald and Solow 1981; Brown and Ashenfelter, 1986; Espinosa and Rhee 1989; Bughin 1995, 1996), or in a sequential way (Manning, 1987a,b).

The first study that compares the negotiation agendas and their profitability in the context of unionised industries is Dowrick (1990). That author finds “that profits under the right-to-manage (RTM) model exceed those under efficient bargaining (EB) (Naylor, 2003, p.59). Furthermore, this conventional result with regard to the RTM agenda is valid irrespective of whether simultaneous or sequential EB (SEB) are considered, as clearly expressed in the International Handbook of Trade Unions: “under unionised monopoly, the firm will prefer to keep employment off the bargaining agenda, whatever

the degree of union influence over employment. In other words, the right-to-manage outcome generates higher profits than either the efficient or sequential bargains, for a given level of union influence over the wage” (Naylor, 2003, p.61).

Following Dowrick (1990) and Naylor (2003), the present paper precisely analyses which institution about the scope of bargaining, and therefore the bargaining agenda, may prevail at market equilibrium and whether firms and unions may (endogenously) agree with this institution. In other words, the present paper presents a game-theoretic approach to the selection of the bargaining agenda. The reference framework of the article is a unionized Cournot duopoly with firms producing homogeneous goods. To determine the equilibrium agendas, three different cases related to the timing of the bargaining game are taken into consideration and their Pareto-efficiency properties are discussed. In this respect, the paper provides an exhaustive treatment of the different strategic effects of the bargaining agenda in relation to the different cases of the timing of the bargaining game, so also integrating the previous literature which treated only partially as well as separately such cases.

In the first case (the only one discussed in Naylor, 2003), in the presence of the mixed duopoly, the firm which introduces the EB agenda negotiate wages and employment when the RTM firm decides output; in other word, the EB firm is Stackelberg wage follower. In the second case, the firm with the EB agenda negotiates wages and employment when the RTM firm bargains the wage with its union, that is, the EB firm is Stackelberg output leader.

Finally, in the third case, despite the diverse agendas, both firms negotiate with unions first wages concurrently, and subsequently the EB firm negotiates with its union over employment, while the RTM firm chooses employment unilaterally; in other words, there is no leadership. This bargaining agenda may be practically relevant, as Manning (1987a, p.124) argues, “Commonly, wages are determined in advance of employment. [...] the fact that over 90% of wage settlements in manufacturing are for one year suggests that the timing of wage bargaining is not very sensitive to economic variables, so it may be reasonable to assume that wage determination is prior to employment determination”, because “in the presence of uncertainty, and the impossibility of complete contingent contracting, it may be optimal to agree on a wage and then let the employer vary employment in response to a changing environment. Here, we are altering this story slightly as the variation of employment is the outcome of a bargain. Wages are agreed in advance to take advantage of insurance possibilities. Even if one does not accept the evidence on the sequential nature of collective bargaining, there is stronger evidence that bar-

gaining about different issues is separated and conducted at different levels. Wage negotiations are commonly conducted at a higher level (e.g. enterprise) than negotiations on job-related issues, which are generally negotiated at plant or workplace level.”

The negotiation processes between unions and firms entail interesting problems not only for the bargaining parties, but also for consumers and society. The comprehension of these issues are significant for the design of policies to regulate labour markets for their appropriate functioning, the organisation of productive activities in different sectors of the economy and product markets as well as the assessment of the effects on social welfare. Consequently, governments and policy-makers as well as guarantees authorities such as antitrust bodies are concerned.

The article contributes to the growing literature on unionised oligopoly (e.g., Horn and Wolinsky, 1988; Dowrick 1989; Bughin, 1995; Correa-López and Naylor, 2004; Fanti and Meccheri, 2011, 2013). However, despite the relevance of the topic for the labour economics as well as the industrial organization, the analysis of the bargaining agenda selection and the impact of different labour market institutions on labour and product markets is scanty in the existing literature. The present paper is one of the few that focuses exactly on these issues. Noteworthy exceptions are the works of Petrakis and Vlassis (2000), Kraft (2006), Bughin (1999), Vannini and Bughin (2000), Buccella (2011) and Fanti (2014, 2015).

Petrakis and Vlassis (2000) build a model where the benchmark equilibrium institution is RTM. Then, the authors test whether the RTM agenda lasts all the possible deviations which need both bargaining parties to agree with. They show that an agreement on the EB agenda can never occur. On the other hand, depending on whether the unions' bargaining power is larger or lower than that of firms, an agreement arises on either RTM or a mixed result where, in equilibrium one firm-union pair selects EB while the other pair chooses RTM.

Kraft (2006) mirrors the assumption of Petrakis and Vlassis (2000) that, when one firm commits to EB, then the rival firm always becomes a Stackelberg follower in the output market. Nevertheless, Kraft (2006) argues, in sharp contrast with Petrakis and Vlassis (2000), that EB is the firms' dominant strategy for a large range of the bargaining power parameter.

In an oligopoly context, Vannini and Bughin (2000) concentrates on the firms' decision whether to adopt cost-raising strategies through the recognition of labour unions. The authors show that in precise circumstances (low union power, low product differentiation and centralised bargaining), firms

can prefer an EB arrangement rather than a RTM arrangement, despite higher unit wage costs. However, Kraft (2006) and Vannini and Bughin (2000) do not take into account the union's utilities to determine the equilibrium agenda and, therefore, abstract from the issue of the endogenous agreement on the bargaining scope.

Thus, the approach of those papers with regard to the bargaining scope and agenda is totally different from the one of the present paper. As a consequence, their results are not directly comparable with those of the present work, which follows and complements Buccella (2011) and Fanti (2014, 2015) who revisited such an issue in a standard game-theoretic approach.

Fanti (2014) first compares the exogenously given equilibria, and then investigates which agenda endogenously arises in equilibrium when the bargaining institutions are the monopoly union and EB. The author finds that, when the equilibrium is exogenous, firms prefer EB when the union bargaining power is lower, equals or slightly higher than that of firms while, for a sufficiently high value of the union bargaining power, unions prefer EB. On the other hand, for a medium-high value of the unions' bargaining power, the parties agree on the monopoly union institution. As for the endogenous equilibria, Fanti (2014) shows that the bargaining parties agree on the selection of the EB agenda for arrangement for a wide medium-high value range of the union's power. In addition, Buccella (2011) and Fanti (2014, Appendix) show that, in the mixed case, when the EB firm negotiates wages and employment when the other firm decides the output level (EB firm Stackelberg wage follower), the endogenous choice of the scope of bargaining by firms is the RTM, irrespective of the unions' bargaining power. By contrast, the endogenous choice by unions is the negotiation over both wages and employment.

Fanti (2015) analyses the endogenous selection of the bargaining agenda in a duopoly in which, in the mixed case, the firm with the EB agenda negotiates wages and employment when the RTM firm bargains the wage with its union (EB firm Stackelberg output leader). He shows that only EB can arise as sub-game perfect Nash equilibrium (SPNE) in pure strategies and under the risk-dominance criterion, provided that the union bargaining power is not too high.

In this paper we revisit the issue of the choice of the negotiation agenda in a twofold way: 1) reconstructing the different equilibrium outcomes due to the three different timing of moves which are implicit in the generic case of Efficient Bargaining and showing the cruciality of such timings for the modification of the conventional results; 2) applying a correct game-theoretic approach to the bargaining agenda issue, which may reveal the lack

of robustness of some conventional results based on the comparison of exogenously given agendas. In particular, as regards the second point, we analyse the “endogenous” equilibrium agenda (Right-to-Manage vs Efficient Bargaining) that can arise under three different timings of the bargaining game in a duopoly market<sup>1</sup> in the following threefold sense: i) “endogeneity” “strictu sensu” because we apply a game-theoretic analysis; ii) “endogeneity” due to the analysis of the SPNEs emerging separately for the hypothetical cases in which alternatively either firms or unions are given the non-cooperative decision on which agenda to implement; iii) “endogeneity” in the sense that it is analysed whether and how the different SPNEs emerging when the decisors are different players may be in “agreement” between them, therefore describing a potential situation of elimination of the conflict of interest on the equilibrium agenda. Finally, the welfare properties of all the emerging SPNEs are investigated, showing the cases in which the equilibrium is Pareto-efficient or inefficient with regard to each player, thus revealing also the cases of strict Pareto-superiority or inferiority (for both players) of such an equilibrium.

The remainder of the article is organized as follows. Section 2 describes the basic duopoly model. Section 3 develops the case of the unionisation of the labour market under the two institutions, EB and RTM, and the three timing characterizations. Section 4 derives the sub-game perfect Nash equilibrium outcomes as well as the key propositions with regard to the choice of the preferred agreement by firms and unions when duopoly is the established market structure. Finally, the last section summarises the main results and implications, and propose plausible directions for further research on the topic.

## 2. The basic model

As Bughin (1999), Petrakis and Vlassis (2000), Naylor (2003), and Kraft (2006), just to mention a few, we consider a duopolistic Cournot model. The two firms, labelled 1 and 2, produce homogeneous goods. The linear (inverse) product demand is

<sup>1</sup> Although in a monopolistic market the strategic interactions are absent, the “endogenous” choice of the bargaining agenda may be a relevant issue also in such a market, as recently shown by Fanti and Buccella (2016 a, b).

$$p = a - q_i - q_j, \quad i, j = 1, 2; \quad i \neq j, \quad (1)$$

where  $p$  is the market price (the slope is normalised to the unity),  $q_i$  and  $q_j$  are the output levels of the two firms. Without loss of generality, it is possible to normalize the parameter  $a = 1$ . Labour,  $l$ , represents the unique input factor. A constant returns-to-scale technology characterises the production function - identical for both firms - with constant (marginal) returns to labour, such that

$$q_i = l_i, \quad i = 1, 2, \quad (2)$$

where  $l_i$  represents the labour force employed by firm  $i$ . The  $i$ -th firm faces an average and marginal cost  $w_i \geq 0$  for every unit of output produced, where  $w_i$  is the wage per unit of labour. Therefore, the firm  $i$ 's cost function is linear and described by:

$$C_i(q_i) = w_i l_i = w_i q_i. \quad (3)$$

Thus, the expression for each firm's profits is

$$\Pi_i = (1 - q_i - q_j - w_i) q_i, \quad i, j = 1, 2; \quad i \neq j. \quad (4)$$

### 3. The unionised labour market

Both firms are unionised, and negotiations are conducted at decentralised level. We assume that the members of the unions are large enough to meet the firm's labour demand. We consider the two typical negotiation models of the trade-union economics (Booth, 1995): 1) the Right-to-Manage model (RTM) (e.g. Nickell and Andrews, 1983), in which wages are the outcome of negotiations between firms and unionised labour; however, once wages are set, the firms have the right to set the employment levels; and 2) the efficient bargaining model (EB) which prescribes that the union and the firm bargain over both wages and employment (or, more realistically, hours of work). Wages and employment negotiations may occur either simultaneously (e.g. McDonald and Solow, 1981; Brown and Ashenfelter, 1986) or sequentially (Manning, 1987a,b).

Each firm-specific union has the following utility function:<sup>2</sup>

$$V_i = w_i l_i. \quad (5)$$

We assume that unions are identical. Therefore, by recalling that  $q_i = l_i$ , (5) becomes:

$$V_i = w_i q_i, \quad (6)$$

that is, the unions' objective function is to maximise the total wage bill. Let us begin by illustrating the cases of RTM and EB, both in the simultaneous and sequential configurations.

### 3.1. Right-to-Manage institution

First, we construct a firm-union two-stage game according to the Right-to-Manage (RTM) model. In the first stage, each firm-union bargaining unit simultaneously negotiates over wages (given the output chosen by firms); in the second stage, firms simultaneously choose their output level (given the wages bargained with the unions). We solve for the equilibrium in the usual backward fashion.

In the second stage of the game, firms compete on quantities (the market game). The first-order conditions (FOCs) of (4) lead to the firms' reaction functions

$$q_i(q_j) = \frac{1}{2}(1 - w_i - q_j), \quad i, j = 1, 2; \quad i \neq j. \quad (7)$$

From (7), we obtain firm  $i$ 's output, for given  $w_i, w_j$

$$q_i(w_i, w_j) = \frac{1}{3}(1 - 2w_i + w_j), \quad i, j = 1, 2; \quad i \neq j. \quad (8)$$

<sup>2</sup> This is a specific case of the more general Stone-Geary utility function, i.e., Pencavel (1984, 1985), Dowrick and Spencer (1994):

$$V(w - w^o)^\theta l,$$

where  $w^o$  is the reservation or competitive wage and  $\theta > 0$  measures the union's wage aggressiveness. A value of  $\theta = 1$  gives the rent-maximising case (i.e., the union seeks to maximise the total rent); values of  $\theta < (>)1$  simply that the union is less (more) concerned about wages and more (less) concerned about jobs. Moreover, the unions aim to maximise the wage bill when  $w^o = 0$  and  $\theta = 1$ . However, values of  $\theta < (>)1$  representing the union's preferences may be important also for the bargaining agenda issue, as Fanti and Buccella (2017) show.

In the first stage of the game, under RTM, the firm-union bargaining unit  $i$  selects  $w_i$  to maximise the following generalized Nash product

$$\underbrace{\max}_{w.r.t. w_i} N_i = (V_i)^b (\Pi_i)^{1-b} = (w_i q_i)^b [(1 - q_i - q_j - w_i) q_i]^{1-b}, \quad i, j = 1, 2; \quad i \neq j, \quad (9)$$

where  $b$  represents the bargaining power of the union. Using (8), the maximisation of (9) with respect to  $w_i$  leads to the sub-game perfect best-reply function in wages of the union-firm pair  $i$  under the assumption of a non-cooperative Cournot-Nash equilibrium in the product market:

$$w_i(w_j) = \frac{b}{4}(1 + w_j), \quad i, j = 1, 2; \quad i \neq j. \quad (10)$$

Solving the system of equations in (10), we obtain the sub-game perfect equilibrium wages:<sup>3</sup>

$$w_i = w_j = w^{RTM/RTM} = \frac{b}{4-b}. \quad (11)$$

Further substitution of (11) into (8), (4) and (6) determines, after the usual algebra, the equilibrium values of output, profits and union's utility

$$q_i = q_j = q^{RTM/RTM} = \frac{2(2-b)}{3(4-b)}, \quad (12)$$

$$\Pi_i = \Pi_j = \Pi^{RTM/RTM} = \frac{4(2-b)^2}{[3(4-b)]^2}, \quad (13)$$

$$V_i = V_j = V^{RTM/RTM} = \frac{2b(2-b)}{3(4-b)^2}. \quad (14)$$

### 3.2. Efficient Bargaining institution.

*Case A:* Simultaneous wage and employment negotiations

Under the simultaneous efficient bargaining, the union and the management of the firm negotiate the wages and employment at the same time. Given the assumption of identical unions having symmetric power across bargaining units, we have to maximise the following generalised Nash product,

<sup>3</sup> The apex - e.g. RTM/RTM - denotes the choice of the type of bargaining arrangement by firms  $i$  and  $j$ , respectively.



## Timing and the selection of the bargaining agenda

$$\begin{aligned} \max_{w_i, q_i} \quad & N_i = (V_i)^b (\Pi_i)^{1-b} = (w_i q_i)^b [(1 - q_i - q_j - w_i) q_i]^{1-b}. \end{aligned} \quad (15)$$

From the system of FOCs of the efficient bargaining game between firms and unions, the firms' reaction functions in output as well as the unions' wages functions are

$$q_i(q_j, w_i) = \frac{1}{2-b} (1 - w_i - q_j), \quad i, j = 1, 2; \quad i \neq j. \quad (\text{contract curve}), \quad (16)$$

$$w_i(q_i, q_j) = b(1 - q_i - q_j), \quad i, j = 1, 2; \quad i \neq j. \quad (\text{rent-sharing curve}) \quad (17)$$

From (16), the firm  $i$ 's output level, for given  $w_i, w_j$ , is

$$q_i(w_i, w_j) = \frac{[b(1 - w_i) - (1 + w_j) + 2w_i]}{(3 - b)(1 - b)}, \quad i, j = 1, 2; \quad i \neq j. \quad (18)$$

Substituting (18) into (17), we get

$$w_i(w_j) = \frac{b[(1 - b) + w_j]}{3 - 2b}, \quad i, j = 1, 2; \quad i \neq j, \quad (19)$$

which defines the sub-game perfect best-reply function in wages of the bargaining unit  $i$ . Solving the system in (19), we obtain the sub-game perfect equilibrium wages

$$w_i = w_j = w^{EB/EB} = \frac{b}{3}. \quad (20)$$

Further substitution of (20) into (18), (4) and (6) allows for the derivation of the equilibrium values of output, profits and union's utility

$$q_i = q_j = q^{EB/EB} = \frac{1}{3}, \quad (21)$$

$$\Pi_i = \Pi_j = \Pi^{EB/EB} = \frac{(1 - b)}{9}, \quad (22)$$

$$V_i = V_j = V^{EB/EB} = \frac{b}{9}. \quad (23)$$

### Case B: Sequential wage and employment negotiations

Under the sequential efficient-bargaining the union and the management of the firm negotiate the wages and employment at different stages. In the first stage, each firm-union pair negotiates over wages; in the second stage, each union-firm

pair bargains the employment level (given the previously bargained wages). In other words, the firm-union bargaining unit  $i$  selects in the first stage  $w_i$  and in the second stage  $q_i$ . We solve for the equilibrium in the usual backward fashion. In this case, we have to maximise the following generalised Nash product,

$$\max_{w.r.t. q_i} N_i = (V_i)^b (\Pi_i)^{1-b} = (w_i q_i)^b \left[ (1 - q_i - q_j - w_i) q_i \right]^{1-b}, \quad (24)$$

where  $b$  is the union's bargaining power. In the first stage, from the system of FOCs of the efficient bargaining game between firms and unions, the firms' output reaction functions are:

$$q_i(q_j, w_i) = \frac{1}{2-b} (1 - w_i - q_j), \quad i, j = 1, 2; \quad i \neq j. \quad (25)$$

From (25), we obtain the firm  $i$ 's output level, for given  $w_i, w_j$ :

$$q_i(w_i, w_j) = \frac{[(1-b) + w_j - (2-b)w_i]}{(1-b)(3-b)}, \quad i, j = 1, 2; \quad i \neq j. \quad (26)$$

Making use of (26), in the first stage the firm-union bargaining unit  $i$  negotiates  $w_i$  and, therefore, maximises the following generalised Nash product:

$$\max_{w.r.t. w_i} N_i = (V_i)^b (\Pi_i)^{1-b} = [w_i q_i(w_i, w_j)]^b \left[ (1 - q_i(w_i, w_j) - q_j(w_j, w_i) - w_i) q_i(w_i, w_j) \right]^{1-b}. \quad (27)$$

The FOCs of the efficient bargaining game between firms and unions lead to

$$w_i(w_j) = \frac{b[(1-b) + w_j]}{2(2-b)}, \quad i, j = 1, 2; \quad i \neq j, \quad (28)$$

which define the sub-game perfect best-reply function in wages of union-firm pair  $i$ . Solving the system in (28), we obtain the sub-game perfect equilibrium wages,

$$w_i = w_j = w^{EB/EB} = \frac{b(1-b)}{4-3b}. \quad (29)$$

Further substitution of (29) into (26) allows for the derivation of the output level:

$$q_i = q_j = q^{EB/EB} = \frac{(2-b)^2}{(4-3b)(3-b)}. \quad (30)$$

Finally, by substituting both (29) and (30) into (4) and (6), we obtain the equilibrium profits and union utilities:

$$\Pi_i = \Pi_j = \Pi^{EB/EB} = \frac{(2-b)^4(1-b)}{[(4-3b)(3-b)]^2}, \quad (31)$$

$$V_i = V_j = V^{EB/EB} = \frac{b(2-b)^2(1-b)}{(4-3b)^2(3-b)}. \quad (32)$$

### 3.3. The mixed case: one bargaining unit chooses EB and the other one chooses RTM

Let us consider the case that the bargaining unit  $i$  chooses EB while the rival  $j$  selects RTM. As regards the mixed case, there are three alternatives for how the rules of the game can be specified, two of them relates to the simultaneous negotiations and one to the sequential negotiations over wage and employment levels.

Simultaneous EB, mixed case 1 (Bughin, 1999; Buccella, 2011; Fanti, 2015): in the second stage, firm  $i$  and union  $i$  bargain over wage and employment taking as given the negotiated wage at  $j$  and firm's  $j$  output. Firm  $j$  chooses employment for given wage and output of the bargaining unit  $i$  and its own wage at  $j$ . In the first stage, the firm and the union  $j$  bargain over the wage taking as given the solutions of the production stage for wage and output in  $i$  and output in  $j$ . In this case, firm  $i$  acts as Stackelberg wage follower,  $i, j = 1, 2, i \neq j$ .

Simultaneous EB, mixed case 2 (Fanti, 2015): in the second stage, firm  $j$  selects its optimal quantity, taking the wage and employment level in  $i$  as given. In the first stage, firm  $i$  and union  $i$  bargain over wage and employment taking as given the negotiated wage in  $j$  and firm's  $j$  reaction function to firm  $i$  output decision in the production stage; firm  $j$  and union  $j$  bargain over the wage taking into consideration its own reaction function to firm  $i$  output choice in the subsequent production stage, and taking as given the wage and employment level in  $i$ . In this case, firm  $i$  is Stackelberg output leader,  $i, j = 1, 2, i \neq j$ .

Mixed sequential EB: in the second stage, firm  $j$  selects its optimal quantity, taking its own wage level and the wage level in  $i$  as given, while firm and union  $i$  bargain efficiently over employment. In the first stage, the firm and the union in  $j$  bargain the wage level taking as given its own reaction function and the negotiated employment in the firm  $i$ ; the firm and the union in  $i$  negotiate efficiently the wage level, taking as given the reaction function of firm  $j$  to the negotiated employment in the firm  $i$ , and its own efficiently negotiated employment,  $i, j = 1, 2, i \neq j$ .

In the following, we examine all the three cases.

*Simultaneous EB, mixed case 1*

The equilibrium outcomes under the mixed oligopoly with the game specification rules in case 1 are here derived. Assume that the union-firm pair  $i$  bargains under EB, while the union-firm pair  $j$  negotiates under RTM. In the second stage, the union-firm pair  $i$  negotiates over wage and employment and, therefore, maximises (15). Solving for  $q_i$  the system of the FOCs in (16) and (17), we derive

$$q_i = \frac{(1 - q_j)}{2}, \quad (33)$$

which represents firm  $i$  output as function of the rival firm's output. On the other hand, still at the second stage, the firm  $j$  chooses the employment level and, thus, from (4), the FOCs for firm  $j$  determine the reaction function in (7). Solving the system composed by (33) and (7) for the firms' quantities, we obtain

$$q_i = \frac{(1 + w_j)}{3}, \quad (34)$$

the firm's  $i$  output as function of the rival firm's wage rate and

$$q_j = \frac{(1 - 2w_j)}{3}, \quad (35)$$

the firm's  $j$  optimal response as function of  $w_j$ .

In the first stage of the game, under RTM, only the bargaining unit  $j$  moves and sets  $w_j$  to maximise (9), taking as given (34) and (35). The FOCs lead to

$$w_j^{EB/RTM} = \frac{b}{4}, \quad (36)$$

the optimal wage of the bargaining unit  $j$  under RTM. Further substitution of (36) into (34) allows for the derivation of

$$q_i^{EB/RTM} = \frac{4 + b}{12}, \quad (37)$$

the firm's  $i$  equilibrium output. Inserting back (36) into (35), we obtain

$$q_j^{EB/RTM} = \frac{2 - b}{6}, \quad (38)$$

the firm's  $j$  output level in equilibrium. Finally, inserting (37) and (38) into (17), the bargained wage in firm  $i$  is

$$w_i^{EB/RTM} = \frac{b(4+b)}{12}, \quad (39)$$

the optimal bargaining unit  $i$  wage under EB. Direct substitutions of the equilibrium quantities and wages into (4)-(6) allow to compute the equilibrium profits and union utilities

$$\Pi_i^{EB/RTM} = \frac{(1-b)(b+4)^2}{144}; \Pi_j^{EB/RTM} = \frac{(2-b)^2}{36}; \quad (40)$$

$$V_i^{EB/RTM} = \frac{b(4+b)^2}{144}; V_j^{EB/RTM} = \frac{b(2-b)}{24}. \quad (41)$$

### *Simultaneous EB, mixed case 2*

The equilibrium outcomes of the mixed duopoly with the simultaneous EB bargain as in case 2 are here derived. Assume again that firm  $i$  negotiates with its union under EB, while firm  $j$  negotiates under RTM. Therefore, at the second stage, the firm  $j$  maximises equation (4) and the first order condition determines the best response function, identical to (7). On the other hand, at the first stage, the bargaining unit  $i$ , making use of (1), (4) and (7), maximises (9) with respect to wage and employment:

$$\max_{w_i, q_i} N_i = (V_i)^b (\Pi_i)^{1-b} = (w_i q_i)^b \left\{ \left[ 1 - q_i - \left( \frac{1 - q_i - w_j}{2} \right) - w_i \right] q_i \right\}^{1-b}, \quad i, j = 1, 2, \quad i \neq j. \quad (42)$$

The FOCs lead to the following expressions

$$q_i = \frac{1 + w_j - 2w_i}{2 - b}, \quad (\text{contract curve}) \quad (43)$$

$$w_i = \frac{b}{2}(1 - q_i + w_j). \quad (\text{rent sharing curve}) \quad (44)$$

Solving the system of equations (43)-(44) for  $w_i$ , we obtain

$$w_i = \frac{b(1 + w_j)}{4}, \quad (45)$$

and consequently

$$q_i = \frac{1 + w_j}{2}, \quad (46)$$

which represent the firm-union pair  $i$ 's output and wage as function of the rival wage level. Taking into account of the own optimal response in terms of

output given by (7), at the first stage the bargaining problem of the firm-union pair  $j$  under RTM is to set  $w_j$  to maximise

$$\max_{w.r.t. w_j} N_j = (V_j)^b (\Pi_j)^{1-b} = \left[ w_j \left( \frac{1 - q_i - w_j}{2} \right) \right]^b \left[ \left( \frac{1 - q_i - w_j}{2} \right)^2 \right]^{1-b}. \quad (47)$$

The FOC leads to

$$w_j = \frac{b(1 - q_i)}{2}, \quad (48)$$

which is the bargained wage in unit  $j$  under RTM as function of the output in unit  $i$ . Therefore, the sub-game equilibrium at the first stage is found by solving the system of the FOCs (48), (46) and (45). Substitution of (46) into (48) allows us for the derivation of the equilibrium bargained wage in unit  $j$  under RTM

$$w_j^{EB/RTM} = \frac{b}{4 + b}. \quad (49)$$

Further substitution of (49) into (45) and (46) leads us to the firm's  $i$  wage and output level in equilibrium, which are given by:

$$w_i^{EB/RTM} = \frac{b(2 + b)}{2(4 + b)}, \quad (50)$$

$$q_i^{EB/RTM} = \frac{2 + b}{4 + b}. \quad (51)$$

Further substitutions of (49) and (51) into the counterpart for  $j$  of (7) lead us to the following firm's  $j$  equilibrium output

$$q_j^{EB/RTM} = \frac{2 - b}{2(4 + b)}. \quad (52)$$

Simple substitutions of (49)-(52) into (4) and (6) allow deriving the equilibrium profits and union utilities

$$\Pi_i^{EB/RTM} = \frac{(1 - b)(2 + b)^2}{2(4 + b)^2}; \quad \Pi_j^{EB/RTM} = \frac{(2 - b)^2}{[2(4 + b)]^2}; \quad (53)$$

$$V_i^{EB/RTM} = \frac{b(2 + b)^2}{2(4 + b)^2}; \quad V_j^{EB/RTM} = \frac{b(2 - b)}{2(4 + b)^2}. \quad (54)$$

#### Mixed sequential EB

Let the firm-union pair  $i$  ( $j$ ) choose EB (RTM). In the second stage, taken as given  $w_i$  and  $w_j$ , the firm-union  $i$  bargains efficiently over employment, while firm  $j$  selects its optimal employment level. Therefore, in the second stage, the

firm  $j$ 's reaction function is (7), while the firm-union pair  $i$ 's reaction function is the following

$$q_i = \frac{1 - w_i - q_j}{2 - b}. \quad (55)$$

Solving the system represented by (7) and (55) we obtain the employment of firms  $i$  and  $j$ , respectively, as a function of  $w_i$  and  $w_j$ :

$$q_i(w_i, w_j) = \frac{1 - 2w_i + w_j}{3 - 2b}, \quad (56)$$

$$q_j(w_i, w_j) = \frac{[1 - b + w_i - w_j(2 - b)]}{3 - 2b}. \quad (57)$$

Further substitutions of (56) and (57) lead to the following Nash Products

$$\max_{w_i, w_j} N_i = (V_i)^b (\Pi_i)^{1-b} = \left[ w_i \left( \frac{1 - 2w_i + w_j}{3 - 2b} \right) \right]^b \left[ (1 - b) \left( \frac{1 - 2w_i + w_j}{3 - 2b} \right)^2 \right]^{1-b}; \quad (58)$$

$$\max_{w_i, w_j} N_j = (V_j)^b (\Pi_j)^{1-b} = \left[ w_j \left( \frac{[1 - b + w_i - w_j(2 - b)]}{3 - 2b} \right) \right]^b \left[ \left( \frac{[1 + w_i - 2w_j - b(1 - w_j)]}{3 - 2b} \right)^2 \right]^{1-b}. \quad (59)$$

Subsequent maximisation of the Nash Products with respect to  $w_i$  and  $w_j$ , leads us to (45) for the bargaining unit  $i$  and

$$w_j = \frac{b(1 + w_i - b)}{2(2 - b)}, \quad (60)$$

for the bargaining unit  $j$ . Solving the system of equations (45) and (60) for  $w_i$  and  $w_j$ , we obtain the equilibrium wages in the mixed case with sequential efficient bargaining:

$$w_i^{EB/RTM} = \frac{b[(2 - b)(2 + b) - b]}{[(4 - b)(4 + b) - 8b]}, \quad (61)$$

$$w_j^{EB/RTM} = \frac{b(4 - 3b)}{[(4 - b)(4 + b) - 8b]}. \quad (62)$$

Further substitutions of (61) and (62) into (56) and (57) allow us for the derivation of the equilibrium quantities

$$q_i^{EB/RTM} = \frac{2(2 - b)[(2 - b)(2 + b) - b]}{(3 - 2b)[(4 - b)(4 + b) - 8b]}, \quad (63)$$

$$q_j^{EB/RTM} = \frac{(2-b)^2(4-3b)}{(3-2b)[(4-b)(4+b)-8b]} \quad (64)$$

Subsequent substitutions of (63)-(64) into (4) and (6) allow for the derivation of the equilibrium values of the profits and union utilities

$$\Pi_i^{EB/RTM} = \frac{4(1-b)(2-b)^2[(2-b)(2+b)-b]^2}{(3-2b)^2[(4-b)(4+b)-8b]^2}; \Pi_j^{EB/RTM} = \frac{(4-3b)^2(2-b)^4}{(3-2b)^2[(4-b)(4+b)-8b]^2}; \quad (65)$$

$$V_i^{EB/RTM} = \frac{2b(2-b)[(2-b)(2+b)-b]^2}{(3-2b)[(4-b)(4+b)-8b]^2}; V_j^{EB/RTM} = \frac{b(4-3b)^2(2-b)^2}{(3-2b)[(4-b)(4+b)-8b]^2}. \quad (66)$$

## 4. The equilibrium agendas

Now we are in a position to conduct the following analysis. We compare firm's profits and unions' welfare under the three labour market institutions, and we investigate which bargaining agenda will *endogenously* emerge in the sub-game perfect Nash equilibrium (SPNE) for both firms and unions.

### 4.1. The game with simultaneous EB and mixed case 1

Let us first consider the game with simultaneous EB and the characterization of the mixed case 1. Given the payoffs in (13), (14), (22), (23), (40) and (41), we obtain the subsequent Tables 1 and 2 which summarise profits and union's utilities, respectively, in the four strategic situations. An analytical inspection allows us to derive the following results.

**Proposition 1** (Buccella, 2011; Fanti, 2014) *In a unionized Cournot duopoly with simultaneous EB and the characterisation of the mixed case 1, RTM is the dominant strategy for firms; this is in explicit conflict with the unions' interests because EB is their dominant strategy as regards the bargaining agenda. Therefore, under the endogenous determination of the SPNE, no agreement on the bargaining agenda may occur between unions and firms.*

Proof: From direct payoff comparison in Tables 1 (for firms) and 2 (for unions).

The rationale for this result is as follows. In the mixed case, the firm with the EB agenda has higher negotiated wages and produces larger output when the competitor is a firm with a RTM agenda, and profits for the RTM firm are higher than those of the EB firm.



**Table 1: Profit matrix, simultaneous EB, mixed case 1**

Profit Matrix		Firm 2	
		RTM	EB
Firm 1	RTM	$\frac{4(2-b)^2}{[3(4-b)]^2}; \frac{4(2-b)^2}{[3(4-b)]^2}$	$\frac{(2-b)^2}{36}; \frac{(1-b)(b+4)^2}{144}$
	EB	$\frac{(1-b)(b+4)^2}{144}; \frac{(2-b)^2}{36}$	$\frac{(1-b)}{9}; \frac{(1-b)}{9}$

**Table 2: Union rent matrix, simultaneous EB, mixed case 1**

Union rent Matrix		Union 2	
		RTM	EB
Union 1	RTM	$\frac{2b(2-b)}{3(4-b)^2}; \frac{2b(2-b)}{3(4-b)^2}$	$\frac{b(2-b)}{24}; \frac{b(4+b)^2}{144}$
	EB	$\frac{b(4+b)^2}{144}; \frac{b(2-b)}{24}$	$\frac{b}{9}; \frac{b}{9}$

The reason is that, because of the common price for the products and lower wages, the margins for the RTM firm are higher than the EB firm. Moreover, lower output and wages for the RTM firm ensure lower total costs. On the other hand, from the unions' point of view, EB is the dominant strategy because of higher negotiated wages and employment levels.

#### 4.2. The game with simultaneous EB and mixed case 2

Let us consider the game with simultaneous EB and the characterization of the mixed case 2. Given the payoffs in (13), (14), (22), (23), (53) and (54), we obtain the following Tables 3 and 4 which sum up profits and union's utilities, respectively, in the four strategic situations.

A closer analytical inspection reveals the following results.

**Proposition 2** (Fanti, 2015) *In a unionized Cournot duopoly with simultaneous EB and the characterisation of the mixed case 2: i) as regards firms, it holds that: when  $1 > b \geq .833$ , the unique SPNE is RTM/RTM; when  $.833 > b \geq .42$ , there exist two SPN equilibria, RTM/RTM and EB/EB; when  $.42 > b \geq 0$  the unique SPNE is EB/EB; ii) as regards unions, there exists a unique SPNE, given by EB/EB, irrespective of the values of  $b$ , that is, EB is the dominant strategy for unions.*

Proof: From direct payoff comparison in Tables 3 (for firms) and 4 (for unions); see also Fanti (2015).

<b>Table 3: Profit matrix, simultaneous EB, mixed case 2</b>			
Profit Matrix		Firm 2	
		RTM	EB
Firm 1	RTM	$\frac{(2-b)^2}{[2(4+b)]^2}; \frac{(1-b)(2+b)^2}{2(4+b)^2}$	$\frac{4(2-b)^2}{[3(4-b)]^2}; \frac{4(2-b)^2}{[3(4-b)]^2}$
	EB	$\frac{(1-b)(2+b)^2}{2(4+b)^2}; \frac{(2-b)^2}{[2(4+b)]^2}$	$\frac{(1-b)}{9}; \frac{(1-b)}{9}$

<b>Table 4: Union rent matrix, simultaneous EB, mixed case 2</b>			
Union rent Matrix		Union 2	
		RTM	EB
Union 1	RTM	$\frac{2b(2-b)}{3(4-b)^2}; \frac{2b(2-b)}{3(4-b)^2}$	$\frac{b(2-b)}{2(4+b)^2}; \frac{b(2+b)^2}{2(4+b)^2}$
	EB	$\frac{b(2+b)^2}{2(4+b)^2}; \frac{b(2-b)}{2(4+b)^2}$	$\frac{b}{9}; \frac{b}{9}$

The results in Proposition 2 are in contrast to those in Proposition 1. The intuition behind this finding is as follows. It can be easily verified that the firm which selects EB with the mixed case 2 produces more output than in the identical situation in case 1 by paying lower wages. Therefore, the firm improves its margins and profitability.

On the other hand, for  $b \in [.42, .833]$ , the firm which adopts EB in the mixed case experiences an increase in its wage to a level such that  $\Pi^{RTM/RTM} \geq \Pi^{EB/RTM}$ , while  $\Pi^{EB/EB} > \Pi^{RTM/EB}$ . As a consequence, the game shows multiple symmetric equilibria. Finally, for  $b \in [.833, 1)$ , the firms' payoff structure is such that  $\Pi^{RTM/RTM} \geq \Pi^{EB/RTM}$  and  $\Pi^{RTM/EB} \geq \Pi^{EB/EB}$ : the RTM is the firms' dominant strategy. With regards to unions, direct payoff comparison shows that, once again, EB is their dominant strategy. Thus, from Proposition 2 and the above analysis, we can derive the following result.

**Proposition 3** (Fanti, 2015) *The unambiguous agreement between unions and firms with regard to the scope of bargaining is on the EB agenda, provided that  $.42 > b \geq 0$ . When  $.833 > b \geq .42$ , there are multiple equilibria: in this case, the game shows the structure of a coordination game for firms.*

Focusing on the case of multiple equilibria for firms (i.e. the situation  $.833 > b \geq .42$ ), we note that there is also one mixed-strategy equilibrium (see Fanti, 2015). Moreover, considering only, for simplicity, the pure-strategy equilibrium selection problem, in principle, it could also be possible the selection

of the EB agenda where there would be an agreement of interest between firms and unions. In particular, in the presence of multiple equilibria, there are refinement techniques for their selection. Two of the most common criteria are the payoff-dominance criterium and the risk- dominance criterium.

A Nash equilibrium is Payoff-dominant if it is Pareto superior to all other Nash equilibria in the game. On the other hand, a Nash equilibrium is considered Risk-dominant if it is less risky (e.g., Harsanyi and Selten, 1988). Two analogous definitions of the Risk-dominance in a 2x2 symmetric coordination game are the following: 1) a strategy is Risk-dominant if it is a best response to a 50-50 randomization by the other player; and 2) a strategy is Risk-dominant if it has the smallest probability in the mixed strategy Nash equilibrium.

Therefore, focusing on the analysis of the selection between multiple equilibria, it is shown the following:

**Proposition 4** (Fanti, 2015) *According to the payoff-dominance criterium, the RTM equilibrium Pareto-dominates the EB equilibrium (that is the firms would coordinate on the RTM equilibrium). By contrast, by using the criterium of the risk-dominance, the strategic pair (EB, EB) risk-dominates the strategic pair (RTM, RTM) if  $.66 > b \geq 0$ .*

Proof: The proof easily follows from the payoffs in Table 3.

Therefore, if we assume that the firms are interested in minimising the risk of coordination failure in the selection of the labour market institution, the unique outcome is the EB agenda for  $b \leq 2/3$ . To sum up, we have shown that the EB agenda is the arrangement on which each firm/union pair may agree with, provided that the union's bargaining power is either sufficiently low ( $b \leq .42$  under pure strategies) or not too high ( $b \leq 2/3$  under risk-dominance criterium).

While the above considerations concern the determination of the SPNE's in the simultaneous EB under the two different timings, as regards the welfare properties of such SPNEs, the following Lemma holds.

**Lemma 1** *When the game considers the simultaneous EB, firms always prefer RTM while unions prefer EB, irrespective of the timing specification.*

Proof: From direct comparison of the RTM/RTM and EB/EB payoffs in Tables 3 (for firms) and 4 (for unions).

Lemma 1 reveals that i) in the case of Proposition 1, the dominant equilibrium strategies are Pareto-efficient for each player; however, there is an ineradicable conflict of interest on the agenda's choice between players; ii) in the case of Proposition 2, the SPNE which would be emerged irrespective of whether the decisors on agenda are the firms or unions (that is, EB for  $.66 > b \geq 0$  in

the case of the risk-dominance criterium) is Pareto-inefficient for firms and Pareto-efficient for unions (while, for  $1 > b \geq .66$ , still in the case of risk-dominance criterium, the SPNE agendas are different for players with a conflict of interest on the agenda's choice but Pareto-efficient for each of them).

### 4.3. The game with sequential EB

Let us finally consider the game characterised by the sequential EB. Given the payoffs in (13), (14), (31), (32), (65) and (66), we obtain the following Tables 5 and 6 which summarise profits and union's utilities, respectively, in the four strategic situations.

Table 5: Profit matrix, sequential EB			
Profit Matrix		Firm 2	
		RTM	EB
Firm 1	RTM	$\frac{4(2-b)^2}{[3(4-b)]^2}; \frac{4(2-b)^2}{[3(4-b)]^2}$	$\frac{(4-3b)^2(2-b)^4}{(3-2b)^2[(4-b)(4+b)-8b]^2}; \frac{4(1-b)(2-b)^2[(2-b)(2+b)-b]^2}{(3-2b)^2[(4-b)(4+b)-8b]^2}$
	EB	$\frac{4(1-b)(2-b)^2[(2-b)(2+b)-b]^2}{(3-2b)^2[(4-b)(4+b)-8b]^2}; \frac{(4-3b)^2(2-b)^4}{(3-2b)^2[(4-b)(4+b)-8b]^2}$	$\frac{(2-b)^4(1-b)}{[(4-3b)(3-b)]^2}; \frac{(2-b)^4(1-b)}{[(4-3b)(3-b)]^2}$

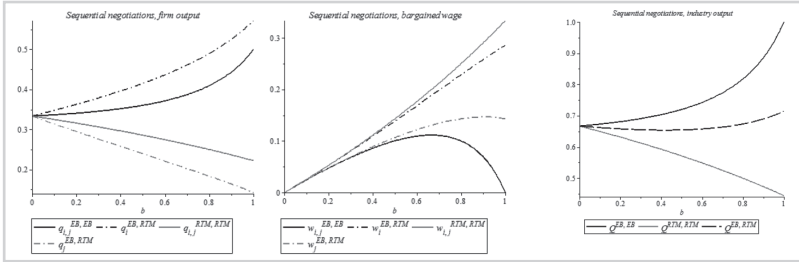
Table 6: Union rent matrix, sequential EB			
Union rent Matrix		Union 2	
		RTM	EB
Union 1	RTM	$\frac{2b(2-b)}{3(4-b)^2}; \frac{2b(2-b)}{3(4-b)^2}$	$\frac{b(4-3b)^2(2-b)^2}{(3-2b)[(4-b)(4+b)-8b]^2}; \frac{2b(2-b)[(2-b)(2+b)-b]^2}{(3-2b)[(4-b)(4+b)-8b]^2}$
	EB	$\frac{2b(2-b)[(2-b)(2+b)-b]^2}{(3-2b)[(4-b)(4+b)-8b]^2}; \frac{b(4-3b)^2(2-b)^2}{(3-2b)[(4-b)(4+b)-8b]^2}$	$\frac{b(2-b)^2(1-b)}{(4-3b)^2(3-b)}; \frac{b(2-b)^2(1-b)}{(4-3b)^2(3-b)}$

An analytical inspection leads to the following results.

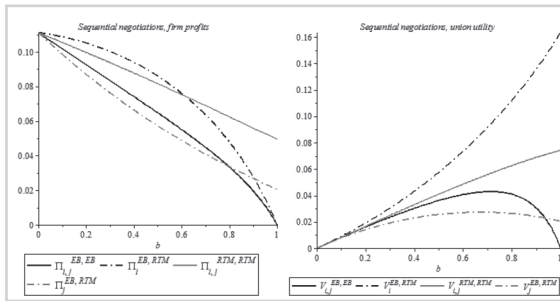
**Proposition 5** In a unionized Cournot duopoly with sequential EB: a) as regards firms, it holds that: when  $1 > b \geq .79$ , the unique SPNE is RTM/RTM; when  $.79 > b \geq .61$ , there exist two SPN equilibria, RTM/RTM and EB/EB; when  $.61 > b \geq 0$  the unique SPNE is EB/EB;

b) as regards unions, there exists a unique SPNE, given by EB/EB, for  $.93 > b \geq 0$ , while for  $1 > b \geq .93$ , there exist two SPNE equilibria, namely RTM/EB and EB/RTM.

The results in Proposition 5 as regards firms are similar to those in Proposition 2, though valid for different ranges of the bargaining power parameter. In fact, the firm which selects EB in the mixed duopoly with sequential bargaining produces more output than the rival with RTM, by paying a higher wage, as Figure 1 shows.



**Figure 1: Firm output (left box), wage (center box) and industry output (right box) in sequential negotiations**



**Figure 2: Firm profits (left box) and union utility (right box) in sequential negotiations**

Nonetheless, for a sufficiently low value of the union bargaining power, the total output is not so large to drive down the product price, and the negotiated wages are low enough to ensure that  $\Pi^{EB/RTM} > \Pi^{RTM/RTM}$ , as Figure 2 depicts. On the other hand, for  $b \in [.61, .79)$ , the firm which adopts EB in the mixed case experiences an increase in its bargained wage to a level such that  $\Pi^{RTM/RTM} \geq \Pi^{EB/RTM}$ , while  $\Pi^{EB/EB} > \Pi^{RTM/EB}$ , leading to multiple equilibria.

Finally, for a sufficiently high value of the union bargaining power ( $b \geq .79$ ), the total output in the mixed duopoly is large enough to reduce the price at a level that RTM is more profitable. In fact, in the mixed case, the common price for the goods lowers while the wage level paid by the EB firm is extremely high with respect to that of the RTM rival.

When both firms negotiate with EB, the total output is even larger than the mixed case and the effect is a price level lower than the mixed case. Consequently we have that  $\Pi^{\text{RTM}/\text{EB}} > \Pi^{\text{EB}/\text{EB}}$  and  $\Pi^{\text{RTM}/\text{RTM}} \geq \Pi^{\text{EB}/\text{RTM}}$ : RTM is the most preferred agenda for firms.

The situation for the unions is not only quantitatively (as for firms, as seen above) but also qualitatively slightly different under sequential (rather than simultaneous) EB. In fact, for  $b \in (0, .93)$ , EB is always the most preferred agenda.

However, for  $b \geq .93$ , every union has an incentive to deviate towards the RTM agenda, provided that the rival bargaining unit negotiates with EB, and vice versa. Therefore, two asymmetric equilibria arise. The result that each union finds convenient to switch to RTM when its power is very high is due to the fact that under sequential EB wages and union's utilities tend to reduce strongly for too high union's power, as clearly displayed in Figs. 1 and 2. This is a feature of the sequential EB game, because a rise in union's power in the stage of employment determination might make the union even worse off,<sup>4</sup> in that such a rise alters the level of employment for any given level of the wage, and in principle "we can not say whether the bargaining set in the first stage alters in a way favourable to the union or the employer. Consequently, we cannot say how the wage will change." (Manning, 1987a, 130).<sup>5</sup> Therefore, we may argue that in the present Cournot model an extremely high union's power alters the level of employment to the extent that in the first stage wages become relatively lower.

Let us focus on the case of multiple equilibria. As regards firms, for  $.79 > b \geq .61$ , the game is a standard coordination game. In addition to the two pure-strategy equilibria, there is also one mixed-strategy equilibrium. We consider again only the pure-strategy equilibrium selection problem, according to the Pareto-dominance and Risk-dominance criteria.

Concerning the Pareto-dominance criterium, the firms would coordinate on the RTM equilibrium: from Table 5, it can be easily checked that RTM dominates EB. On the other hand, according to the risk-dominance criterium, if

<sup>4</sup> This may never occur in a one stage bargaining game such as the simultaneous EB, where an increase in the union's bargaining power always makes unions better off.

<sup>5</sup> As noted by Manning (1987a, 130, where  $q$  is the union's power in the stage of the employment bargaining) "one implication of this result is that it is conceivable that both union and employer could agree to raise or lower  $q$  or, rather more strangely, that the union might want to reduce its influence over employment while the employer might want to increase it."

we assume again that each firm assigns a probability of 1/2 to the RTM and EB each, then the pair (EB,EB) risk-dominates the pair (RTM,RTM) if the expected payoff of EB exceeds the expected payoff of RTM, that is, if  $\Pi_1^{EB/RTM} + \Pi^{EB/EB} \geq \Pi^{RTM/RTM} + \Pi_2^{EB/RTM}$ . Simple algebraic computation shows that the pair (EB, EB) risk-dominates the pair (RTM, RTM) if  $.68 > b \geq 0$ .

On the other hand, concerning unions, for  $1 > b \geq .93$ , the game presents the characteristics of an anti-coordination game where neither the Pareto-dominance nor the risk-dominance criterium allow for the selection of an equilibrium. These findings can be summed up in the next Proposition.

**Proposition 6** *The unambiguous agreement between unions and firms with regard to the scope of bargaining is on the EB agenda, provided that  $.61 > b \geq 0$ . Moreover, the strategic pair (EB, EB) risk-dominates the strategic pair (RTM, RTM) if  $.68 > b \geq 0$ .*

The welfare properties of the SPNEs described in Propositions 5 and 6 are resumed by the following Lemma.

**Lemma 2** *When the game considers the sequential EB, firms always prefer RTM. On the other hand, also unions always prefer RTM. As a consequence, the endogenous agreement between unions and firms on the EB agenda if  $.68 > b \geq 0$  (with the risk-dominance criterium) is Pareto-inefficient for firms and unions and, then, Pareto-inferior. Conversely, in the case in which firms decide on the agenda when  $1 > b \geq .68$  (still in the case of risk-dominance criterium), the emerging RTM as SPNE is Pareto-efficient for firms and unions and, then, Pareto-superior.*

Proof: From direct comparison of the RTM/RTM and EB/EB payoffs in Table 3 and Fig.2 (left-box) for firms and Table 4 and Fig. 2 (right-box) for unions.

Thus, if we assume again that the firms are interested in minimising the risk of coordination failure in the selection of the labour market institution, the unique outcome is the EB agenda for  $b \leq .68$ , a value slightly higher than the mixed case 2. To summarise, also in the presence of sequential negotiations, the EB agenda is the arrangement on which each bargaining pair may agree with, provided that the union's bargaining power is not too high ( $b \leq .61$  under pure strategies and  $b \leq .68$  under the risk-dominance criterium).

The findings of this section have shown that, in the negotiations process between firms and unions, the specification of the timing is crucial to determine the endogenous equilibrium agenda. We have shown that only EB can arise as the endogenous equilibrium arrangement on which each bargaining unit can agree, provided that: 1) the bargaining units concurrently negotiate

wages in the mixed duopoly as in the simultaneous EB, case 2, and sequential EB. In fact, in those cases, as Figure 3 shows, the bargained wage of the EB firm is low enough to make this agenda sufficiently profitable; and 2) the union's power is intermediate under pure strategies ( $b \leq .42$  in the case of simultaneous EB, mixed case 2;  $b \leq .61$  in the case of sequential EB) or not too high under the risk-dominance criterium ( $b \leq .66$  simultaneous EB, mixed case 2;  $b \leq .68$ , sequential EB).

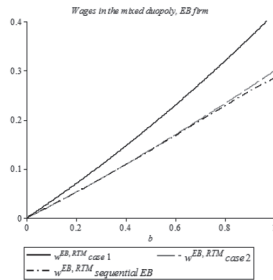


Figure 3: EB firm, negotiated wages in the three timing cases in the mixed duopoly

#### 4.4. Discussion of the outcomes of the three timing cases of the bargaining agenda game

The results of this section have shown that the endogenous choice of the union-firm bargaining agenda crucially depends on the specification of the timing. In fact, when the timing is specified as in the mixed case 1, where the EB firm is Stackelberg wage follower, no endogenous agreement between the bargaining parties arises (Buccella, 2011; Fanti, 2014). However, in the mixed case 2, with the EB firm Stackelberg output leader (Fanti, 2015) and in the case of mixed sequential bargaining, with no leadership, EB is the sole institution that can emerge as the endogenous agenda. These findings reveal that EB emerges as equilibrium in the two cases when wages are concurrently negotiated in the rival units, suggesting that in the mixed case 1, having knowledge of the wage level at the RTM rival firm, the EB unit does not reduce enough the negotiated wage to a level to improve its profitability. Moreover, a new result is that, in sequential bargaining, EB appears as the equilibrium agenda for a range of the unions' power larger than in the mixed case 2 both under pure strategies and the risk-dominance criterium.



New results under sequential bargaining are also that unions always prefer RTM and the mixed agenda is the equilibrium for unions when their power is remarkably high, but, mainly, that when the union's power is large, RTM is a Pareto-superior choice which is robust to an endogenous choice by firms as well as unions.

## 5. Conclusions

The present work has widely revisited a central aspect of labour union(s)-firm management relations: the bargaining scope and the choice of the agenda in negotiations.

The paper has addressed the analysis to the endogenous equilibrium agenda (Right-to-Manage vs Efficient Bargaining) that can arise under three different timings of the bargaining game in a duopoly market. In the first case, when there is a mixed duopoly, the firm which introduces the EB agenda is Stackelberg wage follower; in the second case, the firm with the EB agenda is Stackelberg output leader; in the third case, despite the different agenda, both firms negotiate with unions first wages concurrently, and subsequently the EB firm negotiates with its union over employment.

The key results we have identified in the analysis are the following. The timing of the negotiation process plays a crucial role in determining the endogenous equilibrium agenda between firms and unions. We have shown that, when in the mixed duopoly the EB firm is Stackelberg wage follower, no endogenous equilibrium bargaining agenda arises in union-firm negotiations. On the other hand, when in the mixed duopoly the EB firm is Stackelberg output leader or negotiate sequentially over wages and employment, we have found that only the Efficient Bargaining can endogenously emerge as the equilibrium agenda, in pure strategies and under the risk-dominance criterium, provided that the union's bargaining power is not extremely high. In the case of sequential EB we have shown that also unions would prefer RTM (in contrast to the cases of simultaneous EB).

In general, there is always a conflict of interest in terms of welfare regardless of which is the SPNE as well as the agenda's decisor but the case when union's power is large under sequential EB. In such a case, we have shown that RTM is a robust and Pareto-superior choice.

The results of our model, which extend and modify the established findings (i.e. Naylor, 2003), presents several limitations. We have restricted the analysis to specific forms for the demand and cost functions as well as for

the union utility function. Moreover, the findings presented in the paper are related to the assumptions of quantity competition, homogenous product and a standard wage system. A further step would be to extend the research toward studying the introduction, on the one hand as regards the product market side, of price competition and heterogeneous products, and, on the other hand as regards the labour market side, of performance related payment schemes into the negotiations such as the profit sharing and piece rates. Moreover, the robustness of the present results can be verified in an extended game framework where also managerial delegation, R&D investments, externalities in consumption and production, and capacity choices are considered.

## References

- Booth A.L. (1995), *The Economics of the Trade Union*, Cambridge Univ. Press.
- Brown J.N., Ashenfelter, O. (1986), "Testing the efficiency of employment contracts", *Journal of Political Economy*, 94, 41–87.
- Buccella D. (2011), "Corrigendum to "The strategic choice of union oligopoly bargaining agenda" [Int. J. Ind. Organ. 17 (1999) 1029–1040]", *International Journal of Industrial Organization*, 29, 690–693.
- Bughin J. (1995), "Bargaining over employment as a firm strategic choice", *Review of Industrial Organization*, 10, 723–735.
- Bughin J. (1996), "Trade unions and firms' product market power", *The Journal of Industrial Economics*, 44(3), 289–307.
- Bughin J. (1999), "The strategic choice of union–oligopoly bargaining agenda. International", *Journal of Industrial Organization*, 17(7), 1029–1040.
- Dobson P. W. (1997), "Union–firm interaction and the right to manage", *Bulletin of Economic Research*, 49(3), 213–229.
- Dowrick S. (1989), "Union-oligopoly bargaining", *The Economic Journal*, 99, 1123–1142.
- Dowrick S. (1990), "The relative profitability of Nash Bargaining on the labour demand curve or the contract curve", *Economics Letters*, 33(2), 121–125.
- Dowrick S., Spencer B.J. (1994), "Union attitudes to labor-saving innovation: when are unions luddites?", *Journal of Labor Economics*, 12, 316–344.
- Espinosa M., Rhee S. (1989), "Efficient wage bargaining as a repeated game", *Quarterly Journal of Economics*, 104, 565–588.
- Fanti L., (2014), *When do Firms and Unions agree on a Monopoly Union or an Efficient Bargaining Arrangement?*, Discussion Paper n. 181. Department of Economics and Management, University of Pisa.
- Fanti L., (2015), "Union-Firm Bargaining Agenda: Right-to-Manage or Efficient Bargaining?", *Economics Bulletin*, 35(2), 936–948.
- Fanti L., Buccella D. (2016a), "Bargaining agenda and entry in a unionised model with

- network effects”, *Italian Economic Journal*, 2(1), 91-121.
- Fanti L., Buccella D. (2016b), “Manager-Union Bargaining Agenda Under Monopoly and with Network Effects”, *Managerial and Decision Economics*, online first 20 June 2016, doi: 10.1002/mde.2809.
- Fanti L., Buccella D., (2017), “Union-firm bargaining agenda revisited: When unions have distinct preferences”, *Bulletin of Economic Research*, online first 13 February 2017, doi:10.1111/boer.12113.
- Fanti L., Meccheri N., (2011), “The Cournot-Bertrand profit differential in a differentiated duopoly with unions and labour decreasing returns”, *Economics Bulletin*, 31, 233-244.
- Fanti L., Meccheri N. (2013), “Managerial delegation under alternative unionization structures”, *Labour*, 27 (1) 38-57.
- Harsanyi J. C., Selten R. (1988), *A General Theory of Equilibrium Selection in Games*, MIT Press, Cambridge, MA.
- Horn H., Wolinsky A. (1988), “Bilateral monopolies and incentives for merger”, *RAND Journal of Economics*, 19(3), 408-419.
- Kraft K. (2006), “Wage versus Efficient Bargaining in Oligopoly”, *Managerial and Decision Economics*, 27(7), 595-604.
- López M. C., Naylor, R. A. (2004), “The Cournot-Bertrand profit differential: A reversal result in a differentiated duopoly with wage bargaining”, *European Economic Review*, 48, 681-696.
- Manning A. (1987a), “An Integration of Trade Union Models in a Sequential Bargaining Framework”, *The Economic Journal*, 97, 121-139.
- Manning A. (1987b), “Collective Bargaining Institutions and Efficiency”, *European Economic Review*, 31, 168-176.
- McDonald I.M., Solow R.M. (1981), “Wage bargaining and employment”, *American Economic Review*, 71(5), 896-908.
- Naylor R. A. (2002), “Industry profits and competition under bilateral oligopoly”, *Economic Letters*, 77, 169-175.
- Naylor R. A. (2003), *Economic Models of Union Behavior*, In: Addison J. T. - Schnabel, C. (eds): *International Handbook of Trade Unions*, Cheltenham: Edward Elgar, 44-85.
- Nickell S.J., Andrews, M. (1983), “Unions real wages and employment in Britain 1951-1979”, *Oxford Economic Papers*, 35, 183-206, Supplement.
- Pencavel J.H. (1984), “The trade-off between wages and employment in trade union objectives”, *Quarterly Journal of Economics*, 99, 215-231.
- Pencavel, J.H. (1985), “Wages and employment under trade unionism: microeconomic models and macroeconomic applications”, *Scandinavian Journal of Economics*, 87, 197-225.
- Petrakis E., Vlassis M. (2000), “Endogenous Scope of Bargaining in a Union-Oligopoly Model: When Will Firms and Unions Bargain over Employment?”, *Labour Economics*, 7(3), 261-281.
- Vannini S., Bughin J. (2000), “To be (unionized) or not to be? A case for cost-raising strategies under Cournot oligopoly”, *European Economic Review*, 44(9), 1763-1781



# Paolo Sylos Labini vindicated

SIMONE D'ALESSANDRO, NERI SALVADORI AND RODOLFO SIGNORINO

## 1. Introduction

Franco Modigliani presented his 1958 *Journal of Political Economy* paper as a mere review article of Joe Bain's *Barriers to New Competition* (1956) and Paolo Sylos Labini's *Oligopolio e Progresso Tecnico* (1957, 2<sup>nd</sup> Italian edition). With the benefit of hindsight, it is not an overstatement to claim that Modigliani's paper is a path-breaking contribution to the then blossoming literature on entry-preventing models. In particular, Modigliani elaborated a simple formalization and an intuitive graphical representation which, in his view, embodied "the logical essence" of Bain's and Sylos Labini's approach (Modigliani 1958, p. 217), thus making good an effective lacuna into their works. Moreover, Modigliani coined the (un)fortunate phrase 'Sylos Postulate' –the assumption according to which "potential entrants behave as though they expected existing firms to adopt the policy most unfavorable to them, namely, the policy of maintaining output while reducing the price (or accepting reductions) to the extent required to enforce such an output policy" (*ibidem*)– and ascribed it to Bain's and Sylos Labini's credit.

Ever since the 1959 debate in the *Journal of Political Economy* on Modigliani (1958), interpreters have generally focused on the formal properties of the Sylos Labini-Bain-Modigliani (or S-B-M) model, its drawbacks or possible generalizations, its empirical content etc. Thus far, much less attention has been devoted to assess whether such a model is a legitimate rational reconstruction of Sylos Labini's own thought on oligopoly (a notable exception is Rancan 2015). Our answer to this question is in the negative. In our view, the S-B-M model parts company with Sylos Labini's original oligopoly theory on at least two crucial analytical issues. First, Sylos Labini's concept of competition was akin to the Classical notion of competition that is, a theoretical framework in

which firms set prices and commit themselves to sell their products at those prices. By contrast, the concept of competition underlying the S-B-M model is the standard Cournotian one of firms taking decisions about the quantities to bring to the market, the determination of equilibrium prices being left to a Walrasian-like auctioneer. Second, Sylos Labini was strongly committed to the following two views: i) incumbent firms install productive capacity to strategically deter entry and ii) such capacity does not need to be employed in the no-entry scenario. By contrast, in the usual presentation of the Dixit (1980) model the productive capacity installed by incumbents for entry deterrence purpose is never left idle, both in the entry and in the no-entry scenario (but see Bulow et al. 1985).

To substantiate our claims, in this paper we develop a model in which, as in Dixit (1980), there are three stages: in the first stage the incumbent(s) choose(s) the productive capacity to install; in the second stage a potential entrant chooses whether to enter or not and, in the case of entry, chooses the level of its productive capacity; in the third and final stage incumbent(s) and the entrant (if any) compete on prices (*à la Bertrand*). Here is the main analytical difference between our model and Dixit (1980): in the third stage of the latter, as is well-known, firms compete on quantities (*à la Cournot*).

Besides Dixit (1980), the model here presented is closely related to other two contributions widely discussed in recent oligopoly literature. We refer to Kreps and Scheinkman (1983) and Allen et al. (2000).<sup>1</sup> The model developed by Kreps and Scheinkman (1983) is not a model concerning entry and a fortiori entry deterrence: there are two incumbent firms which choose their productive capacities at the same time and no entrant. In our model, on the contrary, there are one or two incumbents and one entrant; moreover, the entrant chooses its productive capacity after the incumbent(s). Yet, in the two-incumbent version of our model the case in which entry is blockaded corresponds exactly to the case analyzed by Kreps and Scheinkman (1983) with the proviso that at the capacities chosen by the two incumbents the potential entrant chooses not to enter.

As concerns Allen et al. (2000), the one-incumbent version of our model is but a special case of their model, since they allow for different marginal costs

<sup>1</sup> In subsections 4.1 and 4.2, to characterize the equilibria in the pure and mixed strategy we also take advantage of a paper by Baik (1995), that analyzes the horizontal merging of price-setting firms by means of a game which has some subgames identical, apart for notation, to the subgame we study in those subsections.

between the incumbent and the entrant whereas we assume that all firms share the same technology. We present this model in Section 4 for two reasons: first, as a benchmark to compare with the two-incumbent version of our model presented in Section 5; second, since we hope that what is lost in generality with respect to Allen et al. (2000) is gained in transparency. The two-incumbent version of our model confirms all the results by Allen et al. (2000) with one exception. We refer to the fact that the parameters may be such that two Nash equilibria exist: one in which entry is deterred and one in which entry is accommodated and, consequently, there is a problem of coordination among the incumbents. This fact never happens in the one-incumbent version of our model (in the same vein see Gilbert and Vives, 1986.)

To sum up, our model has two distinguishing features. First, competition is à la Bertrand rather than à la Cournot. Second, the capacity installed in the first stage of the game by the incumbent(s) for strategic entry deterrence purpose may be left idle in the third stage in the no-entry scenario.

The structure of the paper is the following. Section 2 deals with Modigliani's interpretation of *Oligopolio e Progresso Tecnico*: we contrast the latter with our rational reconstruction of the original ideas by Sylos Labini. Section 3 briefly summarizes the debate concerning the Sylos Postulate and its reinterpretation by Dixit (1980). Section 4 introduces our model in the case of just one incumbent. Section 5 generalizes the model to the case with two incumbents. Section 6 concludes.

## **2. Sylos Labini's *Oligopolio e Progresso Tecnico* (1957) and Modigliani's 1958 JPE paper**

Sylos Labini never explicitly rejected Modigliani's 1958 interpretation of his 1957 book. Yet, in a few occasions, he appeared to have some reservations on it, particularly as concerns Modigliani's 'invention' of the Sylos postulate (see Arena 2007, Rancan 2015 and Signorino 2016). In the Preface to the Harvard 1962 edition of his book, Sylos Labini wrote that "considering the scientific value of Modigliani's analysis and the original approach to certain specific problems and their solution, his [1958] paper is to be regarded as more than a critical review of the book; it is a new and significant contribution to the theoretical problem of oligopoly. (Sylos Labini 1962, p. vii, emphasis added)". Moreover, he informed his readers that, while he has chosen to revise substantially Part II and III of his former Italian book (devoted to macroeconomic issues), he has left basically unaltered Part I (devoted to equilibrium analysis

in oligopoly). Sylos Labini motivated his choice not to take into consideration Bain (1956) and Modigliani (1958) in the revised edition of his book by claiming that “It seemed to me that if I were to recast my whole analysis, this would not only involve a radical revision of the logic structure of my book, but, just because of the kinship between my analysis and the two others mentioned, might also impair the homogeneous texture of my argument” (Sylos Labini 1962, p. viii, emphasis added).

Sylos Labini’s reservations on Modigliani’s interpretation surfaced in his 2005 personal recollections on Modigliani (see Sylos Labini 2005, p. 42). In particular, as concerns the Sylos postulate, Sylos Labini pointed out:

It is in any case not the rigid assumption one might expect when the term ‘postulate’ is brought in: my point is that the assumption applies in certain market conditions – I begin with a market situation ‘*criée au hasard*’, considering a certain economic space and a given ‘empirical elasticity’ of demand – but not in others. Nevertheless, the fact remains that the existing firms do not necessarily adjust production when other firms attempt entry. Between myself and Franco there are no logical contrasts, but different assumptions, and thus different lines of analysis. (Sylos Labini 2005, p. 43, emphasis added)

Finally, in September 1956 Modigliani sketched his comments to the first provisional Italian edition of Sylos Labini’s book in a long letter where Modigliani anticipated much of the content of his future 1958 paper (see Rancan 2015, pp. 438 ff). Commenting the passage of Modigliani’s letter where the Sylos postulate is introduced, Sylos Labini wrote that “This is not my hypothesis, the interpretation in this specific point is not correct” (quoted in Rancan 2015, p. 441).

We may only speculate on the reasons why Sylos Labini did not wholeheartedly adopt Modigliani’s interpretation of his oligopoly theory. A likely explanation is that Sylos Labini somehow sensed that Modigliani (1958) was not simply his own theory of oligopoly dressed in a mathematical garb. As acutely argued by Rancan (2015), “Modigliani departed from Sylos Labini’s objective approach to the oligopoly problem as an alternative to subjective analysis based on reaction functions, as well as from Sylos Labini’s dynamic approach to economic phenomena” (Rancan 2015, p. 432; see also *idem*, pp. 437-8 and 442). Two issues are here involved, one methodological, objectivism versus subjectivism in the theory of value in oligopoly, and one analytical, classical versus neoclassical notion of competition.

As concerns the methodological issue, Modigliani’s Sylos postulate flies squarely into the face of Sylos Labini’s objectivistic approach to oligopoly theory. Oligopoly models built on such a postulate inevitably turn out i) to place subjective magnitudes right at the core of equilibrium analysis and ii) to gen-



eralize the conjectural variation approach from the traditional Cournot-Edgeworth duopoly models (where entry is not contemplated) to the analysis of new firms' entry into oligopolistic markets. By contrast, as Sylos Labini made clear right at the beginning of Chapter I, 'General Aspects of the Theoretical Problem', he intended to minimize the role played by subjective magnitudes in the equilibrium analysis of an oligopolistic market. Accordingly, he declared his dislike towards the approach of conjectural variations which, in his view, has led to an apparently endless and sterile proliferation of oligopoly models (see Sylos Labini 1962, pp. 19 – 20 and 34). In Sylos Labini's view, the equilibrium analysis of a concentrated oligopoly should be carried out by focusing on objective magnitudes such as technological discontinuities and the structure of the industry, the latter defined as "(a) the absolute size of the market, that is, the volume of sales at a given price; (b) the absorption capacity of the market in the narrow sense previously specified (price elasticity of demand); (c) the distribution of sales among a number of firms of different types" (Sylos Labini 1962, p. 37)<sup>2</sup>

<sup>2</sup> Sylos Labini's reservations on this aspect of Modigliani (1958) raise an interpretative puzzle. Is the Sylos postulate a historical travesty? This is not the place to deal in detail with such an issue (the interested reader may refer to Signorino 2016). It suffices to point out that there is no evidence of such postulate in the first two Italian editions (1956 and 1957) of *Oligopoly and Technical Progress* where Sylos Labini went no further than the following cryptic sentence: "There are no obstacles to the entry of new firms except as implied by the previous assumptions" (Sylos Labini 1957, p. 47 as translated in Sylos Labini 1962, p. 43) to which, in the first 1962 American edition Sylos Labini added: "If new firms enter the market, the existing ones continue to produce as much as before. They do so not only to discourage the entry of new firms, whose additional output in these circumstances depresses the price and so makes the whole market less profitable, but also because by lowering their output the existing firms would raise their total average cost (since on our assumptions total average cost is decreasing up to the limit of plant capacity). For the sake of simplicity, we have ruled out any reduction of output by existing firms as a result of the entry of new firms; if existing firms decide to produce less than maximum output, they do so not under pressure from new entry, but on the basis of independent economic calculations." (Sylos Labini 1962, p. 43). Accordingly, given the strict assumptions on which Sylos Labini developed his analysis of a concentrated oligopoly (that is, average costs quickly raising and profits quickly decreasing as firms produce less than their optimal quantity and a narrow market extension) a sentence of the type 'incumbents do not cut their production in the face of an effective entry into their market' is not a postulate but rather a compelling logical consequence. By contrast, within the S-B-M model a postulate is needed to formalize the potential entrant's conjecture on the likely incumbents' reactions to an effective entry. As remarked by Rancan (2015), Modigliani's commitment to his 'Sylos postulate' most likely derived from his acquaintance with game theory. In fact, that postulate would be consistent with a mini-max approach to uncertainty.

As concerns the analytical issue, Sylos Labini was aware of the theoretical differences existing between the Classical notion of free competition and the Neoclassical notion of perfect competition.<sup>3</sup> The 1962 English text of *Oligopoly and Technical Progress* has a passage which is not to be found in the previous two Italian editions. The passage is the following:

At this point we have to make a brief digression into the nature and the degree of realism of the different concepts of competition, especially in neoclassical and classical theory. The neoclassical economists' perfect competition has never existed anywhere. The static hypothesis and the application of marginal analysis within it led the neoclassical school to ascribe to competition certain characteristics (in particular, perfect homogeneity of goods, horizontal demand curve for the individual firm) which severely limit the realism of the concept. The classical economists' concept of competition, especially Ricardo's is very different. Its principal distinguishing feature is ease of entry. As long as the entry of new firms is easy, no producer can raise prices, and, except for a short time, none can earn above-normal profits. The classical concept ... refers to conditions which might be considered "dynamic": competition is not so much a situation as a process. (Sylos Labini 1962, pp. 52 – 53, Sylos' emphasis)

While the Classical notion of competition is radically different from the Cournot-Walrasian one (see Salvadori and Signorino 2016), Modigliani's interpretation with its Sylos postulate pointed rightly towards a Cournot-type analysis of oligopoly. The likelihood of such an outcome, highly unfit –as we have seen– to Sylos Labini's methodological views, appeared manifest since the 1959 debate in *The Journal of Political Economy*. Both Franklin Fisher (1959) and Farrar and Phillips (1959) stroke a parallel between the Sylos postulate and the Cournotian assumption of given quantities. In both cases, each firm is supposed to choose the quantity to produce by taking as given the quantities produced by all its other competitors, both actual (as in the standard Cournot model) and potential (as in the S-B-M model):

The theoretical model [Modigliani] presents, however, may at first glance

<sup>3</sup> Sylos Labini's acquaintance with Classical economics goes back to the teaching of Alberto Breglia-to whom *Oligopoly and Technical Progress* is dedicated (see Sylos Labini 1962 p. 52 fn 19) and was strengthened, after graduation, by his attendance of Schumpeter's lectures at Harvard and his studies at Cambridge (UK) (Roncaglia 2006, see also Rancan 2015, p. 435). Roncaglia (1991) analyzes the place of Sylos Labini's thought within the rediscovery of the Classical approach engendered by the publication of Piero Sraffa's *Production of Commodities by Means of Commodities* (1960).

cause a feeling of déjà vu, since the behavior assumed in it seems at least reminiscent of Cournot. I discuss this question here primarily with reference to the way in which the results of the Bain - Sylos model as presented by Modigliani compare with the results of a Cournot model enlarged to take account of entry. (Fisher 1959, p. 410)

Modigliani's basic model, we believe, is clearly reminiscent of that developed by August Cournot for his pioneering analysis of the duopoly-oligopoly problem in 1838. [...] And, under the title of "Sylos' postulate" the model also adopts an extremely Cournot-like assumption [...]. (Farrar and Phillips 1959, p. 414)

From such a perspective, the S-B-M model, far from being "a welcome major breakthrough on the oligopoly front" (Modigliani 1958, p. 215), turns out to be but a generalization of a traditional Cournot model: unlike the latter, where the number of incumbents is given and entry is forestalled by natural barriers (Cournot's mineral wells), the former is able to handle new firms' entry. In his reply, Modigliani punctiliously enumerated the differences between a traditional Cournot model and the S-B-M model. Yet, he vigorously upheld his Sylos postulate:

It requires a good deal of obtuseness for the members of the closed Cournot group not to see that they can all improve their lot by co-operating, at least to some extent. But it would seem to be the best part of prudence for a potential entrant to give serious consideration to the possibility that the existing firms will make him as unwelcome as they possibly can (Modigliani 1959, p. 418).

Paraphrasing Farrar and Phillips, Modigliani claimed that Bain and Sylos Labini "have shown a way for coming to grips at least with the problem of price and aggregate output, while sidestepping the traditional 'core of the oligopoly problem', namely, the 'recognition of mutual interdependence'" (*idem*, p. 419). Yet, with the benefit of hindsight, Farrar and Phillips' claim that a major drawback of the S-B-M model is the fact that such a model fails to tackle head-on the problem of oligopolistic firm's interdependence (see Farrar and Phillips 1959, p. 417) appears well-grounded.

### **3. The Sylos Postulate and the Dixit's 1980 model**

Once the problem of strategic interdependence within oligopolistic markets is taken seriously, it is not difficult to show that the Sylos Postulate is the tender spot in the S-B-M model. In fact, the incumbent firm is unable to make

a credible threat to a potential entrant whenever the incumbent's profit associated to a canonical Cournot duopoly equilibrium (that is, an equilibrium with accommodated entry) turns out to be higher than that associated to a duopoly equilibrium with the incumbent keeping its quantity constant. In a game-theoretic jargon, the Sylos Postulate may not be part of a subgame perfect equilibrium.

As is well-known, the way out from the above impasse was provided by Dixit (1980). Dixit started from Schelling's 'paradox of power', that is, "the power to constrain an adversary may depend on the power to bind oneself" (Schelling 1960, p. 22) as applied to Industrial Organization issues by Spence (1977). In Dixit's own words:

[Schelling] suggested that a threat which is costly to carry out can be made credible by entering into an advance commitment which makes its fulfillment optimal or even necessary. This was applied to the question of entry by Spence (1977), who recognized that the established firm's prior and irrevocable investment decisions could be a commitment of this kind. (Dixit 1980, p. 95)

Formally, strategic entry deterrence requires the incumbent to deploy its first move advantage by choosing, within a set of deterrence instruments (such as capitalization rate, brand selection, innovation and advertising etc.), those able to create a binding commitment which irreversibly alters its inter-temporal profit function (see Salop 1979).

Dixit's model analyzes firms' decisions in a two-step setting. In the first step the incumbent chooses capacity, which can be increased but not reduced. In the second step a firm outside the market chooses whether to enter or not. If entry actually occurs, the two firms compete à la Cournot; otherwise the incumbent chooses the quantity to produce as a monopolist. The basic result achieved by Dixit is the identification of a range of quantities produced by the incumbent within which the threat turns out to be credible. Capacity is chosen by the incumbent accordingly. Consequently, "the Spence [1977] excess capacity strategy will not be employed" (Dixit 1980, p. 100).<sup>4</sup>

<sup>4</sup> Though in the final remarks of his paper Dixit is soft in stressing the different conclusions stemming from his and Spence (1977) analysis as concerns the incumbent's choice of excess capacity in the pre-entry phase of the game – "It is not possible to claim universal validity for either of these models" (Dixit 1980, p. 106)– he claims that "in the absence of any asymmetrical advantage possessed by the established firm in the post-entry phase, the Nash solution [Dixit's favorite one] has considerable appeal" (*ibidem*).

The outcome of Dixit (1980) is a generalization of the analytical results achieved by the S-B-M model, while dispensing with the Sylos Postulate (provided the parameters are within some specific values). What is interesting from our perspective is the following. Dixit's self-declared goal is to show that an incumbent has an incentive to manipulate to its own advantage initial conditions by means of a strategic capacity choice, irrespective of the rules of play in the post-entry stage of the game, be it a Cournot-Nash equilibrium or Bertrand-Nash one: see Dixit (1980, pp. 96 and 106). Yet, the Cournotian flavor of the whole thread of analysis is explicitly stated and emphasized. True, Dixit (1980) sketches three extensions of his basic model, one of which refers to "price-setting in the post-entry duopoly, the solution rule being the Bertrand-Nash equilibrium" (p. 105).<sup>5</sup> Yet, a major drawback of this part of the paper is the fact that mixed strategy equilibria are not considered. In our view, the latter is an awkward aspect of Dixit's model since the Cournotian framework is not the most appropriate setting to formalize Sylos Labini's original contribution, which was unambiguously not in terms of quantity competition. Moreover, in Dixit (1980) excess capacity is never observed: the whole accumulated capacity is always actually utilized by the incumbent. This result, however, is a consequence of the assumption on concavity of the demand function. Bulow et al. (1985) have shown, also with a numerical example, that idle capacity in case of no-entry is possible if demand function is not concave (in their example demand has constant elasticity).

Hence, in the following two sections we present a model characterized by both price competition and accumulated capacity which is left idle by the incumbent(s) in the no-entry case. In the next section, we assume that there is just one incumbent firm and one potential entrant. By contrast, in Section 5 we assume that there are two incumbent firms and one potential entrant. In both sections, following Albert Einstein's lead, we try to "make things as simple as possible, but not simpler". Moreover, we use the same notation for both models, even if such a choice makes the model presented in the next section more convoluted than otherwise required.

<sup>5</sup> The other two extensions refer to a post-entry equilibrium in which the entrant acquires the role of quantity leader and to a more general cost function, respectively.

## 4. The Model: one incumbent

There exist two firms: the incumbent firm and the potential entrant  $E$ . Firms play a three-stage game. In the first stage, firm  $M_1$  chooses the level of its capacity,  $k_{M_1} > 0$ . In the second stage firm  $E$  chooses the level of its capacity,  $k_E \geq 0$ ;  $k_E = 0$  implies that firm  $E$  chooses to stay out of the market. In the third stage the two firms compete on prices if firm  $E$  actually enters the market; otherwise the incumbent firm  $M_1$  sets the price.

Each firm  $i$  ( $i = M, E$ ) faces an installing capacity cost in the capacity-setting stage and a constant unit cost of production in the price-setting stage. The installing capacity cost is given by a fixed cost  $F$ , which can be avoided only if  $k_i = 0$ , and a constant unit cost of capacity  $r$ . Moreover, in the third stage, there is a constant unit cost of production  $c$ . For the sake of simplicity we assume that  $c = 0$ . Thus, total cost is given by:

$$C_i(q_i, k_i) = F + cq_i + rk_i$$

for  $i = M, E$ , where  $q_i \leq k_i$  is the quantity produced and sold by each firm, constrained by its capacity. The two firms also face a demand function which, for the sake of simplicity, is assumed to be

$$D(p) = \begin{cases} a - p & \text{if } 0 \leq p \leq a \\ 0 & \text{if } p \geq a \end{cases}$$

with  $a > 2r$ .<sup>6</sup> Since firms may be constrained by their capacity, demand can be rationed. We assume the efficient rationing rule (Levitan and Shubik, 1972). Hence, the pay-offs of the game are given by:

$$\Pi_i(p_{M_1}, p_E, k_{M_1}, k_E) = \begin{cases} p_i \min\{k_i, \max\{a - p_i, 0\}\} - rk_i - F & \text{if } p_i < p_j \\ p_i \min\left\{k_i, \max\left\{\frac{a - p_i}{2}, a - p_i - k_j, 0\right\}\right\} - rk_i - F & \text{if } p_i = p_j \\ p_i \min\{k_i, \max\{a - p_i - k_j, 0\}\} - rk_i - F & \text{if } p_i > p_j \end{cases}$$

The analysis follows by working backwards from the third to the first stage.

<sup>6</sup> The case  $0 < a \leq 2r$  is analytically less interesting and its introduction would imply further complications.

## 4.1. Third Stage

If firm  $E$  did not enter the market in the second stage, firm  $M_1$  chooses  $p_{M_1}$  by maximizing  $\Pi_{M_1} = p(a-p) - rk_{M_1} - F$ , with respect to  $p$ , subject to  $a-p \leq k_{M_1}$ . Hence if  $k_{M_1} > a/2$ ,  $M_1$  chooses the price  $p_{M_1} = a/2$  and produces the quantity  $q_{M_1} = a/2$ : the excess of capacity  $k_{M_1} - a/2$  is left idle. If, on the contrary,  $k_{M_1} \leq a/2$  the whole capacity will be used:  $q_{M_1} = k_{M_1}$  and  $p_{M_1} = a - k_{M_1}$ . Therefore, the profit of  $M_1$  is:

$$\Pi_{M_1} = \begin{cases} k_{M_1}[a - k_{M_1} - r] - F & \text{if } k_{M_1} \leq \frac{a}{2} \\ \frac{a^2}{4} - rk_{M_1} - F & \text{if } k_{M_1} \geq \frac{a}{2} \end{cases} \quad (1)$$

If, on the contrary, firm  $E$  entered the market in the second stage, we have to solve an Edgeworth game in the third stage, since the two firms have already chosen their capacity,  $k_{M_1}$  and  $k_E$ , in previous stages.

It is well-known that an Edgeworth game has three different solutions depending on the size of the two firms. If the two firms are both “large”, that is, if the installed capacity of each firm can match the whole market demand without any rationing ( $k_i \geq D(c) = a$ ), then the usual Bertrand outcome emerges and competition among the two firms pushes their prices down to the marginal costs to be paid in the third stage ( $c$ ), which in our assumptions is zero. If the two firms are both “small”, that is if

$$0 \leq k_E \leq \min \left\{ a - 2k_{M_1}, \frac{a - k_{M_1}}{2} \right\} \quad (2)$$

that is

$$0 \leq k_{M_1} \leq \min \left\{ a - 2k_E, \frac{a - k_E}{2} \right\}$$

both firms charge in equilibrium the price generating a demand equal to the sum of the capacities of the two firms:

$$p_E = p_{M_1} = a - k_{M_1} - k_E \quad (3)$$

As a consequence, the profit of firm  $i$  ( $i = M_1, E$ ) is

$$\Pi_i = [a - k_{M_1} - k_E - r]k_i - F \quad (4)$$

These two equilibria are realized in regions that we call  $A$  and  $B$ , respectively, in Figure 1. In the portion of space between these regions (that is, in regions  $C, D, C'$ , and  $D'$  in Figure 1) equilibria are in mixed strategies. As it is known the largest firm (firm  $M_1$  in regions  $C$  and  $D$  and firm  $E$  in regions  $C'$  and  $D'$ ) gets an expected profit equal to the maximum that it can obtain when the other firm charges a lower price, that is:

$$\Pi_i = \max_p p[a - p - k_j] - rk_i - F = \frac{(a - k_j)^2}{4} - rk_i - F \quad k_i \geq k_j \quad (5)$$

As a consequence, the largest firm will never charge a price lower than

$$p_m = \min \left\{ p : p \min \{ k_i, a - p \} \geq \frac{(a - k_j)^2}{4} \right\}$$

As a consequence, also the other firm will never charge a price lower than  $p_m$ . As is well-known, this fact entails that the expected profit of the smaller firm (firm  $E$  in regions  $C$  and  $D$  and firm  $M_1$  in the other two regions) is:

$$\Pi_j = [p_m - r]k_j - F = \begin{cases} \left[ \frac{(a - k_j)^2}{4k_i} - r \right] k_j - F & \text{if } k_i \leq a - \frac{(a - k_j)^2}{4k_i} \\ \left[ \frac{a - \sqrt{2ak_j - k_j^2}}{2} - r \right] k_j - F & \text{if } k_i \geq a - \frac{(a - k_j)^2}{4k_i} \end{cases}$$

that is

$$\Pi_j = [p_m - r]k_j - F = \begin{cases} \left[ \frac{(a - k_j)^2}{4k_i} - r \right] k_j - F & \text{if } k_i \leq \frac{a + \sqrt{2ak_j - k_j^2}}{2} \\ \left[ \frac{a - \sqrt{2ak_j - k_j^2}}{2} - r \right] k_j - F & \text{if } k_i \geq \frac{a + \sqrt{2ak_j - k_j^2}}{2} \end{cases} \quad (6)$$

Thus, regions  $C$  and  $D$  are separated by the curve

$$k_{M_1} = \frac{a + \sqrt{2ak_E - k_E^2}}{2} \quad (7)$$



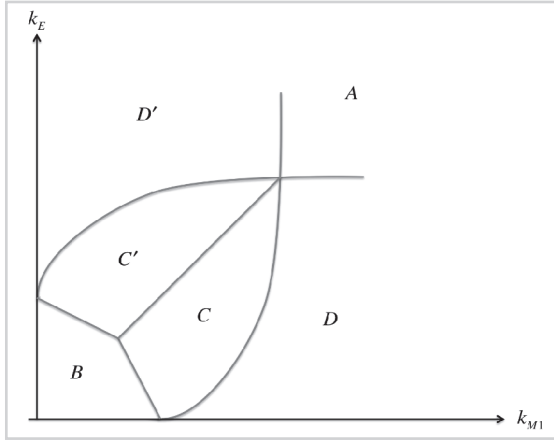


Figura 1. Partition of capacities spacex

## 4.2. Second Stage

In the second stage firm  $E$  chooses the level of its capacity  $k_E$ . In taking this decision it considers as given both  $k_{M1}$  and the functions (5) and (6) relating profits that will be determined in the third stage to  $k_{M1}$  and  $k_E$ . It is easily checked that a situation in which firm  $E$ , confronted with capacity  $k_{M1}$  chosen by firm  $M_1$  in the first stage, chooses in the second stage a capacity  $k_E$  such that  $(k_E, k_{M1})$  is in one of the regions  $A$ ,  $C'$ , and  $D'$  cannot be a Nash equilibrium. Indeed, in region  $A$  the profit of firm  $E$  would be  $\Pi_E = -rk_E - F$ : firm  $E$  would be better off with a lower  $k_E$  and therefore outside this region, in region  $D$ . In regions  $C'$  and  $D'$  we have

$$\Pi_E = \max_p p[a - p - k_{M1}] - rk_E - F = \frac{(a - k_{M1})^2}{4} - rk_E - F$$

once again firm  $E$  would be better off with a lower  $k_E$ , either in regions  $B$ ,  $C$  or  $D$ . A Nash equilibrium in region  $B$  requires that

$$k_E = \frac{a - k_{M1} - r}{2} \quad (8)$$

Indeed, in region  $B$  prices are defined by equation (3) and therefore if

$$k_E \neq \arg \max_k [a - k_{M1} - k - r]k$$

firm  $E$  could have done a better choice. Similarly, a Nash equilibrium in region  $C$  requires that

$$k_E = \frac{2a - \sqrt{a^2 + 12k_{M1}r}}{3} \quad (9)$$

since in region  $C$  profit of firm  $E$  is defined by equation (6) with  $k_i = k_{M1}$ ,  $k_j = k_E$ , and  $k_{M1} \leq \left[ a + \sqrt{2ak_E - k_E^2} \right] / 2$  and therefore if

$$k_E \neq \arg \max_k \left[ \frac{(a-k)^2}{4k_{M1}} - r \right] k$$

firm  $E$  could have done a better choice. For the same reason, a Nash equilibrium in region  $D$  requires that

$$k_E = \hat{k}_E := \arg \max_k \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k. \quad (10)$$

Alternative conditions (8), (9), and (10) are represented as curves  $\alpha$ ,  $\beta$ , and  $\gamma$ , respectively, in Figure 2.

It is easily checked that curves (8) and (9) meet on the boundary between regions  $B$  and  $C$ , whereas curves (9) and (10) do not meet along curve (7) separating regions  $C$  and  $D$ . As a consequence, the curve describing the best choice of firm  $E$  in the assumption that it enters in the market is continuous in the passage between regions  $B$  and  $C$ , but it is not so in the passage between regions  $C$  and  $D$ . Hence, we have to determine the value of  $k_{M1}$  in which it jumps from (9) to (10).

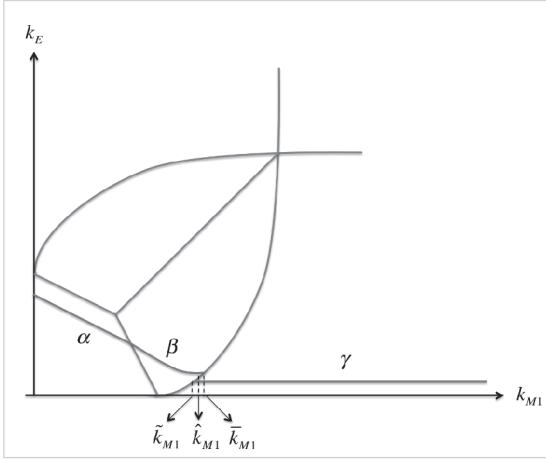


Figure 2. Second stage, best reply of firm E in case of entrance

Let  $(\bar{k}_E, \bar{k}_{M1})$  be the intersection of curves (7) and (9). Similarly, let  $(\hat{k}_E, \tilde{k}_{M1})$  be the intersection of curves (7) with the straight line  $k_E = \hat{k}_E$ . Curve (7) is determined by the equation

$$\left[ \frac{(a - k_E)^2}{4k_{M1}} - r \right] k_E - F = \left[ \frac{a - \sqrt{2ak_E - k_E^2}}{2} - r \right] k_E - F$$

nevertheless

$$\frac{d}{dk} \left[ \frac{(a - k)^2}{4k_{M1}} - r \right] k > \frac{d}{dk} \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k$$

and therefore

$$\arg \max_k \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k = \hat{k}_E < \arg \max_k \left[ \frac{(a - k)^2}{4\bar{k}_{M1}} - r \right] k = \bar{k}_E$$

As a consequence,  $\tilde{k}_{M1} < \bar{k}_{M1}$ , since curve (7) is increasing. If  $\tilde{k}_{M1} \leq k_{M1} \leq \bar{k}_{M1}$ , then (i) firm E chooses to stay on curve  $\beta$ , that is  $k_E = \left[ 2a - \sqrt{a^2 + 12k_{M1}r} \right] / 3$ , whenever

$$\max_k \left[ \frac{(a - k)^2}{4k_{M1}} - r \right] k > \max_k \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k \quad (11)$$

(ii) chooses to stay on curve  $\gamma$ , that is  $k_E = \hat{k}_E$ , whenever

$$\max_k \left[ \frac{(a-k)^2}{4k_{M1}} - r \right] k < \max_k \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k$$

and (iii) is indifferent among the two alternatives if

$$\max_k \left[ \frac{(a-k)^2}{4k_{M1}} - r \right] k = \max_k \left[ \frac{a - \sqrt{2ak - k^2}}{2} - r \right] k \quad (12)$$

In particular, if  $k_{M1} = \bar{k}_{M1}$ , we can take into account the fact that  $(\bar{k}_E, \bar{k}_{M1})$  is on the boundary of region D to obtain that  $\hat{k}_E$  is a better choice than  $\arg \max_k \{ [(a-k)^2 / 4\bar{k}_{M1}] - r \} k$ ; whereas if  $k_{M1} = \tilde{k}_{M1}$ , we can take into account the fact that  $(\hat{k}_E, \tilde{k}_{M1})$  is on the boundary of region C to obtain that  $\arg \max_k \{ [(a-k)^2 / 4\tilde{k}_{M1}] - r \} k$  is a better choice than  $\hat{k}_E$ . Hence, we can define  $\hat{k}_{M1}$  as the unique solution of the equation (12) in  $k_{M1}$ . (Note that the RHS is positive and independent from  $k_{M1}$  whereas the LHS,  $\left[ \frac{a^2 - 12k_{M1}r + a\sqrt{a^2 + 12k_{M1}r}}{4} \right] \left[ \frac{2a - \sqrt{a^2 + 12k_{M1}r}}{54k_{M1}} \right] / 54k_{M1}$ , is a decreasing function of  $k_{M1}$  that in the range  $e \in (0, a^2/4r]$  goes from infinite to 0; note also that inequality (11) certainly holds for  $k_{M1} > a/2$  since the maximand function on the left is higher than the maximand function on the right: hence  $\hat{k}_{M1} > a/2$ .)

Hence the curve describing the best choice of firm E in the assumption that it enters the market is<sup>7</sup>

$$k_E = \begin{cases} \frac{a - k_{M1} - r}{2} & \text{if } 0 \leq k_{M1} \leq \frac{a+r}{3} \\ \frac{2a - \sqrt{a^2 + 12k_{M1}r}}{3} & \text{if } \frac{a+r}{3} \leq k_{M1} \leq \hat{k}_{M1} \\ \hat{k}_E & \text{if } k_{M1} \geq \hat{k}_{M1} \end{cases} \quad (13)$$

This is a correspondence since  $k_E$  can take any of two different values in  $[\hat{k}_{M1}, \infty)$ . Thus, the profit that firm E can obtain is a function of the capacity of firm  $M_1$ , chosen in the first stage and precisely:

<sup>7</sup> Note that  $(a+r)/3$  is the value of  $k_{M1} \geq 0$  in which the graph of equation (8) cuts the constraint (2).

$$\Pi_E(k_{M_1}) = \begin{cases} \left[ \frac{a - k_{M_1} - r}{2} \right]^2 - F & \text{if } 0 \leq k_{M_1} \leq \frac{a+r}{3} \\ \left[ \frac{a^2 - 12k_{M_1}r + a\sqrt{a^2 + 12k_{M_1}r}}{18k_{M_1}} \right] \frac{2a - \sqrt{a^2 + 12k_{M_1}r}}{3} - F & \text{if } \frac{a+r}{3} \leq k_{M_1} \leq \hat{k}_{M_1} \\ \left[ \frac{a - \sqrt{2a\hat{k}_E - \hat{k}_E^2}}{2} - r \right] \hat{k}_E - F & \text{if } k_{M_1} \geq \hat{k}_{M_1} \end{cases} \quad (14)$$

which is a continuous function that is decreasing for  $0 \leq k_{M_1} \leq \hat{k}$  and constant for  $k_{M_1} > \hat{k}_{M_1}$ .

Up to now we have not considered that firm  $E$  has another possible option to choose: to stay out of the market. Firm  $E$  enters the market if and only if the profit it can get in the third stage is positive. If

$$F < \left[ \frac{a - \sqrt{2a\hat{k}_E - \hat{k}_E^2}}{2} - r \right] \hat{k}_E \quad (15)$$

then firm  $E$  will always get positive profits and therefore will enter the market whatever is the choice of firm  $M_1$  in the first stage. If this is not the case, there is a level of incumbent capacity  $Y \leq \hat{k}_{M_1}$  such that  $\Pi_E(Y) = 0$ . In this case the capacity chosen by firm  $E$  is defined by function (13) for  $0 \leq k_{M_1} < Y$  whereas for  $k_{M_1} \geq Y$  the capacity chosen by firm  $E$  is identically equal to 0, that is firm  $E$  stays out of the market. Similarly, the profit of firm  $E$  is defined by function (14) only for  $0 \leq k_{M_1} < Y$  whereas for  $k_{M_1} \geq Y$  the profit of firm  $E$  is identically equal to 0.

### 4.3. First Stage

In the first stage, firm  $M_1$  chooses the level of capacity taking into account the best response of firm  $E$  in the second stage – given by equation (13), if  $0 \leq k_{M_1} < Y$ , and by  $k_E = 0$ , if  $k_{M_1} \geq Y$  – and the equilibria that will be determined in the third stage taking  $k_{M_1}$  and  $k_E$  as given.

If inequality (15) holds, then firm  $M_1$  cannot deter the entry of firm  $E$ . From equations (4), (5), and (13) we obtain that the profit of firm  $M_1$  satisfies the equation

$$\Pi_{M_1} = \begin{cases} k_{M_1} \left[ \frac{a - k_{M_1} - r}{2} \right] - F & \text{if } 0 \leq k_{M_1} \leq \frac{a+r}{3} \\ \left[ \frac{a + \sqrt{a^2 + 12k_{M_1}r}}{6} \right]^2 - rk_{M_1} - F & \text{if } \frac{a+r}{3} < k_{M_1} \leq \hat{k}_{M_1} \\ \left[ \frac{a - \hat{k}_E}{2} \right]^2 - rk_{M_1} - F & \text{if } k_{M_1} \geq \hat{k}_{M_1} \end{cases} \quad (16)$$

which is not a function since there are two distinct values of  $\Pi_{M_1}$  for  $k_{M_1} = \hat{k}_{M_1}$ . Correspondence (16) consists of a segment of parabola, a decreasing function, and a decreasing straight line; it is a continuous function for  $0 \leq k_{M_1} < \hat{k}_{M_1}$ ; and jumps up at  $k_{M_1} = \hat{k}_{M_1}$ , where it takes two different values. Hence there are three candidate pairs  $(k_E, k_{M_1})$  for a Nash equilibrium: the vertex of the parabola,  $((a-r)/4, (a-r)/2)$ , the intersection of the parabola with the decreasing function (if the parabola is increasing there),  $(a-r/(a+r)/3)$ , and  $(\hat{k}_E, \hat{k}_{M_1})$ . There is a problem if the equilibrium is  $(\hat{k}_E, \hat{k}_{M_1})$ : if firm  $M_1$  chooses capacity  $\hat{k}_{M_1}$ , firm E could choose either  $\hat{k}_E$  or  $[2a - \sqrt{a^2 + 12\hat{k}_{M_1}r}]/3$ . But what we are looking for is a Nash equilibrium: only the former choice of firm E leads to a Nash equilibrium; the latter does not. Therefore if firm  $M_1$  chooses capacity  $\hat{k}_{M_1}$ , the Nash equilibrium exists and is  $(\hat{k}_E, \hat{k}_{M_1})$ .<sup>8</sup>

Since both  $((a-r)/4, (a-r)/2)$  and  $(a-r/(a+r)/3)$  are on the parabola, if  $(a-r)/2 < (a+r)/3$ , i.e.  $a < 5r$ ,  $(a-r/(a+r)/3)$  cannot be a Nash Equilibrium, whereas if  $a > 5r$ ,  $((a-r)/4, (a-r)/2)$  cannot be a Nash Equilibrium (if  $a = 5r$ , both points coincide). Hence, if  $a \leq 5r$  firm  $M_1$  has to compare the two local maxima,  $(a-r)^2/8 - F$  and  $[(a - \hat{k}_E)/2]^2 - r\hat{k}_{M_1} - F$ . If  $a \geq 5r$  firm  $M_1$  has to compare  $(a+r)(a-2r)/9 - F$  and  $[(a - \hat{k}_E)/2]^2 - r\hat{k}_{M_1} - F$ . These comparisons are intricate and do not give much insight. Obviously, in each of the two cases, firm  $M_1$  chooses the highest of the two maxima. Figure 3.a shows a numerical example of the incumbent's profits when  $a < 5r$  and the argument of the maximum is  $k_{M_1} = (a-r)/2$ ; Figure 3.b shows a numerical example of the incumbent's profits when  $a < 5r$  and the argument of the maximum is  $k_{M_1} = (a+r)/3$ ; Figure 3.c shows a numerical example of the incumbent's profits when  $a > 5r$  and the argument of the maximum is  $k_{M_1} = \hat{k}_{M_1}$ . Note that if the argument of the maximum is either  $k_{M_1} = (a-r)/2$  or  $k_{M_1} = (a+r)/3$ , then the equilibrium belongs to region A (see Figure 1), hence it is a pure strategy Nash equilibrium whereas if the argument of the maximum is

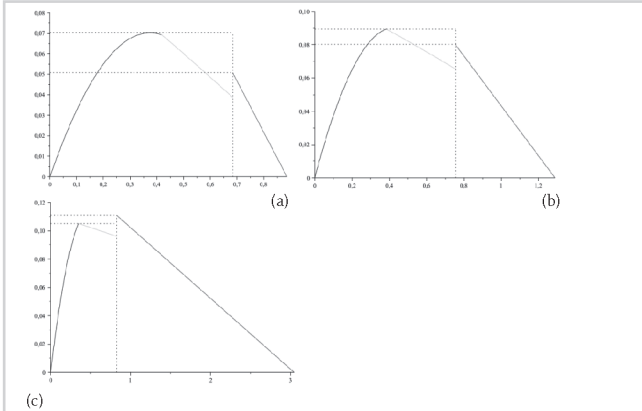
<sup>8</sup> It could be possible to assume that "Whenever the entrant's best response correspondence in (13) is multivalued, the entrant takes the action that maximizes the incumbent's profit." In this way correspondences (13) and (16) become functions and the problem considered in the text vanishes. Actually, this is the route followed by Allen *et al.* (2000). Two remarks are appropriate. First, the assumption is not needed, as the argument in the text proves. Second, the assumption may be justified on the basis of the following two arguments. (i) The assumption simplifies the analysis without hiding a potential Nash equilibrium (but this requires the argument developed in the text above anyway). (ii) As a matter of fact, the problem arises because the price varies in a continuous set; if it would vary in a discrete set, the solution would be close at hand and very close to the action that maximizes the incumbent's profit in the continuous case (this requires a specific analysis). Justifying such an assumption on the economic interpretation (when the entrant is indifferent, it favors the incumbent) does not seem practicable.

$k_{M1} = \hat{k}_{M1}$ , then the Nash equilibrium belongs to region D and it is a mixed strategy equilibrium.

If inequality (15) does not hold there is a level of incumbent capacity  $Y \leq \hat{k}_{M1}$  such that firm E does not enter the market. Thus, the profit of firm M, for  $0 \leq k_{M1} \leq \hat{k}_{M1}$  is determined by (4), (5), (13) and the fact that if  $k_{M1} \geq Y$  the capacity chosen by firm E is identically equal to 0:

$$\Pi_{M1}(k_{M1}) = \begin{cases} k_{M1} \left[ \frac{a - k_{M1} - r}{2} \right] - F & \text{if } 0 \leq k_{M1} < \min \left\{ \frac{a+r}{3}, Y \right\} \\ \left[ \frac{a + \sqrt{a^2 + 12k_{M1}r}}{6} \right]^2 - rk_{M1} - F & \text{if } \frac{a+r}{3} \leq k_{M1} < Y \\ k_{M1}(a - k_{M1} - r) - F & \text{if } Y \leq k_{M1} < \frac{a}{2} \\ \frac{a^2}{4} - rk_{M1} - F & \text{if } \max \left\{ \frac{a}{2}, Y \right\} \leq k_{M1} \end{cases} \quad (17)$$

Function (17) consists of two segments of parabolas (if  $Y > a/2$ , one of the two segments vanishes), a segment of a decreasing function (if  $Y < (a+r)/3$ , it vanishes), and a decreasing half straight line. It is continuous in the ranges  $[0, Y)$  and  $(Y, +\infty)$  and jumps up at  $Y$ , where it assumes only the upper value. It may have local maxima in the vertices of the two parabolas,  $k_{M1} = (a-r)/2$ ; in the intersections of the parabolas with the subsequent decreasing curves,  $k_{M1} = (a+r)/3$  and  $k_{M1} = a/2$ ; in the discontinuity point,  $k_{M1} = Y$ .



**Figure 3. The profit of incumbent when it is not possible to deter the entrance - equation (16). In Figure 3.a,  $a = 1, r = 0.25, F = 0, \Pi_{M1} |_{k_{M1}=(a-r)/2} = 0.0703125$  and  $\Pi_{M1} |_{k_{M1}=\hat{k}_{M1}} \approx 0.0507$ . In Figure 3.b,  $a = 1, r = 0.15, F = 0, \Pi_{M1} |_{k_{M1}=(a+r)/3} = 0.0894$  and  $\Pi_{M1} |_{k_{M1}=\hat{k}_{M1}} \approx 0.0802$ . In Figure 3.c,  $a = 1, r = 0.05, F = 0, \Pi_{M1} |_{k_{M1}=(a+r)/3} = 0.105$  and  $\Pi_{M1} |_{k_{M1}=\hat{k}_{M1}} \approx 0.11087$**

If  $Y \leq (a - r)/2$ , then the entry is blockaded: the incumbent behaves as a monopolist, being sure that the potential entrant will not actually enter. We can distinguish two cases. If  $Y \leq (a - r)/2 \leq (a + r)/3$ , function (17) has a single local maximum: it is increasing on the first parabola, then jumps up on the second parabola at  $Y$ , and is still increasing on the second parabola up to the vertex; then it is decreasing. If  $Y \leq (a + r)/3 < (a - r)/2$ , there is another local maximum at  $k_{M_1} = (a + r)/3$  but the maximum is once again on the vertex of the second parabola.

If  $Y > (a - r)/2$ , function (17) has two local maxima. The first is either in the vertex of the first parabola,  $k_{M_1} = (a - r)/2$ , (if  $(a - r)/2 \leq (a + r)/3$ , i.e.  $a \leq 5r$ ) or in  $k_{M_1} = (a + r)/3$ , if  $a \geq 5r$ . The second local maximum is in  $Y$  and either  $\Pi_{M_1}(Y) = Y(a - Y - r) - F$ , if  $Y \leq a/2$ , or  $\Pi_{M_1}(Y) = a^2/4 - rY - F$ , if  $Y \geq a/2$ . Obviously firm  $M_1$  chooses the highest of the two maxima. If it chooses the first, it accommodates the entry of firm E. If it chooses the second, it deters the entry of firm E. Let us analyze the four cases in turn.

- (a) If  $a \leq 5r$  and  $(a - r)/2 < Y \leq a/2$ , firm  $M_1$  prefers to deter the entry if and only if  $(a - r)^2/8 \leq Y(a - Y - r)$ , that is,  $(2 - \sqrt{2})(a - r)/4 \leq Y \leq (2 + \sqrt{2})(a - r)/4$ . Hence in this case firm  $M_1$  deters the entry if and only if  $(a - r)/2 < Y \leq \min((2 + \sqrt{2})(a - r)/4, a/2)$  and accommodates the entry if and only if  $\min((2 + \sqrt{2})(a - r)/4, a/2) \leq Y \leq a/2$ , which is empty for  $(1 + \sqrt{2})r < a \leq 5r$ .
- (b) If  $a \leq 5r$  and  $Y \geq a/2$ , firm  $M_1$  prefers to deter the entry if and only if  $(a - r)^2/8 \leq a^2/4 - rY$ , that is,  $Y \leq (a^2 + 2ar - r^2)/8r$ . Hence in this case firm  $M_1$  deters the entry if and only if  $a/2 \leq Y \leq \min\left\{(a^2 + 2ar - r^2)/8r, \hat{k}_{M_1}\right\}$  and accommodates the entry if and only if  $\min\left\{(a^2 + 2ar - r^2)/8r, \hat{k}_{M_1}\right\} \leq Y \leq \hat{k}_{M_1}$ .
- (c) If  $a \geq 5r$  and  $(a - r)/2 < Y \leq a/2$ , firm  $M_1$  prefers to deter the entry if and only if  $(a + r)(a - 2r)/9 \leq Y(a - Y - r)$ , which always occur in this case.
- (d) If  $a \geq 5r$  and  $Y \geq a/2$ , firm  $M_1$  prefers to deter the entry if and only if  $(a + r)(a - 2r)/9 \leq a^2/4 - rY$ , that is,  $Y \leq (5a^2 + 4ar + 8r^2)/36r$ .

To present the complete taxonomy, let us first reminder that  $\hat{k}_{M_1}$  is a function of  $a$  and  $r$ ,  $k_{M_1} = g(a, r)$ . Such a function is such that

$$g(a, r) \in \left[ a/2, \left( 3a + \sqrt{5a^2 - 8ar - 4r^2} \right) / 6 \right] \quad g(a, r) > a/2$$

$$\text{since } \hat{k}_{M_1} > \tilde{k}_{M_1} > a/2; \quad g(a, r) < \left( 3a + \sqrt{5a^2 - 8ar - 4r^2} \right) / 6$$

$$\text{since } \left( \left( 3a + \sqrt{5a^2 - 8ar - 4r^2} \right) / 6, (a - 2r)/3 \right)$$



is on the curve separating regions C and D and  $(a - 2r)/3$  is not lower than the best response of firm E in region C (and not larger than the best response of firm E in region B). Then, let  $\hat{a}$  be a solution to the equation

$$\frac{a^2 + 2ar - r^2}{8r} = \hat{k}_{M1}$$

in the range  $\left[ \left( (1 + \sqrt{2})r, 5r \right] \right]$ . Such a solution certainly exists since for  $a = (1 + \sqrt{2})r$

$$\frac{a^2 + 2ar - r^2}{8r} = \frac{(1 + \sqrt{2})r}{2} = \frac{a}{2} < g\left(\left( (1 + \sqrt{2})r, r \right)\right)$$

and for  $a = 5r$

$$\frac{a^2 + 2ar - r^2}{8r} = \frac{17r}{4} > 4r = \left( 3a + \sqrt{5a^2 - 8ar - 4r^2} \right) / 6 > g(5r, r)$$

The above remarks allow us to present the complete taxonomy, that is here presented in Table 1. The first two rows of Table 1 are obvious consequences of points (a)-(b). The last row is a consequence of points (b)-(d) keeping in mind that if  $a \geq 5r$ , then  $(5a^2 + 4ar + 8r^2) / 36r > (a - 2r) / 3 > \hat{k}_{M1}$ .

<b>Table 1. Taxonomy of the choice of firm <math>M_1</math> in the first stage.</b>		
	<b>Deterred Entry</b>	<b>Accommodated Entry</b>
$2r < a < (1 + \sqrt{2})r$	$\frac{1}{2}(a - r) < Y \leq \frac{2 + \sqrt{2}}{4}(a - r)$	$\frac{2 + \sqrt{2}}{4}(a - r) < Y \leq \hat{k}_{M1}$
$(1 + \sqrt{2})r \leq a < \hat{a}$	$\frac{1}{2}(a - r) < Y \leq \frac{a^2 + 2ar - r^2}{8r}$	$\frac{a^2 + 2ar - r^2}{8r} < Y \leq \hat{k}_{M1}$
$a > \hat{a}$	$\frac{1}{2}(a - r) < Y \leq \hat{k}_{M1}$	

Note that anytime entry is deterred firm  $M_1$  chooses the capacity  $k^*_{M1} = Y$  in the first stage, but if  $Y \leq a/2$ , all the capacity is used in the third stage:  $p_{M1} = a - k^*_{M1}$  and  $q = k^*_{M1}$  whereas if  $Y > a/2$ , only part of capacity  $y$  is used in the third stage:  $q_{M1} = a/2 < k^*_{M1}$ ,  $p_{M1} = a/2$ . In this case firm  $M_1$  has idle capacity in equilibrium.

Figure 4 summarizes the results when inequality (15) does not hold by showing in the plane  $(Y, d)$ , the areas in which entry is blockaded, deterred, or accommodated (EB, ED, EA respectively).

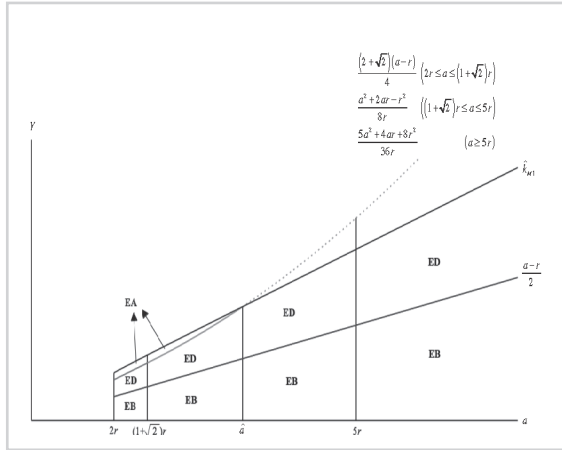


Figure 4. The areas of blocked entry (EB), deterred entry (ED) and accommodated entry (EA) in the plane  $(a, Y)$ .

## 5. The Model: two incumbents

There exist three firms: the incumbent firms  $M_1$  and  $M_2$  and the potential entrant  $E$ . Firms play a three-stage game. In the first stage, firms  $M_1$  and  $M_2$  choose the level of their capacity,  $k_{M_1} > 0$  and  $k_{M_2} > 0$ . In the second stage firm  $E$  chooses the level of its capacity,  $k_E \geq 0$ ;  $k_E = 0$  implies that firm  $E$  chooses to stay out of the market. In the third stage the firms with positive capacities compete on prices.

The structure of costs and the demand function faced by the firms are the same of the model in Section 4. The efficient rationing rule (Levitan and Shubik, 1972) is also assumed. Hence, the pay-offs of the game are given by:

$$\Pi_h^{(\odot)} = \begin{cases} p_h \min\{k_h, \max\{a - p_h, 0\}\} - rk_h - F & \text{if } p_h < p_i \leq p_j \\ p_h \min\left\{k_h, \max\left\{\frac{a - p_h}{2}, a - p_h - k_i, 0\right\}\right\} - rk_h - F & \text{if } p_h = p_i < p_j \\ p_h \min\{k_h, \max\{a - p_h - k_i, 0\}\} - rk_h - F & \text{if } p_i < p_h < p_j \\ p_h \min\left\{k_h, \max\left\{\frac{a - p_h - k_i}{2}, a - p_h - k_i - k_j, 0\right\}\right\} - rk_h - F & \text{if } p_i < p_h = p_j \\ p_h \min\{k_h, \max S\} - rk_h - F & \text{if } p_h = p_i = p_j \\ p_h \min\{k_h, \max\{a - p_h - k_i - k_j, 0\}\} - rk_h - F & \text{if } p_i \leq p_j < p_h \end{cases}$$

where

$$S = \left\{ \frac{a - p_h}{3}, \frac{a - p_h - k_i}{2}, \frac{a - p_h - k_j}{2}, a - p_h - k_i - k_j, 0 \right\}$$

We work backwards and so we should analyze in turn first the third stage, then the second and finally the first. In the analysis of first stage we should build up the reaction functions of the incumbent firms in the space of their capacities and determine the point(s) where they intersect. Since the two incumbents are faced with exactly the same alternatives their reaction functions are symmetrical with respect to the 45° line and, as a consequence, they meet exactly on that line and cannot meet outside it. This means that in the second and in the third stages  $k_{M1} = k_{M2}$ . In next two sub-sessions we analyze the equilibriums that can emerge when this equality holds in the second and in the third stage. This analysis will determine positions that are candidate to be Nash equilibria. The following analysis will be devoted to reduce the set of positions that are candidate to be Nash equilibria.

### 5.1. Third Stage when $k_{M1} = k_{M2}$ and firm $E$ entered the market in the second stage

We follow the results provided by De Francesco and Salvadori (2015, 2016, 2017). Since proofs are very complicated, we refer just to the results as adapted to the simplified demand function here used, with no proof. Readers interested to the proofs can consult the original paper by De Francesco and Salvadori (2015, 2016, 2017).

As we have done in previous section we need a partition of the space  $\{k_{M1}, k_E\}$  since  $k_{M1} = k_{M2}$ . The pure strategy equilibria are in region B ( $k_{M1} \leq a - K$ ) and in region A ( $k_{M1} \leq K - a$ ), where  $K = k_{M1} + k_{M2} + k_E$ . In region B we have that  $p_{M1} = p_{M2} = p_E = a - K$  whereas in region A we have the classical Bertrand solution,  $p_{M1} = p_{M2} = p_E = 0$ . (For a proof see De Francesco and Salvadori, 2015, Proposition 1.) Consequently, in

equilibrium  $\Pi_i = (a - K) k_i - rk_i - F$  in region B and  $\Pi_i = -rk_i - F$  in region A.

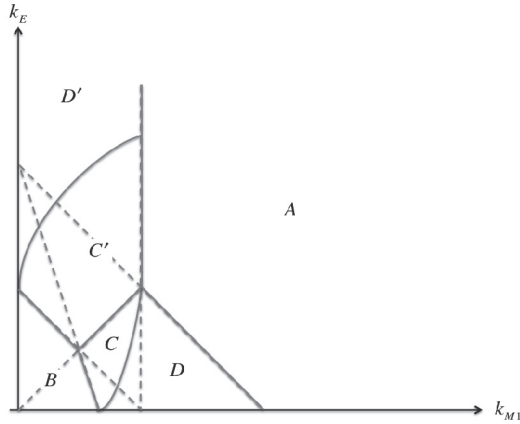


Figure 5. Partition of capacity space.

Region A is delimited by the inequalities

$$k_E \geq a - k_{M1}$$

$$k_{M1} \geq \frac{a}{2}$$

Region B is delimited by the inequalities

$$k_E \leq a - 3k_{M1}$$

$$k_E \leq \frac{a - 2k_{M1}}{2}$$

Regions B' and C' are mixed strategy regions in which  $k_E \geq k_{M1}$ ; they are separated by the curve

$$k_E = \frac{a + \sqrt{4ak_{M1} - k_{M1}^2}}{2}$$

and are irrelevant in this analysis since Firm E will never choose to be in them in the second stage for the same reasons mentioned in the previous section. (Firm E will never choose to be in region A in the second stage either for the same reasons.)

Regions C and D are mixed strategy regions in which  $k_E \leq k_{M1}$ ; they are sep-

arated by the curve

$$k_E = a - k_{M1} - 2\sqrt{(a - 2k_{M1})k_{M1}}$$

which belongs to region C. In region C all firms share the same profit per unit of capacity (apart for the fixed cost), and therefore

$$\Pi_h = p_m k_h - r k_h - F = \frac{(a - k_{M1} - k_E)^2 k_h}{4k_{M1}} - r k_h - F \quad (18)$$

each firm  $h$ . On the contrary, in region D the smaller firm, i.e. the entrant, enjoys a larger profit per unit of capacity (apart for the fixed cost). More precisely, equation (18) still holds for  $h = M_1, M_2$  whereas the profit of firm E is defined by the equation

$$\Pi_E = \max_p p \left[ \varphi_1(p)(1 - \varphi_2(p)) \min\{k_E, a - p - k_{M1}\} + (1 - \varphi_1(p))k_E \right] - r k_E - F \quad (19)$$

where  $\varphi_i(p)$  is the probability that firm  $M_i$  charges a price lower or equal to  $p$  when only firms  $M_i$  and  $M_2$  exist. That is,

$$\varphi_1(p) = \varphi_2(p) = \frac{(p - p_m)k_{M1}}{p(2k_{M1} - a + p)}$$

that is defined in the interval  $[p_m, \tilde{p}]$ , where

$$\tilde{p}(a - \tilde{p} - k_{M1}) = p_m k_{M1} = p_M(a - p_M - k_{M1} - k_E) \quad \tilde{p} < p_M$$

Since

$$p \leq p_M = \frac{a - k_{M1} - k_E}{2}$$

equation (19) can be written

$$\begin{aligned} \Pi_E &= \max_p p \left\{ 1 - \left[ \frac{(p - p_m)k_{M1}}{p(2k_{M1} - a + p)} \right]^2 \right\} k_E - r k_E - F = \\ &= p^* \left\{ 1 - \left[ \frac{(p^* - p_m)k_{M1}}{p^*(2k_{M1} - a + p^*)} \right]^2 \right\} k_E - r k_E - F \end{aligned}$$

where  $p_m < p^* < p^{**} < \tilde{p} < p_M$ . Indeed

$$p \left\{ 1 - \left[ \frac{(p - p_m)k_{M1}}{p(2k_{M1} - a + p)} \right]^2 \right\} = p_m + (p - p_m) \left[ \frac{2k_{M1} - a + p}{k_{M1}} - \varphi_1(p) \right] \frac{k_{M1}}{2k_{M1} - a + p}$$

equals  $p_m$  for  $p = p_m$  and for  $p = p^{**}$  where<sup>9</sup>

$$\frac{2k_{M1} - a + p^{**}}{k_{M1}} - \varphi_1(p^{**}) = 0$$

and is larger than  $p_m$  in the open interval  $(p_m, p^{**})$ . Note that

$$\frac{2k_{M1} - a + p_m}{k_{M1}} > 0 = \varphi_1(p_m)$$

$$\frac{2k_{M1} - a + \tilde{p}}{k_{M1}} < 1 = \varphi_1(\tilde{p}) = \frac{2k_{M1} - a + (a - k_{M1})}{k_{M1}}$$

$$p_M = \frac{a - k_{M1} - k_E}{2} < a - k_{M1}$$

## 5.2. Second Stage when $k_{M1} = k_{M2}$

Firm E enters the market in the second stage if it can get positive profits in the third stage. In this section we first calculate the best choice of firm E for each capacity  $k_{M1} = k_{M2}$  of firms  $M_1$  and  $M_2$ . This allows also to determine, if it exists, a value  $Y$  such that if  $k_{M2} = k_{M1} = Y$ , then firm E does not enter.

In region B the best choice of firm E in the assumption that it enters the market is

$$k_E = \frac{a - r - 2k_{M1}}{2}$$

that cuts the boundary between the region B and the region C in point  $((a+r)/4, (a-3r)/4)$ .

In region C the best choice of firm E in the assumption that it enters the market is

$$k_E = \frac{2(a - k_{M1}) - \sqrt{(a - k_{M1})^2 + 12rk_{M1}}}{3}$$

that cuts the boundary between the region B and the region C in point  $((a+r)/4, (a-3r)/4)$ , so that the best choice of firm E in the assumption that it enters the

<sup>9</sup> Obviously, a straight line and a concave curve meets twice; but function  $\varphi_1(p)$  is defined only in the interval  $[p_m, \tilde{p}]$ .

market is continuous in the regions B and C. This is not the case in the passage from region C and region D. Indeed, we do not calculate the best choice of firm E in region D. We will see that it is not relevant. However, it is relevant to recognize that there is  $\hat{k}_{M_1}$  where the best choice of firm E drops from  $\hat{k}_E$  to  $\hat{k}_E'$ , where  $(\hat{k}_E, \hat{k}_{M_1})$  is in region C and  $(\hat{k}_E', \hat{k}_{M_1})$  is in region D and  $\hat{k}_E < \hat{k}_E'$ . The consequence is that the profit of firms  $M_1$  and  $M_2$  jumps up in  $\hat{k}_{M_1}$ . To see this it is enough to recognize that if we extend the surface describing the profit of firm E in region D to region C, such surface is in region C below the surface describing the profit of firm E in region C; similarly if we extend the surface describing the profit of firm E in region C to region D, such surface is in region D below the surface describing the profit of firm E in region D. The consequence is that along the boundary between region C and region D the partial derivative of the profit of firm E with respect to the capacity of firm E is lower on the side of region D. Hence where along the boundary such derivative is non-positive on the side of region C, it is negative on the side of region D and where along the boundary such derivative is non-negative on the side of region D, it is positive on the side of region C.

In the following we have to consider two cases: one in which there is no  $Y < \hat{k}_{M_1}$  such that when  $k_{M_2} = k_{M_1} = Y$  the best response of firm E implies a naught profit for such a firm; one in which such a  $Y$  exists.

### 5.3. Nash equilibria candidates

In this section we recognize a finite number of candidates to Nash equilibria. To do so we first consider the candidates to Nash equilibria in regions B and C in the assumption that  $k_{M_1}$  is given,  $k_{M_1} - \varepsilon \leq k_{M_2} \leq k_{M_1} + \varepsilon$ , and  $k_E$  reacts to the movements of  $k_{M_2}$ , which are implemented in the previous stage. Second, we will consider three critical points: point  $((a+r)/4, (a-3r)/4)$  point  $(\hat{k}_{M_1}, \hat{k}_E)$ , point  $(Y, 0)$  where  $k_{M_2} = k_{M_1} = Y$  and firm E does not enter since otherwise its profit would be exactly equal to zero; also in these critical points we will keep  $k_{M_1}$  fixed at the requested value,  $k_{M_1} - \varepsilon \leq k_{M_2} \leq k_{M_1} + \varepsilon$ , and  $k_E$  reacts to the movements of  $k_{M_2}$ . The existence of a candidate Nash equilibrium in the interior of region B requires that the following equations hold:

$$\frac{\partial}{\partial k_E} (a - r - k_{M_1} - k_{M_2} - k_E) k_E = a - r - k_{M_1} - k_{M_2} - 2k_E = 0$$

$$\frac{\partial}{\partial k_{M_2}} \left( a - r - k_{M_1} - k_{M_2} - \frac{a - r - k_{M_1} - k_{M_2}}{2} \right) k_{M_2} = \frac{a - r - k_{M_1} - 2k_{M_2}}{2} = 0$$

and  $k_{M_1} = k_{M_2}$

$$k_{M_1} = k_{M_2} = \frac{a - r}{3}$$

$$k_E = \frac{a-r}{6}$$

The existence of such a candidate requires that

$$\frac{a-r}{3} < \frac{a+r}{4} \Leftrightarrow a < 7r$$

To investigate the existence of a candidate Nash equilibrium in the interior region C we need to distinguish two cases, the case in which  $k_{M_2} \leq k_{M_1}$  and the case in which  $k_{M_2} \geq k_{M_1}$ . If  $k_{M_2} \leq k_{M_1}$  the existence of a candidate Nash equilibrium in the interior of region C requires that the following equation and the following inequality hold:

$$\begin{aligned} \frac{\partial}{\partial k_E} \left[ \frac{(a-k_{M_2}-k_E)^2}{4k_{M_1}} - r \right] k_E &= \frac{(a-k_{M_2}-k_E)^2}{4k_{M_1}} - r - \frac{2(a-k_{M_2}-k_E)k_E}{4k_{M_1}} = 0 \\ \frac{\partial}{\partial k_{M_2}} \left[ \frac{\left( a-k_{M_2} - \frac{2(a-k_{M_2}) - \sqrt{(a-k_{M_2})^2 + 12rk_{M_1}}}{3} \right)^2}{4k_{M_1}} - r \right] k_{M_2} &= \\ &= \frac{(a-k_{M_2})(a-3k_{M_2}) \left[ \sqrt{(a-k_{M_2})^2 + 12rk_{M_1}} + (a-k_{M_2}) \right] - 12rk_{M_1} \left[ \sqrt{(a-k_{M_2})^2 + 12rk_{M_1}} - (a-2k_{M_2}) \right]}{18k_{M_1} \sqrt{(a-k_{M_2})^2 + 12rk_{M_1}}} \geq 0 \end{aligned}$$

which never holds (note that the brackets are positive and that in  $Ca/3 \leq k_{M_2} \leq k_{M_1} \leq a/2$ ).

If  $k_{M_2} \geq k_{M_1}$  the existence of a candidate Nash equilibrium in the interior of region C requires that the following equation and the following inequality hold:

$$\begin{aligned} \frac{\partial}{\partial k_E} \left[ \frac{(a-k_{M_1}-k_E)^2}{4k_{M_2}} - r \right] k_E &= \frac{(a-k_{M_1}-k_E)^2}{4k_{M_2}} - r - \frac{2(a-k_{M_1}-k_E)k_E}{4k_{M_2}} = 0 \\ \frac{\partial}{\partial k_{M_2}} \left[ \frac{\left( a-k_{M_1} - \frac{2(a-k_{M_1}) - \sqrt{(a-k_{M_1})^2 + 12rk_{M_2}}}{3} \right)^2}{4k_{M_2}} - r \right] k_{M_2} &= \frac{(a-k_{M_1}) - 2\sqrt{(a-k_{M_1})^2 + 12rk_{M_2}}}{3\sqrt{(a-k_{M_1})^2 + 12rk_{M_2}}} r \leq 0 \end{aligned}$$

which is always satisfied, but never as an equation.

Let us consider now point  $((a+r)/4, (a-3r)/4)$ . If  $k_{M_2} < k_{M_1} = (a+r)/4$ , we are in the pure strategy region and therefore the profit of firm  $M_2$  is decreasing with  $k_{M_2}$  if  $a < 7r$ , otherwise it is increasing. In the latter case  $((a+r)/4, (a-3r)/4)$  is a candi-



date to a Nash equilibrium since if  $k_{M_2} > k_{M_1} = (a+r)/4$ , we are in the mixed strategy region and, as we have seen above, the profit of firm  $M_2$  is decreasing with  $k_{M_2}$ . In the former case  $((a+r)/4, (a-3r)/4)$  is not a candidate to a Nash equilibrium.

Let us consider now point  $(\hat{k}_{M_1}, \hat{k}_E)$  in the assumption that Y does not exist. If  $k_{M_2} < k_{M_1} = \hat{k}_{M_1}$ , then we are in the region in which the profits of firms are proportional to their capacities (apart for fixed cost) and, as we have seen, the profit of firm  $M_2$  is decreasing with  $k_{M_2}$ , but as soon as  $k_{M_2} = \hat{k}_{M_1}$  the profit of firm  $M_2$  jumps up. On the contrary if  $k_{M_2} > k_{M_1} = \hat{k}_{M_1}$ , we are in the region in which the profit of the smaller firm (firm E) is in proportion higher than the profit of the other firms (apart for fixed cost); we have not calculated the best response of firm E in this case, but it is quite obvious to think that the profit of firm  $M_2$  is decreasing with  $k_{M_2}$  and, in any case, since we are just determining candidates to Nash equilibria, what matters is that we cannot exclude that this is a Nash equilibrium.

Let us assume now that Y exists. If  $Y \leq (a-r)/3$ , i.e. if  $F \geq (a-r)^2/36$ , then the entry is blockaded: firms  $M_1$  and  $M_2$  behave like the duopolists in Kreps and Scheinkman (1983) and firm E does not enter.

If  $(a-r)/3 < Y \leq (a+r)/4$  the profit of firm  $M_2$  jumps from  $(a-r-Y-k_{M_2})k_{M_2}/2$  to  $(a-r-Y-k_{M_2})k_{M_2}$  in Y and point  $(Y, 0)$  is a potential Nash equilibrium since the profit of firm  $M_2$  is decreasing on the right of Y. Similarly,  $(Y, 0)$  is a potential Nash equilibrium if  $(a+r)/4 < Y < \hat{k}_{M_1}$ : the profit of firm  $M_2$  jumps in Y and is decreasing on the right of Y.

## 5.4. Nash equilibria

In this section we will complete the analysis of Nash equilibria.

Let us first assume that Y exists, i.e. that

$$F > \left[ \frac{\left( (a - \hat{k}_{M_1}) + \sqrt{(a - \hat{k}_{M_1})^2 + 12r\hat{k}_{M_1}} \right)^2}{36\hat{k}_{M_1}} - r \right] \frac{2(a - \hat{k}_{M_1}) - \sqrt{(a - \hat{k}_{M_1})^2 + 12r\hat{k}_{M_1}}}{3}$$

We have seen in previous section that if  $Y \leq (a-r)/3$ , then the entry is blockaded. If  $a \leq 7r$  and  $(a-r)/3 < Y \leq (a+r)/4$ , then there are two Nash equilibrium candidates,  $((a-r)/3, (a-r)/6)$  and  $(Y, 0)$ . In the former entry is accommodated, in the latter it is deterred. The former is a Nash equilibrium if

$$(a-r-2Y) \left( 2Y - \frac{a-r}{3} \right) < \frac{(a-r)^2}{18}$$

that is if the profit of firm  $M_2$  when  $k_{M_1} = (a - r)/3$ ,  $k_{M_2} = 2Y - (a - r)/3$ , and  $k_E = 0$  is lower than the profit of firm  $M_2$  when  $k_{M_1} = k_{M_2} = (a - r)/3$ , and  $k_E = (a - r)/6$ . Hence the former Nash equilibrium candidate is a Nash equilibrium if  $a < 6(\sqrt{2} - 1)r$  and  $(4 + 2)(a - r)/12 < Y \leq (a + r)/4$ . The latter Nash equilibrium candidate actually is a Nash equilibrium if

$$\frac{(a - r - Y)^2}{8} < (a - r - 2Y)Y$$

that is if the profit of firm  $M_2$  when  $k_{M_1} = Y$ ,  $k_{M_2} = (a - r - Y)/2$ , and  $k_E = (a - r - Y)$  is lower than the profit of firm  $M_2$  when  $k_{M_1} = k_{M_2} = Y$ , and  $k_E = 0$ . Hence the latter Nash equilibrium candidate actually is a Nash equilibrium in the whole range  $(a - r)/3 < Y \leq (a + r)/4$  if  $a > (1 + 16\sqrt{2})r/7$ , otherwise it is a Nash equilibrium only if  $(a - r)/3 < Y \leq (5 + 2\sqrt{2})(a - r)/17$ .

A simple taxonomy for these cases is presented in Table 2. Note that for  $3r < a < 6(\sqrt{2} - 1)r$  there are values of  $Y$ 's in which two equilibria exist. This is not surprising since the comparison is not between two Nash equilibrium candidates, as in the analysis presented in the previous section.

<b>Table 2. Taxonomy of Nash equilibria when <math>a \leq 7r</math> and <math>(a - r)/3 &lt; Y \leq (a + r)/4</math>.</b>		
	Deterred Entry	Accommodated Entry
$3r < a < \frac{1 + 16\sqrt{2}}{7}r$	$\frac{a - r}{3} < Y \leq \frac{5 + 2\sqrt{2}}{17}(a - r)$	$\frac{4 + \sqrt{2}}{12}(a - r) \leq Y \leq \frac{a + r}{4}$
$\frac{1 + 16\sqrt{2}}{7}r \leq a < 6(\sqrt{2} - 1)r$	$\frac{a - r}{3} < Y \leq \frac{a + r}{4}$	$\frac{4 + \sqrt{2}}{12}(a - r) \leq Y \leq \frac{a + r}{4}$
$6(\sqrt{2} - 1)r \leq a \leq 7r$	$\frac{a - r}{3} < Y \leq \frac{a + r}{4}$	Never

If  $(a + r)/4 < Y < \hat{k}_{M_1}$ , then there are two Nash equilibrium candidates: one is either  $((a - r)/3, (a - r)/6)$ , if  $a \leq 7r$ , or  $((a + r)/4, (a - 3r)/4)$ , if  $a > 7r$ ; the other is  $(Y, 0)$ . In the former entry is accommodated, in the latter it is deterred. The analysis of when each of the two actually is a Nash equilibrium is very tiresome with no much insight and is left to the interested reader.

If  $Y$  does not exist, there exist two Nash equilibria candidates: one is either  $((a - r)/3, (a - r)/6)$ , if  $a \leq 7r$  or  $((a + r)/4, (a - 3r)/4)$ , if  $a > 7r$ ; the other is  $(\hat{k}_{M_1}, \hat{k}_E)$ . At least one of the two is a Nash equilibrium, but both can be Nash equilibria. Indeed, we do not have just to compare the two points, but to check what happens when  $k_{M_1}$  is given at the level required and  $k_E$  reacts to the move-

ments of  $k_{M2}$ .

## 6. Concluding remarks

Following the lead of Modigliani (1958), a copious literature has blossomed on limit pricing models also called S-B-M models after Joe Bain, Paolo Sylos Labini and Franco Modigliani. In this paper we have assessed Modigliani's 1958 rational reconstruction of Paolo Sylos Labini's theory of oligopoly and we have found it wanting as concerns two crucial analytical issues. First, Modigliani's invention of the Sylos postulate paves the way towards a Cournot-like interpretation of Sylos Labini's view of market competition which flies in the faces of Sylos Labini's own view of firms setting prices. Accordingly, in our view Sylos Labini's theory of oligopoly must be assessed against the backdrop of a framework based on Bertrand competition. Second, in the standard S-B-M model, such as Dixit (1980), the productive capacity installed by incumbents for entry deterrence purpose is never left idle, irrespective of actual external firms' entry. By contrast, within Sylos Labini's original framework, the entry-deterrence capacity installed by the incumbents does not need to be employed in the no-entry scenario.

To 'vindicate' Sylos Labini's theory of oligopoly and set it free from the Procrustean bed of Cournot competition, we have developed a three stage Bertrand model with entry and capacity setting and we have analyzed its formal properties both in a one-incumbent scenario and in a two-incumbent scenario. In particular, as concerns the one-incumbent variant of our model, we have presented a complete taxonomy of blockaded, deterred and accommodated entry and a graphical representation of the incumbent's choice in the first stage of the game. As concerns the two-incumbent variant, all the results of the one-incumbent scenario have been confirmed, but now a problem of coordination among incumbent firms may emerge since two different Nash equilibria may exist: one in which both incumbents work to deter the entrance and one in which both incumbents work to accommodate the entrance.

## References

Allen B., Deneckere R., Faith T., Kovenock D. (2000), "Capacity precommitment as a barrier to entry: A Bertrand-Edgeworth approach", *Economic Theory*, 15 (3): 501-530.

- Arena R. (2007), "La théorie de l'oligopole de Sylos Labini: diversité des interprétations et prolongements possibles", *Revue d'Économie Industrielle*, 118 (2): 37- 54.
- Baik K.H. (1995), "Horizontal Mergers of Price-Setting Firms with Sunk Capacity Costs", *The Quarterly Review of Economics and Finance*, 35 (3): 245- 256.
- Bain Joe S. (1956), *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*, Cambridge, Mass., Harvard University Press.
- Bulow J., Geanakoplos J., Klemperer P. (1985), "Holding Idle Capacity to Deter Entry", *The Economic Journal*, 95 (1): 178- 182.
- De Francesco M.A., Salvadori N. (2015), "Bertrand-Edgeworth games under triopoly: the payoffs", *Munich Personal RePEc Archive*, <https://mpra.ub.uni-muenchen.de/64638/>
- De Francesco M.A., Salvadori N.(2016), "Bertrand-Edgeworth games under triopoly: the equilibrium strategies when the payoffs of the two smallest firms are proportional to their capacities", *Munich Personal RePEc Archive*, <https://mpra.ub.uni-muenchen.de/69999/>
- De Francesco M.A., Salvadori N. (2017), *Bertrand-Edgeworth games under triopoly: the equilibrium strategies when the payoffs of the two smallest firms are not proportional to their capacities*, unpublished manuscript.
- Dixit A. (1980), "The Role of Investment in Entry-Deterrence". *The Economic Journal*, 90 (1): 95 - 106.
- Farrar D.E., Phillips C.F. (1959), "New Developments on the Oligopoly Front: A Comment". *The Journal of Political Economy*, 67 (4): 414 - 417.
- Fisher F.M. (1959), "New Developments on the Oligopoly Front: Cournot and the Bain-Sylos Analysis". *The Journal of Political Economy*, 67 (4): 410 - 413.
- Gilbert R, Vives X., (1986), "Entry deterrence and the free rider problem", *Review of Economic Studies*, Vol. 53, pp. 71-83.
- Kreps, D.M. and Scheinkman J.A, (1983), "Quantity precommitment and Bertrand competition yield Cournot outcomes", *The Bell Journal of Economics*, 14 (2): 326 - 337.
- Modigliani F. (1958), "New Developments on the Oligopoly Front". *The Journal of Political Economy*, 66 (3): 215 - 232.
- Modigliani F. (1959), "New Developments on the Oligopoly Front: Reply", *The Journal of Political Economy*, 67 (4): 418 - 419.
- Rancan A. (2015), "The Origin of the Sylos Postulate: Modigliani's and Sylos Labini's Contributions to Oligopoly Theory", *Journal of the History of Economic Thought*, 37 (3): 431 - 448.
- Roncaglia A. (1991), "The Sraffian schools", *Review of Political Economy*, 3 (2): 187 - 219.
- Roncaglia, A. (2006), "Paolo Sylos Labini, 1920 - 2005", *Moneta e Credito*, 59 (233): 3 - 21.
- Salop S.C. (1979), "Strategic Entry Deterrence", *The American Economic Review* 69 (2): 335 - 338.
- Salvadori N., Signorino R. (2012), *Competition*. In Gilbert Faccarello and Heinz D. Kurz (eds) (2016). *Handbook on the History of Economic Analysis Volume III Developments in Major Fields of Economics*. Cheltenham, UK and Northampton, MA, USA: Edward Elgar, pp. 70 - 81.
- Schelling T.C. (1960), *The Strategy of Conflict*, Cambridge, Mass.: Harvard University Press.

## Paolo Sylos Labini vindicated

- Signorino R. (2016), *The Sylos Postulate reconsidered*. In Giuseppe Freni, Heinz D. Kurz, Andrea Mario Lavezzi and Rodolfo Signorino (eds). *Economic Theory and its History. Essays in Honour of Neri Salvadori*, London and New York: Routledge, pp. 318 – 335.
- Spence A.M. (1977), "Entry, Capacity, Investment and Oligopolistic Pricing". *The Bell Journal of Economics*, 8 (2): 534 – 544.
- Sylos Labini P. (1957), *Oligopolio e Progresso Tecnico*. 2<sup>nd</sup> and revised edition, 1<sup>st</sup> edition 1956, Milano: Giuffrè.
- Sylos Labini P. (1962), *Oligopoly and Technical Progress*. Cambridge, Mass.: Harvard University Press.
- Sylos Labini P. (2005), Franco Modigliani and oligopoly. *BNL Quarterly Review*, 58 (233 – 234): 41 – 48.



# **Regulation and deregulation policies in an economy with imperfect competition and firms entry**

LORENZO CORSINI

## **1. Introduction**

The aim of this paper is to build a dynamic model that explains the process of firms entry and ties it to the presence of imperfect competition in the markets and to use it to provide policy implications in terms of regulating and deregulating the markets. The idea behind this is rather straightforward: imperfect competition contributes to the determination of profits and through them it influences the entry of firms; as long as policies can affect imperfections and firms entry they can also affect the very working of the economy.

If we imagine that entry is costly and that this cost depends on the number of entrant firms and on some fixed component, then it is possible to build a dynamic process and to study its properties. Moreover, the relationship between imperfect competition and firms' entry does not run in one direction only but it also acts, indirectly, in the other way. In fact, the process of entry contributes to determine the number of firms and thus it is likely to influence the degree of market power. Then, not only we can analyse the entry dynamics but we can also examine the sources of market power.

The entry of firms in an imperfectly competitive environment has been studied in some classical works by Modigliani (1958) and Sylos-Labini (1962) which established the well-known limit pricing model. The focus of that approach was on the entry deterrence and on the (limit) pricing that guarantees such deterrence. The classical two stage, two competitor model was then extended over continuous, infinite time in Gaskin (1971) and to multiple entrants and incumbents in Gilbert and Vives (1986). All the above works aim to

prove the existence of a limit pricing, leaving the study of the dynamic process and of the very role of market imperfection as side questions. For these reasons those models cannot be the base of our analysis.

The study of the dynamic aspects have found more space in recent literature: for example, Das and Das (1997) try to evaluate the entry/exit process with non-homogenous firms, while Datta and Dixon (2002) try to evaluate the impact that improvements in productivity have in the entry dynamics. As for the role of regulation and deregulation, analyses contained in Blanchard and Giavazzi (2003) and Spector (2004) tries to assess the effect of imperfect competition when the number of firms is not given: however, these contributions only introduce firms entry and do not explicitly formalize it. On similar lines, Corsini (2009) analyses theoretically the macroeconomics consequences of regulation and deregulation without considering firms' entry.

What we try to do in this paper is to develop a model where a fixed number of firms operate in an oligopolistic market and where unions bargain on wages and then we explore what happens when new firms can enter the market. We consider the former case as the short run and the latter as the long run<sup>1</sup>. In effect, this allows us to endogenize the number of firms.

The introduction of a dynamic component obviously complicates the analysis and will force us to apply some simplification to make the problem tractable.

The other point on which we focus our attention is the effect that (de) regulation policies have on some key variables (employment, real wages and profits mainly) in a context where the number of firms is endogenous. In particular, we can examine the (de)regulation of product markets through the change in the administrative (entry) costs and of the labour market through a change in unions' bargaining power. Even more interestingly, we are able to compare how those policies affects the markets both in the short and the long run and we can check for the existence of complementarity in the deregulation of different markets through different time horizons.

The paper is organized as follows: in section 2 we build the basic model, in section 3 we study the short run, in section 4 we introduce firms entry and study the dynamics of the process, in section 5 we study the long run, in section 6 we compare the effect of deregulation of the markets in the short and long run and in section 7 we conclude.

<sup>1</sup> We borrow this distinction from Blanchard and Giavazzi (2003).



## 2. Basic Framework

Our model is based on a standard Cournot oligopoly, similar to the one described in Dixon (1988) and to which we add the presence of unions that bargain over wages.

The economy is populated by  $n$  identical firms: each firm  $i$  hires labour to produce a  $y_i$  amount of a homogenous good. Firms are risk-neutral and their production function is

$$y_i = aL_i \quad (1)$$

where  $L_i$  is the amount of labour hired by firm  $i$  and  $a$  is a parameter that measures the average and marginal productivity.

The product market works as in a Cournot oligopoly: the prevailing market price  $P$  is given by the following inverse demand function

$$P = A - \sum_{i=1}^n y_i \quad (2)$$

And for simplicity we normalize the parameter  $A=1$  so that

$$P = 1 - \sum_{i=1}^n y_i, \quad (2a)$$

### Firms behaviour

Firms sell their product in an oligopolistic market: their profits  $\pi_i$  are given by

$$\pi_i = Py_i - W_i L_i \quad (3)$$

where  $W_i$  is the wage paid by firm  $i$ . As it will become clearer later, each firm pay the same wage so that  $W_i$  is the same for any  $i$ .

Given equation (1) and (2a) we can write the above as

$$\pi_i = \left( 1 - \frac{W_i}{a} - \sum_{i=1}^n y_i \right) y_i, \quad (4)$$

In order to obtain positive profits it is necessary to assume that  $W_i < a$ : this is a reasonable assumption that simply implies that the average cost does not exceed average productivity.

Each firm chooses the quantity of  $y_i$  that maximizes the profits, taking as given all the competitors' quantity (this is the standard hypothesis in the Cournot oligopoly), therefore:

$$\frac{\partial \pi_i}{\partial y_i} = 0. \quad (5)$$

Since all firms are identical, they all produce the same amount in equilibrium and the above condition yields

$$y_i = y^* = \frac{\left(1 - \frac{W_i}{a}\right)}{1+n} \quad (6)$$

Where  $y^*$  is the equilibrium amount of production of each firm. Equation (6) is the standard Nash-Cournot solution.

Given that amount of production we can derive the resulting price: combining (2a) with (6) we have

$$P = \frac{\left(1+n \frac{W_i}{a}\right)}{1+n} \quad (7)$$

which, for  $W_i < a$ , is a decreasing function of the number of firm  $n$ .

### Union bargaining

The workers of each firm are organized in unions that bargain over wages: the bargaining happens at the firm level. There is a union for each firm and they are interested in both employment and wages. Their utility function is given by

$$U_i = (W_i - B)L_i \quad (8)$$

where  $B$  is the outside option. The utility depends on the nominal wages because bargaining happens at the firm level and therefore each union thinks its bargaining will not affect the general level of price and takes the latter for granted. If this is the case, then it makes no difference whether it is the nominal or the real wage<sup>2</sup> that enters the utility function<sup>3</sup>. This hypothesis is reasonable as long as the bargaining is decentralized and the number of firms is not too small. We also imagine that each union shares the same bargaining power.

<sup>2</sup> In fact, if the utility function depends on the real values we would have  $U_i = (W_i / P - B / P)L_i$  where  $P$ , if exogenous, has only a scale effect.

<sup>3</sup> See Blanchard and Giavazzi (2003) for an example of a work where such an assumption is made.

During the bargaining, each union tries to obtain the wage that maximizes (8): such wage is given by

$$W_i^C = \operatorname{argmax}_{w_i} (W_i - B)L_i = (a + B) / 2 \quad (9)$$

where  $W_i^C$  represents the claims of each union in terms of wages and  $L_i$  depends on the optimal production chosen by firms according to (6). Obviously, unions would never accept a wage lower than the outside option  $B$  and, if the unions had the right to set unilaterally the wages (the case of the monopoly union) then they would choose  $W_i^C$ . However, firms oppose to this and try to settle on a lower wage (in fact lower wages guarantee higher profits). In the end, the parts will settle on a value between  $B$  and  $(a+B)/2$  where the exact outcome depends on the bargaining power of the parts.

Here we imagine that the outcome is the weighted average between the wage claim and the fall back value, with the bargaining power of unions  $\beta$  (that ranges 0 and 1) being the weight:

$$W_i^B = \frac{a - B}{2} \beta + B \quad (10)$$

where  $W_i^B$  is the bargained wage: when  $\beta=1$  the union claims are fully met and  $W_i^B = W_i^C$  while  $\beta=0$  implies a wage equal to the outside option. Obviously, higher bargaining power leads to higher wages.

Since each union has the same bargaining power, the bargained wage will be the same in each firms and will determine the economy-wide wage  $W^B$ :

$$W^B = W_i^B = \frac{a - B}{2} \beta + B \quad (10a)$$

### 3. The short run analysis

The short run is defined by the fact that the number of firms is fixed. That number is exogenous and can be considered merely a parameter. In this circumstances the only (de)regulation policies that can affect the markets are those affecting the bargaining power of unions. In what follows we derive some key variables and examines how they are affected by the number of firms and by bargaining power

### Market power and profits

The above scheme allows us to derive the relationship between the number of firms, the degree of market power and the resulting profits. If we measure the market power  $\mu$  as the price-cost margin (as suggested in Lerner 1934), we have

$$\mu = \left( 1 - \frac{\frac{a-B}{2}\beta + B}{a} \right) / \left( 1 + n \frac{\frac{a-B}{2}\beta + B}{a} \right) \quad (11)$$

which is inversely related to the number of firms and to the bargaining power of unions.

The profits are given by

$$\pi_i = \left( \frac{1 - \frac{W_i^B}{a}}{1 + n} \right)^2 \quad (12)$$

which is again inversely related to the number of firms.

If we insert the bargained wage (10) in (12) we have

$$\pi_i = \left( \frac{1 - \frac{a-B}{2}\beta - B}{1 + n} \right)^2 \quad (13)$$

that shows that profits are lower when  $\beta$  or  $B$  are higher.

We can also obtain the prevailing price inserting the bargained wage (10) in the equation for price (7):

$$p = \frac{1 + n \left( \frac{a-B}{2a}\beta - \frac{B}{a} \right)}{1 + n} \quad (14)$$

which tells us that a higher number of firms reduces the price level (in fact it reduces the price-cost margin, as we saw above) and that higher bargaining power increases it (through the increase of firms costs).

The picture we get is quite clear, a product market with few firms shows high mark-up and high prices allowing for higher profits. On the other end, a labour market highly regulated (where unions have a high bargaining power) produces lower profits for firms.

### Employment and real wages

Given the fixed number of firms  $n$  we can derive the level of employment and real wages.

The production of each firm is given by (6) and if we combine it with the bargained wage (10a) we have the employment per firm

$$L_i = \frac{1}{1+n} \frac{1 - \frac{a-B}{2} \beta - B}{a} \quad (15)$$

and similarly, the aggregate employment  $L$  is

$$L = \frac{n}{1+n} \frac{1 - \frac{a-B}{2} \beta - B}{a} \quad (16)$$

which depends positively on  $n$  and negatively on  $\beta$ .

The number of firms impact negatively on the employment per firms but positively on aggregate employment. In fact, when more firms are in the market, each of them has a smaller share but the resulting price is lower so that aggregate demand (and employment) is higher.

On the other side, bargaining power increases wages and reduce the demand for labour, leading eventually to lower employment.

As for the real wage, we can obtain it dividing the bargained nominal wage (10a) for the prevailing price (14):

$$\frac{W^B}{P} = \frac{1+n}{\frac{2}{(a-B)\beta - 2B} + n/a} \quad (17)$$

so that both the bargaining power and the number of firms have a positive effects on it<sup>4</sup>. Both results are direct, with stronger unions getting better wages and with markets with more firms producing lower price-wages margins.

Summing up, in the short run, more concentrated product markets (with less firms) yields higher profits and lower aggregate production, employment and real wages; on the other side more regulated labour markets (where unions have more bargaining power) yields higher real wages but lower profits, aggregate production and employment. The picture we have got is the tradi-

<sup>4</sup> The positive effect of  $n$  on the real wages is easily understood if we consider that the bargained nominal wage does not depend on it and that the price depends negatively on it.

tional one with market imperfections granting rents and benefits to incumbent workers and firms at the expense of overall production and employment.

#### **4. Firms' entry and the long run**

Firms entry is determined by profits and by entry costs. The idea is quite simple: high profits attract firms into the market while low profits discourage their entry. In fact, each potential entrant firm observes the profits and, given the entry costs, decides whether to enter or not. If the firm decides to enter it start immediately to operate and to gather the possible profits. We also imagine that firms leave the market according to a stochastic law, with each firm having a given probability to leave the market in each period. A possible explanation for this is that there is a possibility that the producing process of a firm becomes obsolete and it is forced to leave the market. Given this possibility we define  $s < 1$  has the probability that a firm survive to the next period.

The exact sequence of actions that take place in each period is the following: potential firms choose whether to enter the market, the production and selling take place and, finally, some firms leave the market.

The entry process depends on expected profits and we assume that potential firms have static expectations. In other words, we imagine that they look at the amount of profits earned by incumbent firms in the previous period and they base their decision on the assumption that future profits will stay constant<sup>5</sup> at that level.

In each period  $t$ , each potential firm may choose to enter by paying an entry cost  $Q$ : if it does so, it is entitled to gather the profits of current period and, as long as it survives, of following periods. Moreover, at the end of the period, each firm has a probability  $s$  to survive.

Entry costs depends on how many firms are entering and on some administrative fixed costs:

$$Q = C \frac{E_t}{n_{t-1}} + K \quad (18)$$

where  $E_t$  is the number of entrants,  $K$  are the fixed costs and  $C$  is a measure of the variable costs.

More specifically we can think to  $C$  as the cost component related to conges-

<sup>5</sup> In what we are arguing it would be the same if potential entrants assume that future profits will stay, on average, at the same level.

tion effects. When more firms try to enter at the same time, the resources needed to set-up the business become more demanded and more costly, so that costs rise proportionally. Even the bureaucratic procedures that firms have to go through are likely to be more of a burden when many firms are entering at the same time. The degree of congestion effect is normalized by the dimension of the market in the previous period, measured by  $n_{t-1}$ , because we believe that markets that are already large are less likely to suffer from congestion. On the contrary  $K$  are more likely to represent fixed administrative costs, either in the form of fixed fees they have to pay or of loss of time in bureaucratic procedures<sup>6</sup>.

This given, the expected present value of profits  $EPV_t$  for an entrant firm at time  $t$ , is given by

$$EPV_t = \sum_{m=0}^{\infty} (s\delta)^m \pi_{t+m}^e \quad (19)$$

where  $\pi^e$  are the expected profits and  $\delta$  is the discount rate. Given the static expectation of firms we can set the  $\pi_{t+m}^e$  to be equal to  $\pi_{t-1}$  for any  $m$  and solve the above obtaining:

$$EPV_t = \frac{\pi_{t-1}}{1-s\delta}. \quad (20)$$

If firms are risk-neutral, the no-arbitrage condition implies:

$$EPV_t = Q_t \quad (21)$$

In fact, when  $EPV_t > Q_t$  firms would keep entering the market and this would rise the entry cost according to (18), so that the above equality is reached again; if  $EPV_t < Q_t$  firms would not enter the market reducing the entrants, eventually, to zero. Obviously entry costs cannot be negative so that the number of entrants is never less than zero.

The above entry mechanism and the structure of costs allow us to build an equation that describes the entry process: combining (20) and (21) and exploiting (12) and (18) we have

$$\begin{cases} E_t = \left[ \frac{v}{(1+n_{t-1})^2} - k \right] n_{t-1} \text{ for } n_{t-1} \leq \sqrt{t/k} - 1 \\ E_t = 0 \text{ for } n_{t-1} > \sqrt{t/k} - 1 \end{cases} \quad (21a)$$

<sup>6</sup> Obviously part of the fixed costs may be due to technical reason, like the setting up of plants. This makes no difference in what we are arguing.

where  $v \equiv \frac{(1-W_t^n)^2}{(1-s\delta)C}$  and  $k \equiv K/C$ . The above equations simply tell us that entry depends on the difference between expected profits and entry costs. Since equation (21a) determines only the number of entrant it cannot be negative: in the event that expected profits are lower than the costs, no firms will enter and  $E_t=0$ . In other words, as long as expected profits cover the entry costs the firms keep entering the market but, as soon as this is not true, they stop. This mechanism generates a discontinuity in the entry process: we call  $n_E$  the number of firms for which such discontinuity begins:

$$n_E = \sqrt{v/k} - 1. \quad (21b)$$

In order to justify the presence of even only one firm in the markets we have to assume that  $v > k$ : this condition is necessary (but not sufficient) for profits to cover entry opportunity cost. In what follows we assume that  $v > k$ , as in any other case no firm would ever enter and the market would not even exist. In addition, we have also to imagine that at time zero there is already at least one firm operating, otherwise the static expectation hypothesis would not have any sense.

The number of firms  $X_t$  that leave the markets at the end of period  $t$  is easily obtained, in fact if  $s$  is the survival rate then:

$$X_t = (1-s)n_t \quad (22)$$

The number of firms operating in the market at any time  $t$  is then given by

$$n_t = n_{t-1} + X_{t-1} + E_t \quad (23)$$

which combined with (21a) and (22) gives

$$\begin{cases} n_t = \left[ s + \frac{v}{(1+n_{t-1})^2} - k \right] n_{t-1} & \text{for } n_{t-1} \leq n_E \\ n_t = sn_{t-1} & \text{for } n_{t-1} > n_E \end{cases} \quad (24)$$

The above equation describes the dynamic process of the number of firms and is a non-linear difference equation of first order: to make it more compact we may define such process as a function  $f()$  of the number present in the previous period, so that

$$n_t = f(n_{t-1}) \quad (24a)$$

Obviously  $f()$  takes the form described in (24). The dynamics of the process is represented in Figure 1.



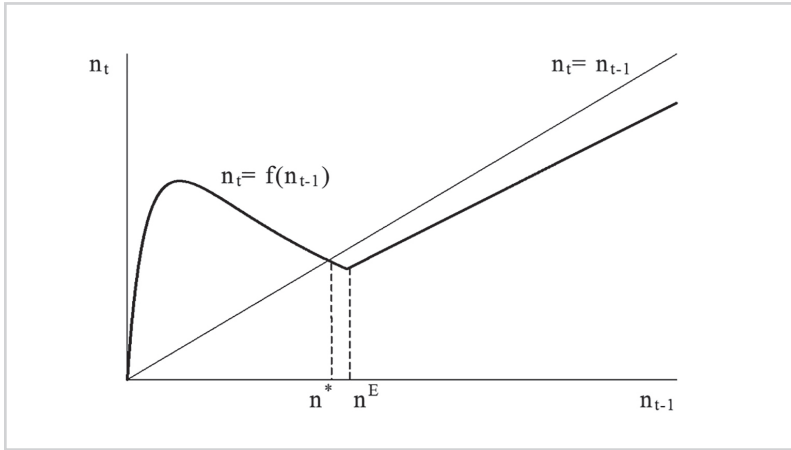


Figure 1. The dynamics of the number of firms

### Equilibrium, dynamics and stability

The equilibrium number of firms  $n^*$  is obtained for  $E_t = U_{t+1}$  (this in fact implies  $n_t = n_{t+1}$ ) so that

$$n^* = \sqrt{\frac{v}{1-s+k}} - 1. \quad (25)$$

Note that condition  $v > k$  guarantees that  $n^*$  is positive and comparing (21b) and (25) we necessarily have that  $n_E < n^*$ .

Equation (24) is a non-linear difference equation of first order. The solution of such equation cannot be derived analytically but we can study the local properties of equilibrium. Given that  $n^*$  is always smaller than  $n_E$ , the local properties of equilibrium can be completely assessed studying  $f(n_{t-1})$  for values lower than  $n_E$ . This said, the equilibrium is locally stable if  $|f'(n^*)| < 1$ : in our case we have

$$f'(n^*) = 1 - 2(1-s+k) \left( 1 - \sqrt{\frac{1-s+k}{v}} \right) \quad (26)$$

so that  $|f'(n^*)| < 1$  for

$$(1-s+k) \left( 1 - \sqrt{\frac{1-s+k}{v}} \right) < 1 \quad (27)$$

The solution of the above has not a straightforward interpretation, but we can show that the condition holds when one of the following is met

$$k < s$$

or

$$v < \left(1 + \frac{1}{k-s}\right) [(1-s)C + K] \quad (28)$$

This said, stability is not always met and when the above conditions do not hold the process may not converge.

### Global stability

We study now the global stability of the process. We start showing that, independently of the local property of the equilibrium, in the long run the number  $n_t$  is necessarily<sup>7</sup> confined in an interval which we call  $[n_A, n_B]$  (we depict this interval in Figure 2).

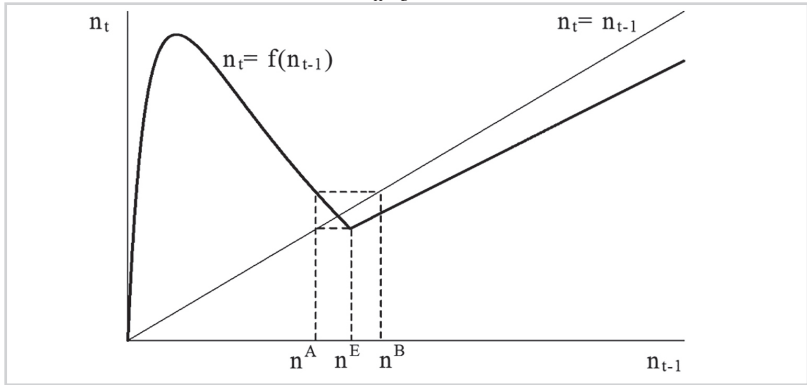


Figure 2. Long run interval for the number of firms

We demonstrate this in three steps:

- 1) for  $n_t < n^*$  we have  $f(n_t) > n_t$  so that as long as  $n_t \leq n^*$  the number of firms keep increasing and it will either reach  $n^*$  or overshoot it;
- 2) for  $n_t \geq n^*$  we have  $f(n_t) \leq n_t$  and  $\min f(n_t) = f(n_E)$  so that once  $n^*$  has been reached or overshoot, the number of firms will keep dropping, reaching a value between  $f(n_E)$  and  $n^*$ ; we define  $n_A \equiv f(n_E)$ .
- 3) if  $f(n_E) < n_t < n^*$  then  $n^* < f(n_t) < f(n_A)$  so that the number of firms can be at the most  $f(n_A) \equiv n_B$ .

<sup>7</sup> This is true only for  $n_0 > 0$ .

We can then assert that the number of firms either converge to  $n^*$  or it keeps staying inside the interval  $[n_A, n_B]$ .

### Range of oscillation

We have just said that the number of firms is, in the long run, confined in the interval  $[n_A, n_B]$ . It could be interesting to give an idea of how broad is this interval. To measure it, we choose the relative increment from the smaller value ( $n_A$ ) to the larger ( $n_B$ ): if we call  $R$  this relative increase we have

$$R = \frac{n_B - n_A}{n_A} = \frac{n_B}{n_A} - 1. \quad (29)$$

The value  $R$  basically represents the largest relative variation that we can observe in the long run number of firms. Alternative (but similar) measures are possible but we opted for this because it is algebraically quite simple, because it is a measure of a range in which  $n^*$  is comprised and because taking as reference point  $n_A$  it delivers a larger value, allowing us to assert that when  $R$  is small we can really be certain that the number of firms remain close to  $n^*$ .

If we insert the value of  $n_A$  and  $n_B$  in the equation<sup>8</sup> for  $R$  we have:

$$R = \frac{n_B}{n_A} - 1 = \frac{1}{\frac{1-s}{\sqrt{\nu}} + \frac{s}{\sqrt{k}}} - k - (1-s). \quad (30)$$

From the above equation it is clear that the range is an increasing function of  $\nu$  and  $k$  and a decreasing function of  $s$ : it thus follows that the dimension of  $R$  cannot be higher than the values that  $R$  assumes when, keeping all the others parameter constant,  $\nu$  tends to infinity. If we compute the limit of  $R$  for  $\nu$  that tends toward infinity we obtain

$$\lim_{\nu \rightarrow \infty} R = \frac{k}{s^2} - k - (1-s) = k \frac{1-s^2}{s^2} - (1-s) \quad (30a)$$

Equation (30a) tells us the highest relative change in the number of firms that we can observe in the long run, independently on the stability properties of the process. Since the interval  $[n_A, n_B]$  comprises  $n^*$  it follows that, even if the process is not stable, in the long run, the relative difference between the number of firms at any instant and the equilibrium value  $n^*$  cannot be higher<sup>9</sup> than  $R$ .

Now that we have specified this aspect, we discuss global stability.

<sup>8</sup> By definition we know that  $n_A = f(n_E)$  and  $n_B = f(n_A)$ .

<sup>9</sup> In fact the difference cannot be greater than  $R$  and it is, instead, smaller.

**Case I: global stability when the process is locally stable**

We start discussing the locally stable case, so that one of the condition expressed in (28) must hold. This means necessarily that  $|f'(n^*)| < 1$  and that  $n^*$  is, at least locally, an attractor. Two possibility may happen in this case, the first one is defined by the fact that  $0 \leq f'(n^*) < 1$ , the second is defined by the fact that  $-1 < f'(n^*) < 0$ .

We start discussing the first one. First of all it is easy to show that we have  $0 \leq f'(n^*) < 1$  for

$$k < s - \frac{1}{2}$$

or

$$v < \frac{(1-s)^3}{(1/2 - s - k)^2} \tag{31}$$

when this is the case we see that for any  $n_t < n^*$  we have  $f(n_t) \geq n_t$  and  $f(n_t) \leq n^*$ . The number of firms keep rising but it never overshoot  $n^*$ , then the number of firms necessarily converges to  $n^*$  and the process is globally stable.

The second case, when  $-1 < f'(n^*) < 0$ , is more problematic and, in effect, we cannot analytically check the global stability. However, we know that even in this case the number of firms will stay in the interval  $[n_A, n_B]$ . If the interval is small enough we may approximate the long run number of firms to  $n^*$  even without deriving global stability. For this reason we compute the highest possible value of  $R$  for different values of  $s$  and  $k$  when  $v$  goes to infinity. The values we obtain represent an upward threshold for the relative difference between  $n^*$  and any other value that  $n_t$  can assume in the long run. We report these in Table 1.

**Table 1: Highest possible values for R**

Values of k	Values of s						
	0.5	0.75	0.9	0.95	0.975	0.99	0.995
0.5	1	0.1388	0.01728	0.004	0.001	0.0002	0.0000
0.6		0.2167	0.0407	0.0148	0.0062	0.0022	0.001
0.7		0.2944	0.0642	0.0256	0.0113	0.0042	0.002
0.8		0.3722	0.0876	0.0364	0.0165	0.0062	0.003
0.9			0.1111	0.0472	0.0217	0.0083	0.004

The above table shows that for reasonable values of  $s$  (when  $s > 0.95$ ) the relative difference between the boundaries of the interval  $[n_A, n_B]$  is, at the very most, 4,72% (and usually much smaller). For this reason we believe that, for reasonable values of  $s$ , the long run number of firms can be approximate to  $n^*$ .

### Case II: locally unstable

In this case we already know that  $n^*$  is a repelling point. However we also know that the number of firms necessarily stays in the interval  $[n_A, n_B]$ : in practice, we observe oscillation in this interval with the number of firms never reaching a stable value. If the interval is small enough we can approximately state that the number of firms stays reasonably close to the stable value  $n^*$ : on the contrary, if the interval is big then the number of firms shows large oscillations. To assess this we compute the highest possible value of  $R$  for different value of  $s$  and  $k$  when  $v$  goes to infinity. We report the results in Table 2.

**Table 2: Highest possible values for R**

Values of k	Values of s						
	0.5	0.75	0.9	0.95	0.975	0.99	0.995
0.75	1.75	0.3333	0.0759	0.0310	0.0139	0.0052	0.0026
1	2.5	0.5278	0.1346	0.0583	0.0212	0.0103	0.0051
1.5	4	0.9167	0.2518	0.1120	0.0418	0.0204	0.0101
2	5.5	1.3056	0.3691	0.1661	0.0625	0.0306	0.0151
3	8.5	2.0833	0.6037	0.2651	0.1037	0.0591	0.0252
5	14.5	3.639	1.0728	0.4902	0.1862	0.0915	0.0454
25	74.5	19.194	5.7642	2.6508	1.0108	0.4976	0.2469

Whether the above results suggest that the number of firms keep staying near  $n^*$  depends on what we mean with the word “near”. However, we can state that for  $s > 0.975$  and for  $k < 3$  the difference between the two boundaries of the interval is, at the very most, 10% something we could reasonably consider small. Then for  $s > 0.975$  and for  $k < 3$  we believe we could approximate the instable case to the stable one. For other values of the parameter this would be unwise.

### **Concluding remarks on the equilibrium and its stability**

In the previous part we have shown that, depending on some key parameters, the process either converge to  $n^*$  or it is trapped in the interval  $[n_A, n_B]$ . The dimension of such interval is therefore extremely important: if it is small then the number of firms is always very close to  $n^*$  and can be approximated to it. This dimension depends on the values of the parameters  $\nu$ ,  $k$ ,  $s$  and we have found values of them for which the interval is sufficiently small. Obviously those threshold values are subjective to what can really be called sufficiently small.

## **5. Long run analysis**

As we have just seen, we can expect the number of firms to converge in the long run towards  $n^*$  or, for many other reasonable values of the parameters, to stay close to that value. We have then all the elements to evaluate the long run outcomes in terms of our key variables. We can also assess the effects that the degree of regulation of markets has, comparing the short run effects with the long run. Obviously this analysis is correct only if the parameters have such values that the long run number of firms is  $n^*$  or a value close to it. When this is not the case the number of firms keeps moving from values greater than  $n^*$  to value lower than it: in this sense  $n^*$  can still be considered a sort of average of the long run number of firms, but not a precise measure of it.

This said, if we consider  $n^*$  as the long run number of firms, we can compute the equilibrium values of the index of Lerner, profits, employment and real wages. To make notation simpler, we derive those variables as functions of the bargained wage  $W^b$  which depends only on exogenous parameters and is positively correlated to unions' bargaining power  $\beta$ .

We start from the index of Lerner, which measures the market power of firms. If we insert in the price-cost margin (11) the equilibrium value of firms  $n^*$  we have

$$\mu = \frac{1 - W^B}{1 + \frac{W^B}{a} \sqrt{\frac{(1 - W^B)^2}{(1 - s\delta)C} \frac{1}{1 - s + k}}} \quad (32)$$

so that the market power is tied in a positive way to the fixed costs  $K$  and the variable costs  $C$ . The above equation allows us to identify in the entry costs the ultimate source of firms market power and to tie it to something concrete and measurable: something on which a policy maker could, in effect, intervene.

The next step is to determine the long run profits of firms. If we insert  $n^*$  as in equation (25) in the equation (13) we have

$$\pi_i = (1 - s\delta) [(1 - s)C + K] \quad (33)$$

which tells us that profits depends positively on the entry cost (on both  $K$  and  $C$ ) and, surprisingly, are independent on the bargaining power.

The first fact is quite obvious, higher entry cost discourages entry and reduce the number of firms in the market: this increases the market power and rises the profits. Note that also the variable component of entry costs influences the level of profits, however, given the fact that is multiplied by  $(1-s)$  its effects is, in most of the cases, small.

The absence of the bargaining power in the equation of profits may at first seems surprising. The truth however is that the effects of the bargaining power is two- folds: on one side, it reduces the share of revenues that go to the firms but in doing this, it also discourages entry reducing the number of firms and rises the profits. In the end the two effects cancel each other out.

In other words the bargaining power effectively reduces the degree of competition in the product markets so that its effects on profits is not necessarily negative and, actually, is neutral.

Finally, we derive the long run aggregate employment  $L^{LR}$  (combining short run employment from 16 with 25)

$$L^{LR} = \frac{1}{a} \left\{ \left( 1 - \frac{W^B}{a} \right) - \sqrt{(1 - s\delta) [(1 - s)C + K]} \right\} \quad (34)$$

and the long run real wage ( $W^{LR}/P$ ) is given by (combining short run real wage 17 with 25)

$$\frac{W^{LR}}{P} = \frac{1}{a} \left\{ \frac{\sqrt{(1-s\delta)[(1-s)C+K]}}{W^B} + \frac{1}{a} \right\}. \quad (35)$$

The above results allow us to make two assertions: first, in the long run bargaining power maintains its short run effects, it reduces employment and increases real wages; second, entry costs decreases both employment and real wages (with variable entry costs having a small role).

The mechanism that makes this happen is still the same: entry costs and bargaining power discourage entry and reduce competition. In this, the effect of the bargaining power of union is two-folds: it reduce the equilibrium number of firms (rising the aggregate price) but it increase the nominal bargaining wage. The latter effect however is stronger than the former so that real wages are in the end positively related to the bargaining power.

## 6. A comparison of deregulation policies in the short and long run

Now that we have examined both the short and the long run we can compare how the deregulation of markets affects employment and real wages in the two different time horizons. When we refer to the long run we use the results we obtained for the stable case. We have seen that these results are reasonably similar to the other cases for many values of the parameters: however, they may differ when the parameters assumes some extreme values. We summarize now the effects that (de)regulation of labour and product markets have in the short and long run and we search for the existence of a (de)regulation policy mix that, affecting both markets, could improve the working of the economy without causing a loss to any economic agent.

For simplicity we report only the effects of deregulation policies: in the labour market this would happen through a legislation that reduces the bargaining power (a stricter law on strikes for example) while in the product market it would mean a reduction of the entry costs (in the fixed or variable component). While we do not describe directly regulation policies, their effects would simply be the opposite than those of deregulation. Table 3 presents the effects that a deregulation of labour market (a decrease in the union bargaining power  $\beta$ ) has on the key variables.



<b>Table 3: The effect of a deregulation of labour market</b>		
	Short Run	Long Run
Employment	Increase	Increase
Real Wages	Decrease	Decrease
Profits	Increase	None

Table 4 does the same, showing the effects of a deregulation of the product market, which could be brought forth with a reduction of the fixed costs  $K$  or the variable costs  $C$ .

<b>Table 4: The effect of a Deregulation of product market</b>		
	Short Run	Long Run
Employment	None	Increase
Real Wages	None	Increase
Profits	None	Decrease

The above tables suggest that the same deregulation policies may produce different effects in the different time horizons. The following assertions seem to be particularly relevant:

- In the long run, a combination of deregulation in both the markets increases employment but necessarily reduces both real wages and profits: we can then assert that deregulation policies on both markets are not able, on their own, to fully compensate the drawbacks that each one brings.
- A deregulation policy on both markets however, produces different effects at different time horizons and the loss of a part in the short run may be compensated by its gain in the long run. Only if the parts find this intertemporal trade-off beneficial the deregulation of both markets seems to bring benefit to all the economic agents. Interestingly, in this case incumbent firms would have a gain in the short run and a loss in the long run, while the positive and negative effects for workers would have the opposite timing.
- Since usually economic agents bestow a higher value to the short than to the long run<sup>10</sup>, the above suggests that the intensity of deregulation in the labour market should be lower than the one in the product market.
- While in principle the same trade-off could be possible through a regula-

<sup>10</sup> This simply means that they have a positive intertemporal substitution rate.

tion of both markets, this would actually happen at the cost of the reduction of employment in both the short and long run.

- As a general fact, (de)regulation policies that affect the product markets do not have any effect in the short run. However, if a policy reduces the variable parts of the entry costs then it would be possible to accelerate the achievement of the long run equilibrium.

To summarize, it seems that deregulation policies generate intertemporal trade-offs, offering better or worse economic conditions depending on which is the time horizons we consider: exploiting this trade-off could improve the working of the economy but only if workers accept to pay some short run costs to obtain long run benefits. This fact induces to suggest that deregulation policies should be stronger in the product market than in the labour market.

## **7. Conclusion**

The present paper builds a macro model and uses it to explore three main aspects: i) the role of imperfect competition in the short run, ii) the process of firms' entry and its dynamics properties and iii) the effects of (de)regulation policies in the short and long run. Our analysis on the short run confirms the standard view, where less competitive markets reduce output and employment but increase rents, allowing for higher profits (when the product market is more concentrated) and higher real wages (when unions detain higher bargaining power). We then introduce firms' entry and the analysis of this process allows us to determine the equilibrium number of firms and to study its properties. The stability of this process proved to be troublesome and we have shown that for some values of the parameters it may not converge to a stable value. However, we have been able to show that even when the process does not converge, the number of firms stays confined in an interval which for reasonable values of the parameters is rather small (and contains the equilibrium value), so that in the long run we can approximate the number of firms with its equilibrium value.

The entry mechanism and the equilibrium value we found allow us to endogenize the number of firms and to explore the effects of imperfect competition in the long run. These effects are in some case different from the short run and, in particular, we have showed that in the long run the bargaining power does not have any effect on the level of profits.

Finally, we derived the effect of deregulation policies and we explored the existence of complementarity in the deregulation of these markets: while there are no complementarities if we consider a time horizon only, some intertem-

poral complementarities do arise. In effect, the deregulation of both markets could improve the overall working of the economy but it would induce a gain for firms and a loss for workers in the short run and exactly the opposite in the long run. The loss in a time horizon could be compensated by the gain in the other so that we observe an intertemporal trade-off. Since usually short run is more valued than the long run, deregulation policies should probably be stronger in the product market than in the labour market. The present paper could also be extended to allow for different assumptions and to analyse their consequences. A possible development could be the introduction of a more formal bargaining model and the use of a Nash maximand: this would probably not change the results and the policy implications but would provide more formal treatment of the subject and could encompass new elements of analysis. Finally, different assumptions on firms' expectations should be explored: while they would probably provide similar results in terms of the long run equilibrium, they could provide relevant differences in the convergence process.

## References

- Blanchard O. J., Giavazzi F. (2003), "Macroeconomic effects of regulation and deregulation in goods and labor markets", *Quarterly Journal of Economics*, 118, 879-907.
- Corsini L. (2009), "Optimal Degree of Regulation in an Unionised Economy", *Economia Politica - The Journal of Analytical and Institutional Economics*, 26, 105-120.
- Das S., Das S. (1997), "Dynamics of entry and exit of firms in the presence of entry adjustment costs", *International Journal of Industrial Organization*, 15, 217-241.
- Datta B., Dixon H. (2002), "Technological Change, Entry, and Stock-Market Dynamics: An Analysis of Transition in a Monopolistic Industry", *American Economic Review*, 92, 231-235.
- Dixon H. (1988), *Oligopoly theory made simple*, in S. Davies et al (eds), *Economics of Industrial Organization*, Harlow, Longman.
- Gaskins D. (1971), "Dynamic limit pricing: optimal pricing rule under the threat of entry", *Journal of Economic Theory*, 2, 306-322.
- Gilber, R., Vives X. (1986), "Entry deterrence and the free rider problem", *Review of Economic Studies*, 53, 71-83.
- Lerner A. P. (1934), "The concept of monopoly and the measurement of monopoly power", *Review of Economic Studies*, 1, 157-175.
- Modigliani F. (1958), "New developments on the oligopoly front", *Journal of Political Economy*, 66, 215-232.
- Sylos-Labini P. (1962), *Oligopoly and Technical Progress*, Cambridge, Mass., Harvard University press.
- Spector D., (2004), "Competition and the capital-labor conflict", *European Economic Review*, 48(1), 25-38.



# Price Competition in an International Mixed Market

ALESSANDRA CHIRCO AND MARCELLA SCRIMITORE

## 1. Introduction

In this paper we assume that a public (welfare-maximizing) firm competes on the home market against a number of domestic and foreign private (profit-maximizing) firms. Such firms compete over prices, sharing the same linear technology and offering imperfectly substitutable products. A mixed oligopoly is the reference framework we develop in a single-country perspective in order to capture the extent to which competition exerted by the foreign firms affects domestic welfare. These assumptions are consistent with recent competitive patterns observed on the international scenarios. In these, the presence of state-owned firms, still massive on the domestic markets, is increasing despite the worldwide waves of privatization that have taken place over the past twenty years. Public ownership, which is commonly observed in network industries like telecommunications, transports, energy and utilities, characterizes also a range of services like insurances and banking, postal services, health care and education. Private and public firms frequently coexist in these sectors, which are moreover becoming increasingly exposed to international competition, in response to the international liberalizations and demand growth. A characterization of the equilibria arising in such a scenario is the basis for an assessment of the welfare effects caused by changes in the market structure, such as cross-border acquisitions and privatization policy. Our main result is that, under price competition and differentiated products, the presence of international competitors in the domestic market induces an overall convergence to efficient pricing, and is consistent with binding budget constraints of the public firm

even under constant average and marginal costs. This also holds under Stackelberg price competition, provided that the presence of foreign competitors is not too large. The paper is organized as follows. Section 2 puts our results in context by sketching the main relevant related literature. The model is then developed in Sections 3, while Section 4 gathers some conclusions.

## **2. The related literature**

Mixed oligopoly theory has remarkably developed in the last two decades by pointing out those situations in which interactions between private and public firms lead to a higher social desirability as compared to a fully privatized context, thus contributing to the debate on privatization. Efficiency, strategic and political arguments have been invoked to provide a theoretical support to the idea that mixed public-private markets can dominate the alternative of pure private markets. The most commonly adopted framework is one in which public ownership is associated with pure welfare maximization objectives,<sup>1</sup> and firms compete over quantities under increasing marginal costs. In a framework of strategic interactions, a number of works attributes the welfare-enhancing character of a mixed market to the incentives of the public firm to expand total output, thus indirectly regulating markets. Indeed, under quantity competition a publicly owned firm produces the quantity at which the clearing price equals the marginal cost; this implies a greater aggregate market production than that observed in a market with only profit-concerned firms, and therefore a higher efficiency. In this context, however, De Fraja and Delbono (1989) make an important warning about the role of increasing marginal costs: the presence of this high producing firm may impact negatively on welfare due to its higher average and marginal costs. While welfare is positively affected by the boost to production, it is negatively affected, under quadratic costs, by an unequal sharing of production between the public and the private firms.<sup>2</sup>

<sup>1</sup> Some authors assume that public firms maximize a weighted average of welfare and their own profits, thus tackling partial privatization. The search for the optimal degree of government ownership in the privatized firms is a core issue of this literature. See Matsumura (1998) as a major reference.

<sup>2</sup> In this case, though public and private firms share the same technology, at equilibrium the public firm is endogenously less efficient, the efficiency gap basically reflecting the differences in firms' objectives.

The spread in the produced quantities and the associated welfare losses are higher, the larger is the number of private firms in the market. In these circumstances privatization may lead to welfare improvements. In the context of mixed oligopolies, privatization has to be positively considered also when it induces a firm's restructuring which improves the overall efficiency and productivity. Many works focus on these cost-saving reasons which justify a change of ownership from public to private, on the assumption, often derived from a conventional wisdom, that public firms are less efficient.<sup>3</sup> As pointed out by De Fraja (1991), in these circumstances the beneficial effects of privatization crucially depend on the size of the technological efficiency gains, which must be high enough to offset the welfare loss due to the firm's output contraction once the privatization is realized. A rationale for public ownership has also been found in its being an instrument to achieve political, social, industrial and environmental goals. White (2002) shows that governments can strategically manipulate the public firms' objective functions in order to disguise their real political orientation and to actually pursue aims which differ from those publicly stated.<sup>4</sup> Moreover, the presence of public firms has been considered socially beneficial when in the bargaining process they allow for higher wages (Willner, 1999), or when their positive impact on social welfare is through their contribution to investment in R&D (Poyago-Theotoky, 1998). Finally, public more than private ownership is to be invoked when governments pursue an environmental policy: public firms may internalize their environmental damages and ensure higher revenues from environmental taxes (Bárcena-Ruiz and Garzón, 2006). The recent extension of the analysis of mixed oligopoly to an international framework has raised several issues,

<sup>3</sup> Managerial slackness and higher agency costs are often invoked as reasons for the lower efficiency attributed to public companies. Willner (1999) points to the higher wages paid under public ownership as an explanation for lower cost efficiency. This belief, however, has been often challenged and does not receive unanimous consensus, neither in the theoretical nor in the empirical literature (see Björkroth et al. (2006), p.180, and the papers referred therein).

<sup>4</sup> Most of economic literature, however, looks at the political interference within state-owned enterprises as a reason for their privatization. By assigning external objectives to public firms, Estrin and Perotin (1991) show how politician may contribute to amplify the agency problem at the firm or state level, and to weaken the capital market pressure which would motivate public managers towards efficiency. Public enterprises are also viewed by Boycko et al. (1996) as a means to pursue political objectives, such as excess employment; also in this case privatization is invoked for the efficiency gains it may generate in underperforming firms.

which are of interest for both industrial organization and international trade theory: from the impact of privatizations in international markets to the effectiveness of open-door policies and cross-borders acquisitions. Several papers analyse international competition in a single-country mixed market. Among these, Fjell and Pal (1996) extend the model by De Fraia and Delbono to allow for the competition of foreign private firms in addition to the domestic ones, while Pal and White (1998) analyse privatizations in the presence of subsidies or tariffs. More recent works explore competition in international mixed markets using a two-country approach in which interactions involve self-interested governments. In this line, Dadpay and Heywood (2006) offer an analysis of the equilibria under quantity competition and decreasing returns to scale, showing that welfare gains are typically associated to coordinated privatization, though the governments' self-interest does not provide the appropriate incentives to unilaterally privatize. Using a similar framework, Han and Ogasawa (2008) examine the optimal extent of privatization, while Bárcena-Ruiz and Garzón (2005) develop a setting with asymmetric constant marginal costs where the decision of one government to privatize depend on the relative cost advantage of the private firms over the publicly-owned firm. With the exception of the duopoly model by Ohnishi (2010), Matsumura et al. (2009) and Matsumura and Matsushima (2012),<sup>5</sup> the analysis of international mixed oligopolies has been developed under quantity competition in a homogeneous product market. This calls for the additional assumption of decreasing returns to scale, constant returns being inconsistent in that framework with a non-negative profit condition for the public firm.

This paper contributes to the existing literature by developing a general international mixed-oligopoly model under price-competition and imperfect product substitutability. The main theoretical advantage of this set-up is that price competition, by enlarging significantly the set of market configurations in which the public firms' budget constraint is consistent with constant average and marginal costs, allows us to rule out any exogenous or endogenous technological asymmetry between public and private firms, so that the properties of equilibria and their policy implications rely exclusively on the characteristics of firms' strategic interactions.

<sup>5</sup> Among the studies dealing with price competition on international markets, Ohnishi (2010) includes an analysis in single-country perspective, which is however confined to a duopoly. An analysis carried out in a two-country perspective is conversely found in Matsumura and Matsushima (2012), while a monopolistic competition model with foreign competitors is in Matsumura et al. (2009).



### 3. The model

We consider a market for a differentiated product in which  $S$  varieties are produced. The demand accruing to this market stems from the home country only, while on the supply side a public domestic firm, called firm  $i$ , interacts with  $m$  domestic and  $n$  foreign private firms. In the sequel, the set of domestic private firms will be denoted by  $\{H\}$  and its generic element by  $h$ , the set of foreign private firms by  $\{F\}$  and its generic element by  $f$ . Since each firm produces one variety,  $m+n+l=S$ .

The representative domestic consumer exhibits the following semi-linear quadratic preferences:

where  $q_0$  is a composite good produced in a perfectly competitive market,

$$U(q) = \sum_{s=1}^S q_s - \frac{1}{2} \left( (1-\gamma) \left( \sum_{s=1}^S q_s^2 \right) + \gamma \left( \sum_{s=1}^S q_s \right)^2 \right) + q_0$$

which absorbs all income effects of price changes. Maximization of  $U(q)$  yields a linear direct demand function for the generic variety  $s$ :

$$q_s^d = \frac{1-\gamma - (1+\gamma(m+n-1))p_s + \gamma\bar{P}_s}{(1-\gamma)(1+\gamma(m+n))} \quad (1)$$

where  $\bar{P} = \sum_{v \neq s} p_v$  is the sum of the prices of all the other varieties and  $g$  is the degree of product substitutability, ranging from 0 (absence of substitutability) to 1 (homogeneous products).<sup>6</sup>

As standard in mixed oligopoly models, the public firm maximizes social welfare, while all private firms are profit-maximizing. In a single-country framework, the social welfare is defined as the sum of consumer surplus (CS) over all the  $m+n+l$  varieties, and the aggregate profits of domestic firms (i.e., the sum of the profits of the public firm  $\pi_i$ , and those of the domestic private firms,  $\Pi_h = \sum_{h \in \{H\}} \pi_h$ ): obviously the profits of private foreign firms do not enter the public firm's objective function. Firms compete simultaneously over prices.

Let us consider the optimal behaviour of the public firm. It solves the following problem:

$$\max_{p_i} (CS + \pi_i + \Pi_h)$$

<sup>6</sup> Our analysis is also robust to a model specification with the Shubik and Levitan (1980) demand function, which embodies product substitutability under the hypothesis that the market size is independent of the number of varieties and the degree of product substitutability.

where the consumer surplus:

$$CS = \frac{1}{2} \left( (1-\gamma) \left( q_i^2 + \sum_{h \in \{H\}} q_h^2 + \sum_{f \in \{F\}} q_f^2 \right) + \gamma \left( q_i + \sum_{h \in \{H\}} q_h + \sum_{f \in \{F\}} q_f \right)^2 \right)$$

can be expressed in terms of prices by using (1) for all domestic and foreign varieties. The solution yields the best reply of the public firm as a function of the rivals' prices:

$$p_i = \frac{c(1+\gamma(n-1)) + \gamma \sum_{h \in \{H\}} p_h}{1 + \gamma(m+n-1)} \quad (2)$$

Notice that this reaction function exhibits the standard strategic complementarity of price decisions. However, this occurs with respect to the private domestic firms only: the public firm's price is strategically independent of the price of the foreign private firms, notwithstanding the fact that the latter enters the public firm's objective function through both the consumers' surplus and the aggregate domestic profits. Under constant average and marginal costs, the foreign rivals' behaviour affects the impact of a marginal change of the public firm's price on the consumers' surplus with the same magnitude but opposite sign as it affects the marginal impact of the public firm on aggregate domestic profits. In the extreme case in which the public firm interacts with foreign firms only ( $m=0$ ), its optimal reaction is to set a price equal to marginal cost, independently of  $n$  and independently of the prices set by the rivals.<sup>7</sup>

As far as private domestic firms are concerned, for the generic firm  $h$  profit maximization yields the following reaction function:

$$p_h = \frac{1+c + \gamma(c(m+n) - (1+c)) + \gamma \left( p_i + \sum_{k \in \{H-h\}} p_k + \sum_{f \in \{F\}} p_f \right)}{2(1 + \gamma(m+n-1))} \quad (3)$$

where  $\sum_{k \in \{H-h\}} p_k$  denotes the sum of the prices of the private domestic firms other than  $h$ . Similarly, the reaction function of the generic private foreign firm  $f$  is:

<sup>7</sup> This result extends to a general oligopoly setting the mixed-duopoly model by Ohnishi (2010).

$$p_f = \frac{1+c+\gamma(c(m+n)-(1+c))+\gamma\left(p_i + \sum_{h \in \{H\}} p_h + \sum_{g \in \{F-f\}} p_g\right)}{2(1+\gamma(m+n-1))} \quad (4)$$

where  $\sum_{g \in \{F-f\}} p_g$  denotes the sum of the prices of the private foreign firms other than  $f$ .

Aggregating (3) over  $h$  and (4) over  $f$ , we get:

$$\sum_{h \in \{H\}} p_h = \frac{m\left(1+c+\gamma(c(m+n)-(1+c))+\gamma\left(p_i + \sum_{f \in \{F\}} p_f\right)\right)}{2+\gamma(m+2n-1)} \quad (5)$$

$$\sum_{f \in \{F\}} p_f = \frac{n\left(1+c+\gamma(c(m+n)-(1+c))+\gamma\left(p_i + \sum_{h \in \{H\}} p_h\right)\right)}{2+\gamma(2m+n-1)} \quad (6)$$

Equations (5)-(6) and (2) can be solved simultaneously for  $p_i$ ,  $\sum_{h \in \{H\}} p_h$  and  $\sum_{f \in \{F\}} p_f$ . By substituting the solutions into (3) and (4) and recalling the definitions of  $\sum_{k \in \{H-h\}} p_k$  and  $\sum_{g \in \{F-f\}} p_g$ , we obtain the following Bertrand-Nash equilibrium prices:

$$p_i^{BN} = \frac{c\gamma m(\gamma m + 2\gamma(n-1) + 2) + n\gamma c((n-2)\gamma + 3) + c(1-\gamma)(2-\gamma) + \gamma m(1-\gamma)}{\gamma^2 m(m+2n) + n\gamma((n-2)\gamma + 3) + (1-\gamma)(\gamma(3m-1) + 2)} \quad (7)$$

$$p_z^{BN} = \frac{(1-\gamma)^2 + c(1-\gamma) + c\gamma m(\gamma m + 2\gamma(n-1) + 2) + \gamma m(1-\gamma) + \gamma n(1-\gamma) + n\gamma c(\gamma(n-1) + 2)}{\gamma^2 m(m+2n) + n\gamma((n-2)\gamma + 3) + (1-\gamma)(\gamma(3m-1) + 2)} \quad (8)$$

with  $z=h,f$ .

Inspection of eqts (7) and (8) and simple comparative statics exercises allow to establish the following proposition.

**Proposition 1** *In a single-country mixed oligopoly with price competition, the public firm sets a price  $p_i^{BN}$  such that the mark-up over the marginal cost is decreasing in the share of foreign firms, and converges to zero when all private firms are foreign. The price of the public firm is lower than the price of the private firms – the difference  $(p_z^{BN} - p_i^{BN})$  being decreasing in  $g$ .*

In order to understand the interplay between the mode of competition (price-setting) and market internationalization in this single-market mixed oligopoly, it may be useful to disentangle their effects and analyse their implications separately. By setting  $n=0$  in (7)-(8), our solution confirms the result by Anderson et al. (1997) and Ghosh and Mitra (2010) that in a closed-economy framework a public firm competing over prices sets a price higher than

the marginal cost, though lower than that of its private rivals. This is in sharp contrast with the behaviour we would observe under quantity competition. In a closed-economy and under the same preferences and technological assumptions, in a quantity setting framework the public firm would react to whatever quantities produced by the private firms, by producing the amount of its own variety at which its market clearing price equals the marginal cost: a marginal increase in the public firm's production would not affect, for given quantities of the rivals, the contribution of the private firms to welfare, while it would increase the specific contribution of the public firm – i.e.,  $q_i^2/2 + \pi_i$  – so long as its clearing price exceeds the marginal cost. Under price competition, for given choices of the rivals, a price reduction by the public firm reduces the marginal contribution of the private firms to welfare, and this mitigates its incentive to price at marginal cost. Aggressiveness in prices is more detrimental for the private firms contribution to welfare than aggressiveness in quantities, and this explains why we do not observe efficient pricing of the public firm under price competition. Strategic complementarity implies that the differences between public and private prices are lower under price competition.

The existence of foreign firms obviously makes for a stronger aggressiveness of the public firm both under quantity and under price competition. Foreign firms contribute to the public firm's objective function only through the consumer surplus, so that the negative effect of a reduction of  $p_i$  on their profits is neglected when domestic welfare is maximized. However,  $p_i$  converges to the marginal cost only when no domestic private firms are active in the market. The relative weight of foreign firms in satisfying consumers' demand exerts a marked downward pressure on prices, but this effect of market internationalization does not rely on the competition among private firms; rather, it derives from the objectives pursued by the public firm. Notice that under quantity competition and constant marginal costs, market openness would be inconsistent with a non-negative profit constraint of the public firm.<sup>8</sup>

The properties of the above equilibrium have clear consequences in terms of the welfare evaluation of cross-border acquisitions and privatization policies. In particular, the following propositions can be established.

<sup>8</sup> In the presence of foreign private firms, a marginal increase in the production of the public firm would raise the contribution of private firms to welfare, thus inducing the former to expand its production beyond the zero-profit threshold. Notice that our result is consistent with the findings of Fjell and Pal (1996) in a Cournot setting with homogenous product and convex costs.

**Proposition 2** *Regardless of  $\gamma$ , the acquisition of a foreign firm by the domestic country implies a decrease in the domestic CS and an increase in the domestic welfare. The welfare gain is higher, the lower the overall number of private firms, and the higher  $m$  with respect to  $n$ .*

**Proposition 3** *Regardless of  $\gamma$ , the privatization of public firm is detrimental for both CS and domestic welfare. The welfare losses are higher, the lower the overall number of private firms and the higher  $n$  with respect to  $m$ .*

Both propositions can be verified computationally. However, they have both intuitive explanations. If the domestic country acquires a foreign firm, the domestic CS decreases, due to a generalized increase in prices. As far as welfare is concerned, since ex-post it includes the profits of the acquired firm, the increase in overall profits outweighs the decrease in the consumers' surplus. The welfare gain is therefore higher, the higher the private firm's profits. The latter are high when the overall number of firms is low and – according to Proposition 1 – when the number of foreign firms is low with respect to that of domestic firms. If the public firm is privatized, both CS and domestic welfare are negatively affected due to the generalized increase in price. Clearly, the welfare losses are higher, the higher the impact of the public firm on welfare, i.e., when the overall number of private firms is low enough to prevent that a fully private market be anyway close to a competitive outcome, and when the number of foreign firms is relatively high so that the share of private profits on domestic welfare is low.

### 3.1. The sequential game

The above discussion should have clarified that the properties of equilibria under price competition, and the differences with the corresponding outcome under quantity competition, rely upon the very basic fact that in a simultaneous game the public firm strategy is defined for given choices of its private rivals. Under quantity competition, the existence of a public firm implies a reduction of the allocative inefficiency on the given quantities sold by the private firms; under price competition it reduces the demand faced by its private rivals at their given prices, but cannot affect the price over cost margin at which these quantities are sold. Strategic complementarity and substitutability do not play a relevant role in defining the key features of equilibria in simultaneous moves. On the contrary, the nature of strategic interaction becomes extremely relevant when we extend the above set-up to allow for price competition in a sequential game, with the public firm in the role of first-mover and with the  $m$  domestic and the  $n$  foreign private firms in the role of followers. Without any loss of generality, in this analysis we assume  $c=0$ .

Following the usual solution procedure of this Stackelberg game, we obtain the following equilibrium prices:

$$p_i^{SN} = \frac{(1-\gamma)(\gamma m(1+\gamma(m-1))) - \gamma n(1+\gamma(n-1))}{\gamma^3(m+n)^3 + \gamma^2 n^2(5-3\gamma) + 2\gamma^2 mn(5-4\gamma) + (1-\gamma)(\gamma m(5\gamma(m-1)+8) + \gamma n(8-3\gamma) + (2-\gamma)^2)}$$

$$p_z^{SN} = \frac{(1-\gamma)((1-\gamma)(2-\gamma) + \gamma(\gamma(m^2+n^2) + 3(m+n)(1-\gamma)) + 2\gamma^2 mn)}{\gamma^3(m+n)^3 + \gamma^2 n^2(5-3\gamma) + 2\gamma^2 mn(5-4\gamma) + (1-\gamma)(\gamma m(5\gamma(m-1)+8) + \gamma n(8-3\gamma) + (2-\gamma)^2)}$$

with  $z=h, f$ . The main implications of this solution are gathered in the following proposition.

**Proposition 4** *In a single-country mixed-oligopoly with price competition and sequential moves with public leadership, for any  $\gamma, m, n$ , we get  $p_i^{SN} < p_z^{SN}$ ; moreover,  $p_i^{SN} \leq p_i^{BN}$  and  $p_z^{SN} \leq p_z^{BN}$ . As in the simultaneous case,  $p_i^{SN}$  is decreasing in the share of foreign firms, but it equals the marginal cost when  $m=n$ .*

Sequentiality of price decisions with the public firm in the role of the leader would create a more competitive and welfare enhancing environment even in a closed-economy framework (Bárcena-Ruiz, 2007). The public firm anticipates that a reduction of its own price will be followed by a reduction of the unit profit margin of the private firms, due to strategic complementarity; this lowers the marginal negative impact on private profits of the public firm's price decrease, allowing for  $p_i^{SN}$  to be closer than  $p_i^{BN}$  to marginal cost. The same positive effect on welfare would be observed under quantity competition, though in that case it would be associated, due to strategic substitutability, to a decrease in the production of the public firm (the price of which would exceed the marginal cost) and an increase in that of the private firms.<sup>9</sup> The presence of foreign firms causes a further decrease of the price set by the public firm at equilibrium, along the same lines described for the simultaneous game; again, market openness creates an further incentive towards a generalized price reduction. The more aggressive attitude shown by the public firm when it takes the role of the market price leader is such that the latter behaves like a public monopolist when faced with an equal number of foreign

<sup>9</sup> This occurs because in a sequential quantity-setting characterized by strategic substitutability, the public firm leader perceives a trade-off between its own contribution to welfare and that of the rivals. This result relies on increasing marginal costs if product homogeneity is assumed (Ejell and Heywood, 2002), but it is also observed under constant marginal costs if we allow for product differentiation - case in which affects the extent to which the simultaneous and sequential solution differ.

and domestic private competitors: only a market with  $n < m$  is consistent with a positive price-over-cost margin, while a market where foreign firms play a dominant role is one in which a public price leader suffers from profit losses, and is therefore inconsistent with a budget-balancing policy.

## 4. Concluding remarks

This paper provides an analysis of simultaneous and sequential price competition with imperfect product substitutability in an international mixed market. By exploring the properties of equilibria in a single-country model, we confirm the role of market regulators a public firm plays on a market populated by domestic and foreign firms. Indeed, we have shown that the public firm always succeeds in enforcing internal market discipline by inducing all private firms to keep lower prices and by reacting to international competition with further beneficial price reductions, regardless of the degree of product substitutability and of whether firms move simultaneously or with the public firm in the role of leader. Moreover, the paper highlights how privatizations are welfare detrimental in such a context and how policies of market openness and acquisitions of foreign firms should be conversely assigned priority to achieve welfare gains.

## References

- Anderson S.P., De Palma A., Thisse J.-F. (1997), "Privatization and efficiency in a differentiated industry", *European Economic Review*, 41, 1635-1654.
- Bárcena-Ruiz J.C., Garzón M.B. (2005), "International Trade and Strategic Privatization". *Review of Development Economics*, 9, 502-513.
- Bárcena-Ruiz J.C., Garzón M.B. (2006), "Mixed Oligopoly and Environmental Policy", *Spanish Economic Review*, 8, 139-160.
- Bárcena-Ruiz J.C. (2007), "Endogenous Timing in a Mixed Duopoly: Price Competition", *Journal of Economics*, 91, 263-272.
- Björkroth T., Grönblom, S., Willner, J. (2006), *Liberalisation and Regulation of Public Utility Sectors: Theories and Practice*. In: Bianchi, P., Labory, S. (eds): *International Handbook on Industrial Policy*. Cheltenham: Edward Elgar, 180-197.
- Boycko M., Schleifer A., Vishny R.W. (1996), "A Theory of Privatisation", *Economic Journal*, 106, 309-19.

- Dadpay A., Heywood J.S. (2006), "Mixed oligopoly in a single international market", *Australian Economic Papers*, 45, 269-280.
- De Fraja G. (1991), "Efficiency and privatization in imperfectly competitive industries", *Journal of Industrial Economics*, 39, 311-321.
- De Fraja G., Delbono F. (1989), "Alternative Strategies of a Public Enterprise in Oligopoly", *Oxford Economic Papers*, 41, 302-311.
- Estrin S., Pérotin V. (1991), "Does Ownership Always Matter?" *International Journal of Industrial Organization*, 9, 55-71.
- Fjell K., Pal D. (1996), "A Mixed Oligopoly in the Presence of Foreign Private Firms", *The Canadian Journal of Economics*, 29, 737-743.
- Fjell K., Heywood J.S. (2002), "Public Stackelberg leadership in a mixed oligopoly with foreign firms", *The Australian Economic Papers*, 41, 267-281.
- Ghosh A., Mitra M. (2010), "Comparing Bertrand and Cournot in Mixed Markets", *Economics Letters*, 109, 72-74.
- Han L., Ogawa H. (2008), "Economic Integration and Strategic Privatization in an International Mixed Oligopoly", *FinanzArchiv / Public Finance Analysis*, 64, 352-363.
- Matsumura T. (1998), "Partial privatization in mixed duopoly", *Journal of Public Economics*, 70, 473-483.
- Matsumura T., Matsushima N. (2012), "Airport Privatization and International Competition", *The Japanese Economic Review*, 63, 431-450.
- Matsumura T., Matsushima N., Ishibashi I. (2009), "Privatization and entries of foreign enterprises in a differentiated industry", *Journal of Economics*, 98, 203-219.
- Ohnishi K. (2010), "Domestic and international mixed models with price competition", *International Review of Economics*, 57, 1-7.
- Pal D., White M.D. (1998), "Mixed Oligopoly, Privatization, and Strategic Trade Policy", *Southern Economic Journal*, 65, 264-281.
- Poyago-Theotoky J. (1998), "R&D Competition in a Mixed Duopoly under Uncertainty and Easy Imitation", *Journal of Comparative Economics*, 26, 415-428.
- Shubik M., Levitan, R. (1980), *Market Structure and Behavior*, Cambridge, MA: Harvard University Press.
- White M. D. (2002), "Political manipulation of a public firm's objective function" *Journal of Economic Behavior and Organization*, 49, 487-499.
- Willner J. (1999), "Policy Objectives and Performance in a Mixed Market With Bargaining", *International Journal of Industrial Organization*, 17, 137-145.



# La dimensione temporale del processo produttivo in Nicholas Georgescu-Roegen e Gordon C. Winston

GIUSEPPE VITTUCCI MARZETTI E MARIO MORRONI

## 1. Introduzione

Sebbene l'attività di produzione richieda tempo e si dispieghi nel tempo, gli aspetti temporali dei processi produttivi non sono di solito discussi o esplicitati in modo chiaro nella maggior parte delle analisi economiche della produzione. I modelli di produzione in economia tendono generalmente a sottovalutare o ignorare completamente tutte le informazioni riguardanti il profilo temporale degli eventi nel processo produttivo. È importante sottolineare che, in questi modelli, gran parte dell'informazione temporale viene persa “non perché venga selettivamente tralasciata dopo essere stata valutata irrilevante alla comprensione del problema economico; l'informazione temporale è persa in modo automatico, filtrata via dal modo in cui l'economia tratta il tempo” (Winston, 1982, p. 3).

Due importanti eccezioni sono costituite dal modello fondi-flussi di Georgescu-Roegen (1969, 1970, 1971, 1990) e dall'analisi *time-specific* della produzione di Winston (1974a, b, c; 1977, 1982).<sup>1</sup>

Nicholas Georgescu-Roegen (1906-1994) e Gordon C. Winston (1929-2013) hanno proposto due diverse rappresentazioni analitiche del processo produttivo che consentono di analizzarlo come un processo che si svolge nel tempo,

<sup>1</sup> Una considerazione esplicita del tempo nella teoria della produzione all'interno di un quadro neoclassico si trova anche in Frisch (1964).

giungendo a risultati interessanti circa il ruolo che il tempo svolge nella produzione e i riflessi che questo ha sull'organizzazione dei processi, le economie di scala e di gamma, il grado di flessibilità produttiva, la domanda dei fattori, le tecniche *cost-minimizing* e i prezzi dei prodotti.<sup>2</sup>

Il presente lavoro analizza gli aspetti salienti e le implicazioni principali dei contributi di questi due autori, mettendone in risalto i caratteri di complementarietà.

Il lavoro è organizzato come segue. Nella Sezione 2 presentiamo il modello fondi-flussi di Georgescu-Roegen. Nella Sezione 3 consideriamo l'analisi *time-specific* della produzione di Winston. In particolare, dopo aver discusso gli aspetti relativi alla relazione tra scelta dell'unità temporale elementare e perdita informativa, ci concentriamo sulla sua analisi temporalmente specifica dei valori di produzione, che riteniamo costituisca il naturale complemento del modello fondi-flussi, dato che in quest'ultimo i prezzi dei prodotti e i costi dei fattori non vengono considerati esplicitamente. La Sezione 4 conclude il lavoro riassumendone brevemente i punti principali.

## 2. Il modello fondi-flussi di Nicholas Georgescu-Roegen

Il modello fondi-flussi di Georgescu-Roegen (1970, 1971) pone in evidenza il rapporto che intercorre tra efficienza economica e organizzazione della produzione grazie a una rappresentazione degli aspetti temporali e organizzativi dei processi produttivi.<sup>3</sup> Lo studio del profilo temporale dell'impiego degli elementi di produzione permette di analizzare il modo in cui gli input e i processi, indivisibili e complementari, si combinano al fine di minimizzare i tempi di inattività degli strumenti, aumentare il loro grado di utilizzo e ridurre la durata dei processi.

<sup>2</sup> La questione diversa ma correlata del ruolo del tempo nell'attività di consumo e nelle teorie del consumo in economia è trattata da Steedman (2001), che propone una teoria *time-specific* del consumo sulla base delle intuizioni di Gossen (discusse in dettaglio da Georgescu-Roegen (1983) nel suo saggio introduttivo all'opera di Gossen).

<sup>3</sup> Georgescu-Roegen presentò il modello fondi-flussi nel 1965 alla conferenza della International Economic Association tenutasi a Roma. Il modello è stato poi ripreso senza modifiche di rilievo in alcuni lavori successivi (Georgescu-Roegen, 1969, 1970, 1971, 1990). Sull'approccio fondi-flussi si vedano anche Morroni (1992, 1999, 2014), Scazzieri (1993, 2014), Piacentini (1995), Mir-Artigues e González-Calvet (2007), Vittucci Marzetti (2013).

## 2.1. Fondi, flussi e funzionale di produzione

Georgescu-Roegen divide gli elementi di produzione in flussi e fondi. I flussi sono utilizzati in un solo processo come input, oppure emergono da un singolo processo come prodotto finito. Un flusso di input o output corrisponde sempre a una certa quantità di materiale, sostanza, o energia, che entra nel processo, o che ne esce, in un dato istante di tempo. Al contrario, un fondo fornisce i suoi servizi in numerosi processi che si succedono nel tempo, entrando nel e uscendo dal processo produttivo. I fondi quindi non vengono mai incorporati nel prodotto e non possono essere accumulati o decumulati in un istante, come può avvenire al contrario nel caso dei flussi (Georgescu-Roegen, 1969, 1976, pp. 72, 83-86).<sup>4</sup>

Gli elementi di produzione sono definiti sulla base del loro ruolo all'interno del processo produttivo e del profilo temporale del loro utilizzo. Quindi la stessa merce può essere un flusso nel processo che porta alla sua produzione e un fondo in un processo che la utilizza come strumento di produzione. Per esempio, un computer è un flusso nel suo processo di produzione, ma un fondo nei processi in cui fornisce i suoi servizi produttivi. Tuttavia, come nota lo stesso Georgescu-Roegen, non si corre il rischio di confondere fondi e flussi se si presta attenzione all'unità di misura in cui sono espressi: i fondi sono espressi in termini di "sostanza" per tempo (o di servizi prestati da questa "sostanza" nel processo produttivo all'interno di un certo intervallo), mentre i flussi sono espressi in termini di "sostanza" per unità di tempo, ovvero quantità accumulata o decumulata in un certo intervallo temporale.

I fondi possono rimanere inattivi non solo durante le pause del processo produttivo, ma anche quando il processo è in corso. Inoltre, durante i periodi in cui un fondo entra attivamente nella produzione, il grado di utilizzo della sua capacità produttiva può variare. La produttività del fondo e i relativi costi di produzione sono pertanto, naturalmente, influenzati dal grado e dal tempo di utilizzo dei fondi.

Georgescu-Roegen definisce il processo elementare, come "il processo attraverso il quale un'unità di output ... è prodotta" (1971, p. 5). Un processo elementare è definito come il processo che inizia al tempo 0, con l'ingresso

<sup>4</sup> Un flusso può risultare sia dal decumulo di uno stock sia dalla trasformazione effettuata dal processo di produzione. Va però notato che, come rileva lo stesso Georgescu-Roegen (1969, 1970), mentre tutti gli stock si accumulano o de-cumulano in un flusso, non tutti i flussi implicano l'aumento o la riduzione di uno stock (ad esempio nel caso dell'elettricità).

delle materie prime, e termina all'istante  $T$ , con l'uscita di un'unità di prodotto finale. Per ogni singolo elemento del processo di produzione, sia esso fondo sia flusso, possiamo definire una funzione del tempo nell'intervallo chiuso e limitato  $t \in [0, T]$ .

Nell'approccio fondi-flussi, il processo di produzione è rappresentato dal seguente funzionale (relazione tra una funzione e un insieme di funzioni) (Georgescu-Roegen, 1971 p. 236):

$$O(t) = \Lambda[G_1(t), G_2(t), \dots, G_I(t), F_1(t), F_2(t), \dots, F_H(t), U_1(t), U_2(t), \dots, U_K(t)], \quad (1)$$

dove  $O(t)$  è una funzione che descrive in ogni istante  $t$  la quantità cumulata di prodotto finale;  $G_i(t)$  è la funzione dell' $i$ -esimo flusso di output di scarto che lascia il processo ( $i=1, 2, \dots, I$ );  $F_h(t)$  indica la funzione dell' $h$ -esimo flusso di input ( $h=1, 2, \dots, H$ ); infine  $U_k(t)$  è la funzione che indica il grado di utilizzo del fondo  $k$ -esimo ( $k=1, 2, \dots, K$ ).

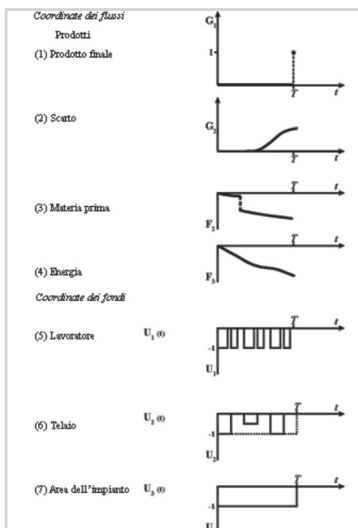
Per convenzione, diamo un segno positivo alle funzioni di output  $O(t)$  e  $G_i(t)$ , e un segno negativo alle funzioni dei flussi di input  $F_h(t)$  e dei fondi  $U_k(t)$ .  $O(t)$  e  $G_i(t)$  sono funzioni non-decrescenti del tempo, mentre  $F_h(t)$  sono funzioni non-crescenti del tempo. Dalla definizione di processo elementare segue poi che  $O(t) = 0$  per  $t \in [0, T)$  e  $O(T) = 1$ .

Le funzioni  $U_k(t)$  variano da 0, quando il fondo  $k$ -esimo rimane inattivo, a -1, quando la capacità produttiva del fondo  $k$ -esimo è pienamente utilizzata (Morrone, 1992, pp. 50-60).

La Figura 1 illustra un esempio di un possibile andamento di queste funzioni per un processo elementare generico con i seguenti elementi: i) flussi di output:  $G_1(t)$  prodotto,  $G_2(t)$  rifiuti; (ii) flussi di input:  $F_1(t)$  dalla natura,  $F_2(t)$  materia prima,  $F_3(t)$  energia; iii) fondi:  $U_1(t)$  lavoratore,  $U_2(t)$  telaio;  $U_3(t)$  spazio dell'impianto.

È possibile confrontate i vari profili temporali dell'impiego dei fondi guardando alle funzioni  $U_k(t)$ . Nella Figura 1, i profili temporali dei tre fondi sono differenti a causa della distribuzione irregolare dei tempi di presenza e attivazione. Il tempo di presenza è indicato con una linea tratteggiata, mentre quello di utilizzo con una linea continua. In particolare, risulta che il lavoratore è presente solo quando il processo è in funzione. Per contro, il telaio è presente per tutta la durata del processo elementare, anche se inattivo durante le pause, quando il processo è sospeso.

**Figura 1: Coordinate di fondi e flussi di un generico processo elementare**

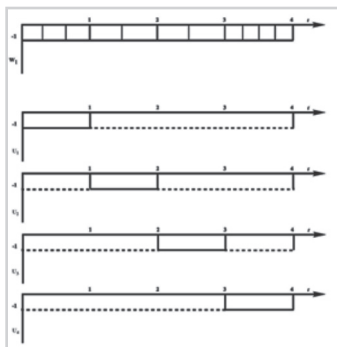


Fonte: Georgescu-Roegen (1970, p. 89)

## 2.2. Organizzazione temporale dei processi elementari e sistemi di produzione

Il modello fondi-flussi è molto utile per analizzare le differenze tra la produzione artigianale e la produzione di fabbrica. Nella produzione artigianale, i processi elementari sono organizzati in serie: uno di seguito all'altro. L'organizzazione in serie permette di produrre un bene alla volta secondo le specifiche del cliente. Questo favorisce un'elevata flessibilità produttiva, vale a dire la capacità di variare il mix di prodotti nel tempo, in base alle specifiche caratteristiche della domanda.

**Figura 2. Produzione artigianale: un lavoratore, quattro attività e quattro fondi**



Fonte: Morroni (2104, p. 9)

Nell'esempio illustrato nella Figura 2, la funzione  $W_1(t)$  indica, in ogni istante, il grado di utilizzo del lavoratore nel processo elementare. Le funzioni  $U_k(t)$  ( $k = 1, 2, 3, 4$ ) indicano il grado di utilizzo di quattro fondi rappresentati da quattro differenti strumenti disponibili nell'impianto. L'artigiano esegue un processo elementare svolgendo quattro attività diverse, ciascuna composta da diverse operazioni, che richiedono l'esecuzione di uno o più compiti elementari, che consistono in operazioni che, per definizione, non sono ulteriormente divisibili (per esempio, il carico o lo scarico di un prodotto intermedio o il taglio di un pezzo di tessuto). Nella Figura 2, la prima attività richiede l'esecuzione di tre compiti elementari, la seconda e la terza richiedono lo svolgimento di due compiti elementari, infine, la quarta richiede l'esecuzione di quattro compiti elementari. Il lavoratore utilizza quattro fondi, uno dopo l'altro. Tre su quattro dei fondi rimangono costantemente inattivi durante tutto il processo, potendo il lavoratore svolgere una sola attività alla volta (Morrone, 1992, pp. 60-67, 2014, pp. 8-9).

In sintesi, l'organizzazione artigianale della produzione ha tre caratteristiche fondamentali: i) alta flessibilità; ii) lunghi tempi di inattività per gli strumenti; iii) lunghi tempi di addestramento e di formazione, in quanto i lavoratori, dovendo svolgere tutti i compiti elementari, devono possedere una vasta conoscenza produttiva (Morrone, 2014, p. 8).

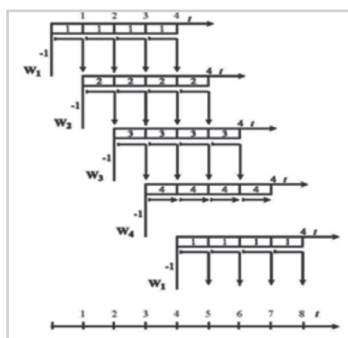
Supponiamo ora che un aumento della domanda comporti una crescita del volume della produzione tale da rendere necessario impiegare quattro lavoratori ( $W_1, W_2, W_3, W_4$ ). Se si mantenesse il modo di produzione artigianale, ciascun lavoratore, dei quattro assunti, svolgerebbe un intero processo elementare, eseguendo tutte le quattro attività una dopo l'altra, fino al completamento del processo elementare, in parallelo con gli altri quattro lavoratori. Di conseguenza, il numero dei lavoratori, degli strumenti e il volume della produzione sarebbero tutti moltiplicati per quattro. Rimarrebbero così inalterati sia i tempi d'inattività di ciascuno strumento sia il numero di conoscenze che ciascun lavoratore deve possedere.

Per ridurre i tempi di inutilizzo dei fondi (gli strumenti) e la conoscenza produttiva richiesta a ciascun lavoratore, è necessario passare a una produzione in linea tipica del sistema di fabbrica, avviando i processi elementari uno dopo l'altro con un certo ritardo predeterminato corrispondente ad un'uguale frazione della durata del processo elementare. Più precisamente, per eliminare completamente tutte le inefficienze causate dall'inattività dei fondi in un processo elementare di durata  $T$ , il tempo di ciclo (*cycle time*)  $\delta$  ( $\leq T$ ) deve essere uguale al massimo comune divisore (misura massima comune) delle durate degli intervalli di tempo di utilizzazione effettiva di ciascun fondo nel processo elementare. In questo caso, la grandezza minima efficiente del processo, cioè il numero di

processi elementari attivati contemporaneamente, è  $T/\delta$ . Il processo si stabilizza dopo  $(T - \delta)$  unità di tempo, o  $(T/\delta - 1)$  cicli; e la scala minima efficiente, vale a dire la quantità media di produzione per unità di tempo nel processo stabilizzato, diventa  $1/\delta$  (Tani, 1986, p. 219; Vittucci Marzetti, 2013).

Come messo in luce da Tani (1986), in un'organizzazione della produzione in linea si possono avere diversi livelli di divisione del lavoro compatibili con il pieno utilizzo dei fondi (Tani, 1986, pp. 222-24; cfr. Morroni, 1992, p. 63; Vittucci Marzetti, 2013). La Figura 3 illustra una possibile organizzazione in linea che annulla completamente i tempi di inattività e porta alla completa specializzazione (in questo caso  $\delta$  è uguale a 1). In questa nuova organizzazione del processo produttivo nel tempo, ogni lavoratore svolge una sola attività in sequenza. Ciò comporta una riduzione del numero di compiti elementari eseguiti da ciascun lavoratore con conseguente diminuzione del numero di capacità e competenze richieste. Il passaggio dalla produzione artigianale al sistema di fabbrica rende inoltre possibile un sensibile risparmio di capitale fisso. Infatti, l'insieme degli strumenti necessari a un artigiano per eseguire un processo elementare è sufficiente per effettuare quattro diversi processi elementari, perché adottando il sistema di fabbrica i tempi di inattività degli strumenti sono eliminati. Naturalmente, l'eliminazione dei tempi di inattività degli strumenti è tanto più importante quanto maggiore è il costo del capitale fisso (Morroni, 2014, pp. 8-10), anche se va sottolineato che il modello fondi-flussi non prende in considerazione direttamente il ruolo svolto dai prezzi degli input nella scelta della tecnica produttiva. È questo uno dei punti di complementarità tra il contributo di Georgescu-Roegen e quello di Winston che discuteremo più avanti (Sezione 3.2).

**Figura 3. Produzione in linea: quattro lavoratori, quattro attività e quattro fondi**



Fonte: Morroni (2104, p. 10)

### 2.3. Profili temporali di utilizzo dei fondi, scala di produzione e flessibilità

L'analisi della produzione in linea rende evidente che l'esistenza di una scala efficiente di produzione, e di eventuali inversioni di efficienza oltre certi volumi della produzione, deriva non solo dalla presenza d'indivisibilità degli elementi della produzione e non solo dal fatto che alcuni processi sono attivabili solo in una determinata scala, ma anche dalla particolare distribuzione dei tempi di utilizzo dei fondi in relazione alla durata del processo produttivo. Il profilo temporale di utilizzo dei singoli fondi dipende ovviamente dalla particolare tecnica produttiva adottata.<sup>5</sup>

Il rapporto tra organizzazione ed efficienza analizzato nel modello fondi-flussi assume particolare rilevanza nell'analisi di tutti i problemi connessi alla scala e al grado di flessibilità della produzione (Morrone 1992, cap. 11 e 12).

La dimensione dell'unità produttiva riflette la necessità di ridurre al minimo sia i tempi d'inattività sia il sottoutilizzo della capacità produttiva dei fondi, sincronizzando i vari processi e bilanciando le capacità produttive delle varie fasi. In presenza di fondi specializzati e indivisibili, tipici dei processi industriali, è spesso necessario raggiungere un elevato volume di produzione per rendere compatibili le diverse capacità produttive dei fondi e risolvere il problema di eventuali strozzature (*bottleneck*) da una fase all'altra del processo. Una volta stabilita una scala tale da eliminare i tempi di inutilizzo per i vari elementi fondo, qualsiasi espansione della capacità produttiva deve avvenire con salti discreti, vale a dire "secondo multipli interi della scala corrispondente alla piena utilizzazione della capacità produttiva". Se questa regola non è rispettata, la produzione sarà condotta in modo meno efficiente e si avranno "cadute di efficienza al di là di certi livelli di aumento della produzione" (Landesmann, 1986, pp. 308-9; Scazzieri, 2014).<sup>6</sup>

<sup>5</sup> A causa delle particolari caratteristiche tecniche dei processi, i profili temporali di utilizzo dei fondi sono generalmente specifici e limitati. Queste rigidità temporali di attivazione dei fondi fanno sì che il processo in linea abbia per lo più una gamma limitata di scale efficienti di attivazione, anche quando tutti gli elementi sono perfettamente divisibili. La divisibilità degli elementi della produzione è quindi una condizione necessaria, ma non sufficiente, per la divisibilità del processo.

<sup>6</sup> Se il processo di produzione elementare è scomponibile in diverse fasi intermedie, la necessità di coordinare le diverse capacità produttive dei vari elementi fondo non comporta necessariamente un aumento delle dimensioni dell'impresa laddove i costi di transazione siano comunque inferiori ai costi di internalizzazione delle attività interdipendenti. Di fatto, diversi processi intermedi possono essere eseguiti da una singola unità produttiva o da più unità produttive, che a loro volta possono appar-



Il sistema di fabbrica rappresenta proprio il tentativo di superare il problema dell'indivisibilità e complementarità, in modo da ridurre al minimo i periodi d'inattività dei vari elementi fondo. Secondo Georgescu-Roegen (1969, p. 68), il sistema di fabbrica, per questa proprietà straordinaria di raggiungere la massima economia di tempo, "merita di essere considerato insieme alla moneta come una delle due più importanti innovazioni economiche del genere umano".

In condizioni di mercato mutevoli, la flessibilità svolge un ruolo fondamentale nell'influenzare la competitività. L'analisi del grado di flessibilità richiede un modello che tenga conto esplicitamente della dimensione temporale dei processi produttivi. In particolare, l'approccio fondi-flussi mostra chiaramente come il grado di flessibilità sia legato a tempi brevi di settaggio delle macchine e alla dimensione dei magazzini. La riduzione dei tempi di settaggio e dei costi dei magazzini è un elemento chiave per migliorare la flessibilità (Morrone, 1992, pp. 180-186). Se si riducono i costi e i tempi di settaggio e di magazzino, è infatti possibile produrre un'ampia gamma di prodotti differenziati, godendo al contempo di economie di varietà e di scala.

### **3. L'analisi temporale della produzione di Gordon C. Winston**

Nella sua analisi della produzione, Gordon C. Winston combina la rappresentazione temporalmente specifica del processo produttivo di Georgescu-Roegen con le implicazioni di costo di "prezzi specifici al tempo" (*time-specific price*) e il nuovo concetto di "prezzi specifici alla durata" (*duration-specific price*) dei servizi di capitale (Winston, 1974a, b, c; 1977, 1982; Winston e McCoy, 1974).

Le implicazioni di costo dei prezzi specifici al tempo dei servizi del lavoro erano state precedentemente analizzate da Marris (1964), mentre il concetto di Winston di prezzi specifici alla durata si basa sull'analisi del "costo d'uso del capitale" (*owner cost of capital*) proposta da Jorgenson e Griliches (1967).

tenere alla stessa impresa o ad imprese differenti. È questa la sostanza della "critica transazionale" di Coase (1937) e Williamson (1975) alla spiegazione tecnologica della dimensione dell'impresa proposta, tra gli altri, da Stigler (1951), in cui l'impresa è vista come un insieme di attività indipendenti, complementari o rivali, soggette a diversi rendimenti di scala, e la dimensione dell'impresa deriva dalla soluzione di un problema di minimizzazione dei costi medi totali.

### 3.1. Unità temporale elementare e perdita informativa

Winston (1982) individua tre diversi significati del concetto di tempo utilizzato in economia: i) il “tempo analitico” (*analytical time*), che descrive “il contesto temporale di qualsiasi analisi economica”; ii) il “tempo prospettico” (*perspective time*), legato alla prospettiva temporale degli agenti economici (passato, presente e futuro); iii) il “tempo merce” (*commodity time*), il tempo visto come bene scarso da assegnare tra impieghi alternativi, come in Becker (1976).

Per quanto riguarda il tempo analitico, Winston (1982) propone la distinzione tra: i) “unità temporale di calendario” (*calendar unit time*) propria delle analisi convenzionali ( $T$ ); ii) “unità temporale elementare” (*elementary unit time*) ( $t$ ), vale a dire “la più grande unità temporale che può essere utilizzata per decomporre  $T$  e tale da permettere ancora di far emergere tutti i rapporti economici rilevanti” (1982, p. 24). Questa distinzione tra unità di calendario e unità temporale elementare si basa sulla dicotomia di Hood (1948) tra misure economiche del tempo esogene ed endogene. Mentre l’unità di calendario è una misura esogena, data esternamente dal processo analizzato, l’unità temporale elementare è un “intervallo endogeno funzionale determinato dal particolare processo economico preso in esame” (Winston, 1982, p. 24) ed è pertanto in genere differente per processi economici differenti.

L’unità temporale elementare determina il livello di “astrazione temporale” nell’analisi del processo produttivo: più grande è l’intervallo di tempo coperto dall’unità elementare, maggiori sono l’astrazione e la conseguente perdita di informazioni relativamente all’organizzazione nel tempo del processo produttivo, ma maggiore è anche la trattabilità analitica della rappresentazione e minore la quantità di dati necessari per descrivere analiticamente il processo.

A un estremo troviamo i casi in cui l’impiego di unità temporali elementari eccessivamente ridotte genera una “massa indigesta di informazioni temporali su eventi economici” (Winston, 1982, p. 19). Questo può essere considerato il caso della rappresentazione del processo di produzione propria del modello fondi-flussi di Georgescu-Roegen, in cui il tempo è continuo e il livello di astrazione temporale nella descrizione del processo elementare è praticamente nullo, ma al tempo stesso il modello risulta estremamente esigente in termini di dati richiesti e trattabilità analitica, risultando per molti versi intrattabile.<sup>7</sup>

<sup>7</sup> I contributi di Morroni (1992) e Piacentini (1995, 1997) possono essere letti in tale ottica come tentativi di ridurre la complessità analitica del modello fondi-flussi cercando al contempo di mantenere per quanto possibile l’impianto teorico e le intuizioni di base del modello stesso.

All'altro estremo affrontiamo al contrario il “problema più sottile e grave” che la durata eccessivamente lunga del periodo elementare, non adeguata al processo produttivo sottostante, “cancella la temporalità degli eventi” (Winston, 1982, p. 19). È questo, ad esempio, il caso dell'acritica estensione del periodo di produzione in agricoltura, proprio delle analisi seminali degli economisti proto-classici (per esempio, Petty, 1662; Cantillon, 1755; Quesnay, 1759) in cui l'analisi economica era focalizzata sulla creazione e distribuzione del surplus sociale nel settore agricolo e in cui pertanto l'adozione dell'anno solare come unità temporale elementare era giustificata, ai processi produttivi manifatturieri fatta ad opera di Smith (1776) e Ricardo (1817).<sup>8</sup> Come nota Hicks (1985), con l'implicita estensione di questo modello di produzione dai processi agricoli a quelli manifatturieri, Smith e Ricardo perdono “il periodo naturale dell'agricoltura, che nel Modello Originale aveva fatto sì che il singolo periodo fosse auto-contenuto”, e trasformano l'unità temporale da endogena a convenzionale: questa non ha più “nulla a che fare con i periodi di produzione dei vari beni. Così lo stock che è reso disponibile all'inizio del periodo non può più essere considerato come composto interamente di prodotti finiti, ma deve necessariamente consistere di semi-lavorati” (Hicks, 1985, p.32).

Come notato da Leontief (citato da ten Raa, 1986a), la definizione dell'unità temporale elementare di riferimento, laddove non si analizzi la produzione in tempo continuo, è una discretizzazione che genera sempre un problema di aggregazione, con conseguente distorsione (*bias*). La distorsione deriva dal fatto che alcune funzioni continue del tempo di input (output) sono trattate come funzioni a gradini (ten Raa, 1986b, p.812), e in cui pertanto input (output) che sono differenti rispetto al tempo di effettiva disponibilità (consegna) sono messi assieme e trattati come uguali.

Va notato come, in modo abbastanza curioso, mentre gli economisti classici commettono l'errore di estendere l'unità temporale elementare propria della produzione agricola alla produzione manifatturiera, gli economisti

<sup>8</sup> Nel processo produttivo del “grano”, l'intero processo si assume contenuto in un singolo “periodo naturale”, l'anno agricolo. Citando Mill (1826, p. 185), “un anno è assunto nell'economia politica come il periodo che comprende un ciclo di rotazione di produzione”. Tutti gli input (il capitale circolante, come le sementi e i mezzi di sussistenza dei lavoratori) vengono considerati anticipi fatti all'inizio del periodo e quindi entrano nel processo dall'inizio. L'output (il grano) è prodotto solo alla fine del periodo e il surplus è distribuito. I salari di sussistenza sono pagati ante factum, all'inizio del periodo di produzione.

neoclassici commettono l'errore opposto di utilizzare funzioni di produzione istantanee, che implicitamente assumono un processo in linea continuo stabilizzato, a cui si avvicinano alcuni processi manifatturieri, per descrivere analiticamente i processi agricoli.<sup>9</sup>

Ciò che è importante sottolineare è che la rappresentazione analitica di un processo di produzione tramite una funzione di produzione o un vettore di input-output rispetto a una data unità temporale elementare, come fatto generalmente nell'economia della produzione standard (ad esempio, Mas-Colell *et al.*, 1995, cap. 5; Varian, 1992, cap. 4; Coelli *et al.*, 2005, cap. 2), porta necessariamente ad escludere dall'analisi tutte le problematiche relative al tempo del processo all'interno dell'unità in questione. L'ipotesi implicita alla base di questa scelta è che o il processo di produzione è un processo in linea continuo stabilizzato, tale pertanto da risultare invariante al tempo, oppure che, nell'unità di tempo elementare scelta, il processo ha un profilo temporale banale, cioè tale da escludere qualsiasi necessità di coordinamento temporale tra i diversi input per evitare che si generino inefficienze.

Se si assume ad esempio che, dato un processo elementare definito all'interno del periodo di tempo elementare ( $t \in [0,1]$ ): i) tutti gli input entrano all'inizio del periodo; ii) tutti i fattori sono utilizzati lungo tutto l'intervallo; iii) l'output si ottiene alla fine del periodo; una rappresentazione in termini fondi-flussi di questo processo all'interno del periodo elementare è la seguente:

$$\begin{aligned} F_h(t) &= F_h; U_k(t) = U_k; 0 \leq t \leq 1 \\ O(t) &= 0; G_i(t) = 0; 0 \leq t < 1 \\ O(1) &= O; G_i(1) = G_i \end{aligned} \quad (2)$$

dove  $F_h$ ,  $U_k$ ,  $O$  e  $G_i$  sono costanti date. Questa rappresentazione fa emergere in modo chiaro che il profilo temporale assunto per il processo produttivo esclude *ipso facto* dall'analisi tutte le questioni relative al coordinamento temporale degli elementi di produzione. Di fatto, nel processo elementare descritto dal sistema (2) le uniche inefficienze che possono sorgere sono quelle relative alla sotto-utilizzazione di fondi non perfettamente divisibili, e non sono ipotizzate

<sup>9</sup> Sull'idea implicita del processo di produzione che sta dietro la funzione di produzione si vedano Georgescu-Roegen (1970) e Vittucci Marzetti (2013, pp. 7-9).

inefficienze associate ai tempi di inutilizzo dei fondi.<sup>10</sup> L'assunzione implicita in tale rappresentazione è che il problema della scelta dell'organizzazione più efficiente sotto il profilo temporale dell'utilizzo degli input entro il periodo di produzione è già stato affrontato e risolto nel modo migliore. Ovviamente, proprio per questo, tale schema non è adatto ad analizzare le problematiche dell'organizzazione temporale della produzione.

Va quindi notato, come già sottolineato nell'Introduzione, che l'informazione temporale è qui “persa in modo automatico, “filtrata via” dal modo in cui questi modelli trattano il tempo (Winston, 1982, p. 3).

### **3.2. L'analisi temporale della produzione e dei prezzi: prezzi invariati nel tempo, prezzi specifici al tempo e prezzi specifici alla durata**

Come abbiamo discusso nella Sezione 2, il modello fondi-flussi di Georgescu-Roegen mostra come una disposizione in linea dei processi elementari possa essere impiegata (ed è di fatto utilizzata nel sistema di fabbrica) per ottenere il pieno utilizzo degli elementi fondo caratterizzati da un particolare profilo temporale di utilizzo legato alle particolari caratteristiche del processo elementare che implica alcuni tempi morti a causa della specifica tecnica adottata.

Nell'analisi di Georgescu-Roegen, il pieno utilizzo è sempre l'uso più efficiente che può farsi di un elemento fondo,<sup>11</sup> e questo necessariamente segue dall'assunzione di “invarianza economica” dei fondi: per definizione, gli elementi fondo sono quelli che entrano ed escono dal processo identici, fornendo servizi produttivi nel processo stesso. A questo proposito, il modello è stato duramente criticato da Lager (2000, 2006) e da Kurz e Salvadori (2003), che hanno evidenziato correttamente come tale ipotesi escluda *ipso facto* dall'analisi questioni importanti legate ai mezzi durevoli di produzione, come il profilo temporale ottimale di utilizzo e la loro vita economica (cfr. Vittucci Marzetti, 2013, Sezione 4).<sup>12</sup>

<sup>10</sup> È questo il motivo per cui, come osserva Morroni (1992), l'indivisibilità degli elementi di produzione è l'unica fonte di rendimenti di scala crescenti generalmente considerata dalle teorie microeconomiche standard.

<sup>11</sup> Come osserva lo stesso Georgescu-Roegen, “uno degli aspetti più importanti dell'economia della produzione è come minimizzare questi periodi di inattività del fondo, sia che stiamo pensando a un uomo, a beni strumentali, o a un terreno. ... L'economia della produzione si riduce a due comandamenti: primo, produrre con il sistema di fabbrica; secondo, lasciare operare la fabbrica giorno e notte” (Georgescu-Roegen, 1970, pp. 6-8).

<sup>12</sup> Ad esempio, i problemi di cosiddetto “troncamento prematuro”, cioè la possibilità che un mezzo durevole di produzione diventi economicamente obsoleto prima della fine della sua durata massima in termini tecnici (cfr., per esempio, Kurz e Salvadori, 1997, 2003).

È interessante notare che Georgescu-Roegen non analizza il ruolo dei prezzi nei processi di produzione. L'analisi temporale della produzione di Winston (1982) colma questa lacuna del modello fondi-flussi e può essere considerata pertanto naturale complemento di questo modello.

Per iniziare è importante osservare che Winston (1982) non adotta la distinzione di Georgescu-Roegen tra fondi e flussi. Come afferma lo stesso Winston, “pur mantenendo lo spirito dell'analisi, non è necessario preoccuparsi della distinzione tra decumulo delle scorte (i suoi “stock”) e utilizzo nella produzione dei servizi di beni strumentali o di una squadra di lavoratori (i suoi “fondi”): le loro differenze rilevanti sono riflesse nell'analisi temporalmente specifica nel loro diverso comportamento in termini di prezzo” (1982, p. 48, enfasi nostra). Pertanto, secondo Winston, la distinzione tra fondi e flussi non è necessaria, in quanto emerge naturalmente dai diversi schemi normalmente adottati per determinare i prezzi dei vari elementi di produzione.<sup>13</sup>

Winston distingue due tipi di flussi di produzione (entrambi definiti rispetto a un'unità temporale elementare predeterminata): i) i “flussi di servizi” (*service flow*), “tra cui i servizi forniti dal capitale e gli input come l'energia elettrica”; ii) i “flussi di beni” (*goods flow*), “che includono i dadi, i bulloni, la stoffa e lo zucchero” (1982, p. 47).<sup>14</sup>

Per quanto riguarda l'analisi temporale delle tipologie di prezzo dei diversi fattori di produzione, Winston identifica tre diverse tipologie: i) “prezzi invariante al tempo” (*time-invariant price*), prezzi di flussi di merci che non variano nel tempo unitario  $T$ ; ii) “prezzi specifici al tempo” (*time-specific price*), i prezzi dei flussi di beni e servizi che variano nell'unità di tempo  $T$  (ad esempio l'energia elettrica e i servizi del lavoro); e, infine, iii) “prezzi specifici alla durata” (*duration-specific price*), cioè i prezzi dei flussi di servizi generati dagli stock quando il prezzo di mercato è relativo allo stock (il prezzo che si paga è relativo alla proprietà dello stock o alla sua effettiva disponibilità nel corso del periodo  $T$ ), e non all'ammontare dei servizi produttivi effettivamente prestati all'interno del processo produttivo.

<sup>13</sup> Di fatto comunque la definizione di “stock di capitale e di lavoratori” di Winston (1982, p. 42), ovvero fattori di produzione che forniscono servizi nel processo di produzione senza deprezzamenti causati da utilizzo, è praticamente equivalente a quella di fondo data da Georgescu-Roegen.

<sup>14</sup> Winston (1982, p. 52) distingue ulteriormente tra: i) “flussi immagazzinabili” (*storable flow*), cioè i flussi che possono essere accumulati (flussi accumulati, *accumulated flow*); ii) “flussi deperibili” (*perishable flow*), cioè flussi che non possono essere accumulati. Mentre i primi sono meno collegati al tempo effettivo di utilizzo entro l'unità di calendario, i secondi sono temporalmente specifici.

La nozione di prezzi specifici alla durata è fondamentale per quella parte dell'analisi della produzione di Winston che riguarda il capitale fisso. A questo proposito, Winston (1982, pp. 54-56) distingue tre diversi "prezzi del capitale": i) il prezzo di acquisto di un'unità di stock di capitale,  $P_m$  (per esempio, il prezzo di acquisto di un camion); ii) il prezzo di acquisto della disponibilità dell'unità di capitale nel corso del periodo  $T$ ,  $P_k$ ; questo è quello che viene solitamente indicato come "prezzo del capitale" (*rental price of capital*), o "costo d'uso del capitale" (*owner cost of capital*) (per esempio, il canone di locazione di un camion per una settimana); iii) il prezzo del flusso di servizi del bene capitale effettivamente generati nella produzione,  $p_k$ . Quest'ultimo prezzo è un prezzo specifico alla durata, in quanto inversamente proporzionale al tasso di utilizzo effettivo dello stock di capitale nel processo produttivo (ad esempio, assumendo che i servizi erogati dal camion siano proporzionali alla distanza percorsa e quest'ultima sia a sua volta proporzionale al tempo di utilizzo effettivo del veicolo, il prezzo per chilometro percorso dal camion in una settimana).<sup>15</sup>

Ipotizzando che il tasso di deprezzamento del bene capitale nel periodo  $T$ ,  $\delta_p$ , sia determinato solo da fattori che non dipendono dall'uso reale del bene (dipendano, ad esempio, solo da obsolescenza tecnologica), il prezzo dei flussi di servizi effettivamente generati dall'unità di capitale sarà:<sup>16</sup>

$$p_k = P_k/T_p = P_m(r_T + \delta_p)/T_p \quad (3)$$

dove  $T_p (\leq T)$  è il periodo di effettivo utilizzo del bene capitale in  $T$  e  $r_T$  è il tasso di interesse di mercato, pari al costo opportunità delle risorse impiegate.

Il punto importante sollevato da Winston (1982) è che l'esistenza di mercati di locazione (*rental market*) per gli stock non elimina la distinzione tra stock

<sup>15</sup> Ciò comporta supporre che il carico del camion sia sempre lo stesso, indipendentemente dal tempo effettivo di utilizzo, e che questo viaggi sempre alla stessa velocità.

<sup>16</sup> Se si assume che il tasso di deprezzamento del bene capitale dipenda a sua volta dal grado di utilizzo effettivo nel processo,  $\delta T$  diventa esso stesso funzione di  $T_p$ . Un limite importante di questo quadro teorico è tuttavia che il tasso di deprezzamento di un bene capitale in realtà non è esogeno, come assunto implicitamente nei modelli che ipotizzano "deprezzamento per evaporazione", o "deprezzamento da decadimento radioattivo", ma dipende dalla struttura complessiva del sistema di produzione e dalla distribuzione del reddito. Per una discussione approfondita del problema nell'alveo dell'approccio von Neumann-Sraffa si veda Kurz e Salvadori (1997, cap. 9).

e flusso di servizi dello stock; di volta in volta ciò che è effettivamente preso in locazione sul mercato è (o è anche) la disponibilità della risorsa produttiva nel corso del periodo, in modo indipendente da qualsiasi utilizzo effettivo della risorsa nel periodo in questione. Così, nel nostro esempio del camion, quando il prezzo pagato per noleggiare il camion per una settimana è indipendente da qualsiasi utilizzo effettivo del camion nel corso della settimana o, in ogni caso, il canone include qualche componente fissa indipendente da qualsiasi utilizzo effettivo del camion, lo schema di prezzo implica che sia il conduttore a pagare per i periodi di inattività del camion.

È interessante osservare che, nel caso in cui quelli acquistati fossero solo i servizi produttivi effettivamente generati dagli stock (nell'esempio del camion, paghiamo per il camion solo per i chilometri effettivamente percorsi nel corso della settimana), i costi di produzione non sarebbero dipendenti dai tempi di inattività e non ci sarebbe da parte del conduttore alcuna necessità di coordinare temporalmente i processi produttivi passando dalla produzione in serie a quella linea per ridurre i tempi di inutilizzo degli stock (dei fondi nel gergo di Georgescu-Roegen), perché questo schema di prezzo trasferirebbe i costi di tale inutilizzo dal concessionario al concedente. Al contrario, con prezzi specifici alla durata, ogni volta che non vi sono elementi con prezzi specifici al tempo coinvolti nella produzione, la minimizzazione dei costi comporta sempre la minimizzazione dei tempi morti per quegli elementi della produzione che hanno prezzi specifici alla durata. La minimizzazione dei tempi di inattività di questi stock comporta sempre una qualche forma di organizzazione nel tempo dei processi produttivi elementari quando sono presenti rigidità nel loro profilo temporale di utilizzo nel singolo processo elementare. Questi problemi tendono ad essere completamente trascurati nella rappresentazione *mainstream* del processo produttivo.

## 4. Conclusioni

Sebbene i modelli di produzione in economia di solito tendano a sottovalutare o ignorare completamente tutte le informazioni circa il profilo temporale degli eventi nel processo di produzione, vi sono due eccezioni di rilievo: il modello fondi-flussi di Nicholas Georgescu-Roegen (1969, 1970, 1971, 1990) e l'analisi della produzione *time-specific* di Gordon C. Winston (1974a, b, c; 1977, 1982). Georgescu-Roegen e Winston hanno proposto due diverse rappresentazioni analitiche del processo produttivo che consentono di analizzarlo come



un processo che si svolge nel “micro-tempo” (utilizzando l’espressione di Winston, 1982).

In questo lavoro abbiamo discusso i tratti salienti delle due teorie, evidenziandone gli aspetti originali, i risultati principali e gli elementi di complementarità.

Nella Sezione 2 abbiamo preso in esame il modello fondi-flussi. Georgescu-Roegen (1970, 1971) identifica i diversi possibili modelli di organizzazione della produzione nel tempo (in serie, in parallelo e in linea), mostrando formalmente come e perché la produzione in linea “merita di essere affiancata al denaro, come le due innovazioni economiche più importanti per l’umanità” (Georgescu-Roegen, 1970, p. 8), e dà il via agli studi sulle relazioni tra sistemi di produzione in linea, scala di produzione, specializzazione dei fattori e conoscenza produttiva, dando un contenuto analitico preciso al “famoso teorema di Adam Smith che la divisione del lavoro dipende dall’estensione del mercato” (Young, 1928, p. 529; cfr. Scazzieri, 1993; Morroni, 1992, 2014).

Nella Sezione 3 abbiamo discusso gli aspetti salienti del contributo di Gordon C. Winston. All’interno dell’accezione del termine tempo come “tempo analitico” (*analytical time*), che descrive il contesto temporale dell’analisi economica, Winston (1982) introduce la distinzione tra “unità temporale di calendario” (*calendar unit time*) e “unità temporale elementare” (*elementary unit time*). Mentre la prima è un periodo convenzionale dato esogenamente, la seconda rappresenta l’intervallo temporale unitario della rappresentazione analitica del processo di produzione, e come tale, nell’analisi della produzione *time-specific*, è endogena, poiché definita sulla base delle caratteristiche del processo in esame, affinché sia tale “da permettere ancora di far emergere tutti i rapporti economici rilevanti” (1982, p. 24). Cosa che non avviene nell’analisi della produzione standard, in cui il periodo di produzione elementare è dato esogenamente e fissato in modo convenzionale.

Nella sezione conclusiva abbiamo presentato la teoria della produzione di Winston (1982), che riteniamo costituisca il naturale complemento del modello fondi-flussi di Georgescu-Roegen. Il modello fondi-flussi è stato concepito per analizzare i processi in linea, permettendo di comprendere il collegamento tra scala di produzione, rendimenti di scala, specializzazione dei fondi, distribuzione dei compiti e della conoscenza produttiva, ma non analizza, come invece fa la teoria di Winston, il modo in cui i costi degli input e i prezzi dei prodotti determinano la tecnica produttiva intesa anche come organizzazione temporale della produzione.

L’analisi di Winston (1982) unisce la rappresentazione *time-specific* del processo produttivo del modello fondi-flussi con le implicazioni di costo di

prezzi specifici al tempo (*time-specific price*) e alla durata (*duration-specific price*) dei servizi produttivi generati dai fondi. La sua analisi mostra in particolare quando, come e perché i diversi schemi di prezzo di fatto adottati rendono rilevanti gli aspetti temporali del processo produttivo, con questo mettendo in relazione l'organizzazione dei processi produttivi nel tempo esaminata da Georgescu-Roegen, gli aspetti temporali dei prezzi e le tecniche *cost-minimizing*, e dando così origine a una teoria della produzione duale *time-specific*.

## Riferimenti bibliografici

- Becker G. S. (1976), *The Economic Approach to Human Behavior*, University of Chicago Press, Chicago.
- Cantillon R. (1755), *Essai sur la nature du commerce en Général*, ristampato in H. Higgs (a cura di) (1931), Macmillan, Londra.
- Coase R. H. (1937), "The nature of the firm", *Economica*, 1937, 4(16), 386-405.
- Coelli, T. J., Rao D. S. P., O'Donnell C. J. e Battese G. E. (2005), *An Introduction to Efficiency and Productivity Analysis*, Springer, New York.
- Frisch R. (1964), *Theory of production*, Kluwer Academic Publishers.
- Georgescu-Roegen N. (1969), "Process in farming versus process in manufacturing: a problem of balanced development", in U. Papi e C. Nunn (a cura di), *Economic Problems in Agriculture in Industrial Societies*, (Atti della conferenza della International Economic Association, Roma, Settembre, 1965), Macmillan, Londra; ristampato in N. Georgescu-Roegen (1976), *Energy and Economic Myths*, Pergamon Press, New York, 71-102.
- Georgescu-Roegen N. (1970), "The economics of production", *American Economic Review*, 60(2), 1-9.
- Georgescu-Roegen N. (1971), *The Entropy Law and the Economic Process*, Harvard University Press, Cambridge Mass.
- Georgescu-Roegen N. (1976), *Energy and Economic Myths*, Pergamon Press, New York.
- Georgescu-Roegen N. (1983), "Hermann Heinrich Gossen: his life and work in historical perspective", in H. H. Gossen, *The Laws of Human Relation and the Rules of Human Action Derived Therefrom*, MIT Press, Cambridge MA.
- Georgescu-Roegen N. (1990), "The production process and dynamic economics". in M. Baranzini e R. Scazzieri (a cura di), *The Economic Theory of Structural Change*, Cambridge University Press, Cambridge.
- Hicks J. R. (1985), *Methods of Dynamic Economics*, Clarendon Press, Oxford.
- Hood W. C. (1948), "Some aspects of the treatment of time in economic theory", *Canadian Journal of Economics and Political Science*, 14:453-468.
- Jorgenson D. W., Griliches Z. (1967), "The explanation of productivity change", *Review of Economic Studies*, 34, 249-283.
- Kurz H. D., Salvadori N. (1997), *Theory of Production. A Long-Period Analysis*, Cambridge University Press, Cambridge.
- Kurz H. D., Salvadori, N. (2003), "Fund-flow versus flow-flow in production theory: re-

- flections on Georgescu-Roegen's contribution", *Journal of Economic Behavior & Organization*, 51, 487-505.
- Lager C. (2000), "Production, prices and time: a comparison of some alternative concepts", *Economic Systems Research*, 12(2), 231-253.
- Lager C. (2006), "The treatment of fixed capital in the long period", *Economic Systems Research*, 18(4), 411-426.
- Landesmann M. A., R. Scazzieri (1996), "The production process: description and analysis", in M. A. Landesmann e R. Scazzieri (a cura di), *Production and Economic Dynamics*, Cambridge University Press, Cambridge.
- Marris R. (1964), *The Economics of Capital Utilization: A Report on Multiple-Shift Work*, Cambridge University Press, Cambridge.
- Mas-Colell A., Whinston, M. D., Green, J. R. (1995), *Microeconomic Theory*, Oxford University Press, New York.
- Mill J. (1826), *Elements of political economy*, Londra.
- Mir-Artigues, P. e González-Calvet, J. (2007), *Funds, Flows and Time. An Alternative Approach to the Microeconomic Analysis of Productive Activities*, Springer, Berlino.
- Morroni M. (1992), *Production Process and Technical Change*, Cambridge University Press, Cambridge.
- Morroni M. (1999), "Production and time: a flow-fund analysis" in K. Mayumi e J. Gowdy (a cura di), *Bioeconomics and Sustainability. Essays in Honor of Nicholas Georgescu-Roegen*, Elgar, Cheltenham, 194-228.
- Morroni M. (2014), "Production of commodities by means of processes. The flow-fund model, input-output relations and the cognitive aspect of production", *Structural Change and Economic Dynamics*, 29, 5-18.
- Neumann J. von (1945), "A model of general equilibrium", *Review of Economic Studies*, 9, 1-9.
- Petty W. (1662), *A Treatise of Taxes and Contributions*, ristampato in C.H. Hull (a cura di) (1986), *The Economic Writings of Sir William Petty*, Kelley, New York.
- Piacentini P. P. (1995), "A time-explicit theory of production: analytical and operational suggestions following a 'fund-flow' approach", *Structural Change and Economic Dynamics*, 6, 461-83.
- Piacentini P. (1997), "Time-saving and innovative processes". in G. Antonelli e N. De Liso (a cura di), *Economics of Structural and Technological Change*, Routledge, Londra, 170-183.
- Quesnay F. (1759), *Tableau Économique*, 3<sup>o</sup> ed., ristampato in M. Kuczynski e R. Meek (a cura di) (1972), Macmillan, London.
- Ricardo D. (1817), *On the Principles of Political Economy and Taxation*, ristampato in P. Sraffa e M. H. Dobb (a cura di) (1951), *The Works and Correspondence of David Ricardo*, vol. I, Cambridge University Press, Cambridge.
- Scazzieri R. (1993), *A Theory of Production. Tasks, Processes and Technical Practices*, Clarendon Press, Oxford.
- Scazzieri R. (2014), "A structural theory of increasing returns", *Structural Change and Economic Dynamics*, 29, 75-88.
- Smith A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, ristampato in R. H. Campbell, A. S. Skinner e W. B. Todd (a cura di) (1976), *The Glasgow Edition of the Works and Correspondence of Adam Smith*, vol. II, Oxford University Press, Oxford.

- Sraffa P. (1960), *Production of Commodities by Means of Commodities*, Cambridge University Press, Cambridge.
- Steedman I. (2001), *Consumption Takes Time. Implications for Economic Theory*, Routledge, Londra.
- Stigler G. J. (1951), "The division of labor is limited by the extent of the market", *The Journal of Political Economy*, 59(3), 185-193.
- Tani P. (1986), *Analisi Microeconomica della Produzione*, Nuova Italia Scientifica, Roma.
- ten Raa T. (1986a), "Dynamic input-output analysis with distributed activities", *Review of Economics and Statistics*, 68, 300-310.
- ten Raa T. (1986b), "Applied dynamic input-output with distributed activities", *European Economic Review*, 30, 805-831.
- ten Raa T. (2006), *The Economics of Input-Output Analysis*, Cambridge University Press, Cambridge.
- Varian H. R. (1992), *Microeconomic Analysis*, W. W. Norton & Company, New York.
- Vittucci Marzetti G. (2013), "The flow-fund approach: a critical survey", *Journal of Economic Survey*, 27 (2), pp. 209-33.
- Williamson O. E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York.
- Winston G. C. (1974a), "Capital utilization and optimal shift work", *Bangladesh Economic Review*, 2, 515-558.
- Winston G. C. (1974b), "Factor substitution, ex ante and ex post", *Journal of Development Economics*, 1, 145-163.
- Winston G. C. (1974c), "The theory of capital utilization and idleness", *Journal of Economic Literature*, 12, 1301-1320.
- Winston G. C. (1977), "Capacity: an integrated micro and macro analysis", *American Economic Review*, 67, 418-423.
- Winston G. C. (1982), *The Timing of Economic Activities. Firms, Households, and Markets in Time-Specific Analysis*, Cambridge University Press, Cambridge.
- Winston G. C. and McCoy, T. O. (1974), "Investment and the optimal idleness of capital", *Review of Economic Studies*, 41, 419-428.
- Young A. A. (1928), "Increasing returns and economic progress", *The Economic Journal*, 38(152), 527-542

# Free from what?

## Competition, regulation and antitrust in American economics, 1870–1914

NICOLA GIACOLI

### Introduction: the history of economic thought viewpoint

If asked to summarize in a single decision and a single doctrine the constitutional jurisprudence of the Gilded Age and Progressive Era,<sup>1</sup> most American jurists would mention *Lochner v New York* and freedom of contract. *Lochner*<sup>2</sup> is among the Supreme Court's most famous, and most reviled, decisions – the symbol of the era of so-called laissez faire constitutionalism, when the Court substantively applied the due process clauses of the Constitution to strike down various state and federal laws that infringed constitutional rights to property and freedom of contract. The latter freedom was at the core of such jurisprudence, as all the laws canceled by the Court would have limited the contractual liberty of American citizens in the effort to further some public interest or social goals.<sup>3</sup>

The literature on *Lochner* and laissez faire constitutionalism is immense; for obvious reasons, it is entirely written from the viewpoint of legal or political history. The goal of this essay is to offer a different perspective for examining the freedom of contract saga.

<sup>1</sup> By the terms Gilded Age and Progressive Era it is meant the period of American history going from the 1870s to the end of WWI. No exact boundaries exist, but the Gilded Age is usually delimited by the turn of the century, with the remaining years ascribed to the Progressive Era.

<sup>2</sup> *Lochner v New York*, 198 U.S. 45 (1905).

<sup>3</sup> For example, at stake in *Lochner* was a law by the state of New York limiting the working hours of bakers for health and sanitary reasons.

The history of economic thought is well acquainted with freedom of contract too. The notion lay at the heart of one of the two interpretations of free competition offered by classical economists. As the embodiment of classical competition, freedom of contract remained crucial for the next generations of economists, most notably the American scholars who faced the theoretical and policy challenges raised by the impressive economic transformations of the post-Civil War years. In fact, the classical economists' view of contractual freedom also played a central role in the heated legal and political controversies about competition and regulation that occurred in the US between 1870 and 1910, following the spectacular growth in industrial size and concentration. In other words, it may be demonstrated that, when debating about the good and the bad of a free market system vis-à-vis government interference, American jurists and politicians were actually, if often unwittingly, referring to the freedom of contract ideal developed by the Classics and still endorsed by most economists of the time.

My thesis is that, independently of the jurisprudential trajectory that led the *Lochner* Court to embrace freedom of contract, we may conceive of the classical notion of freedom of contract – together with its classical alter ego, freedom to trade – as a red thread connecting the legal, political and economic discourse of the Gilded Age and Progressive Era about free markets, the limits of competition and the desirability of regulation. Thus, the history of economic thought may have a lot to say about the intellectual roots of so crucial a period of America's social and economic history and of its constitutional law. And given that in the fields of regulation and antitrust we still live under many respects in the shadow of the legislation developed at the time, it may even be argued that the late 19<sup>th</sup>-century “freedom of contract versus freedom of trade” dichotomy still holds sway.

In view of this thesis, the essay has a simple structure. In the first part (sections 1 to 4), I present an overview of the classical notions of freedom of contract and freedom to trade, together with an analysis of the classical market mechanism and theory of monopoly. Then, in the second part, I show how the dialectic between the two freedoms shaped the jurisprudential, political and economic debates about regulation (section 7) and the protection of competition (secs. 8-10). The Conclusion argues for the present-day relevance of these debates.

## 1. Two notions of free competition

The United States of the Gilded Age was allegedly the land of *laissez faire*. Filling the popular expression “*laissez faire*” with precise analytical content is not so simple, though. The most immediate solution is to equate it with free competition and the absence of state intervention. Yet, both notions are themselves less than obvious. To start with, what exactly does “free competition” mean? Free from what? How to assess when and if this “freedom” exists? This basic question admits of two answers.

According to the commonest interpretation, the yardstick of free competition is, quite literally, the existence of the freedom to compete, or *freedom to trade*. Under free competition, individuals should be free to pursue any kind of market activity without external interference or constraint, whatever its source. This interpretation is bound up with economic theory. A freely competitive market is depicted as an atomistic structure of diffused property, made up of prevalently small businesses; a universe of “small dealers and worthy men”,<sup>4</sup> whose independence and possibility to compete is guaranteed by the market mechanism itself. Competition, both actual and potential, is the key economic force through which the market polices itself, warranting freedom to trade. Any market power is only temporary: whenever a business obtains, by either luck or merit, a supra-competitive profit, free entry brings profit back to the competitive level. No market position is in fact permanent or safe under freely competitive conditions; everything is fluid and subject to change under competitive pressure. Both free entry and market atomization highlight the *horizontal* character of competition: free competition is a market structure characterized by multiplicity and free entry of firms all placed at the same level of the production and marketing chain. The magic of free competition is that it guarantees everybody’s freedom to trade while rewarding with temporary extra gains those smart enough to beat their rivals.

Competitive forces may stop functioning in two cases. The first is when the state disturbs their working by hindering actual or potential competition. The most general form of government interference is the creation of an artificial privilege or constraint – that is to say, an obstacle to the freedom to trade. In those cases, some wealth exists that only certain individuals designated by government acts can gain and that, conversely, all other individuals are

<sup>4</sup> To borrow Justice Peckham’s dictum in *United States v Trans-Missouri Freight Association*, 166 U.S. 290 (1897), at 323.

legally prevented from acquiring. The second hindrance to freedom to trade may arise from the market itself, whenever, contrary to the general principles of free competition, a business becomes so powerful that competition, both actual and potential, cannot effectively work against it anymore. Due to technological and organizational factors, this was an ever more frequent situation in the American economy of the post-Civil War decades.

Familiar as it is, the structural/horizontal characterization of free competition is not unique. Another common view of free competition also focuses on the idea that individuals should be totally unconstrained in pursuing their economic activities, but emphasizes the *vertical* dimension of contract and exchange, rather than the horizontal ones of atomization and free entry. Competition is free, then, when the utmost *freedom of contract* is guaranteed.

By freedom of contract is meant the possibility for any individual to bargain about her own property rights without any hindrance or constraint. The notion stems directly from the legal protection of property rights, including the right to the value of property itself. Since value is determined in the market and materializes through exchange (i.e., contractual activity), property is truly protected only when the individual is free to enter any kind of contract she deems proper to reap the value of her property. John Stuart Mill effectively put it: “The right of property includes then, the freedom of acquiring by contract. The right of each to what he has produced, implies a right to what has been produced by others, if obtained by their free consent; since the producers must either have given it from good will, or exchanged it for what they esteemed an equivalent, and to prevent them from doing so would be to infringe their right of property in the product of their own industry” (Mill 1909 [1848], II.2.3).

Buyer-and-seller bargaining involves agents placed at different levels of the supply-and-demand chain. Hence, depicting free competition as freedom of contract highlights the vertical character of competition. This, rather than the structural/horizontal dimension, was the image of competition classical economists – starting with Adam Smith – primarily emphasized. For them competition meant first and foremost the contrast of interests between buyers and sellers, each trying to get the most from the bargain. Indeed, their notion of free competition as full contractual liberty transcended economics and drew most of its attractiveness from higher values, such as the classical liberal goals of enhancing personal freedom and autonomy and ensuring complete control over the fruits of one’s own labor. It was mainly because their political economy put at center stage these liberal values that classical economists granted so great an importance to vertical buyer/seller relations and, conse-



quently, to the freedom-of-contract version of free competition.

For competition to be totally free, contractual liberty should be absolute, under the only proviso of not violating someone else's contractual liberty. Thus, classical liberal economists and jurists admitted as the only exception to freedom of contract – and thus the only rationale for government interference – the Latin dictum *sic utere tuo ut alienum non laedas*. After the Civil War, this common law maxim would become the cornerstone of American police power jurisprudence.<sup>5</sup> For the rest, freedom of contract ought to reign undisturbed.

## 2. Freedoms on collision path

The two freedoms, with their respective associated images of competition, could happily co-exist in the classical system. While classical economists emphasized the vertical view of competition more than the horizontal one – indeed, they did not even conceive of a market structure in the modern sense of the word, this being a notion dating back no earlier than the 1920s – the bargaining between a buyer and a seller, typically both of small dimension, could cause no trouble to the atomistic character of markets. In any case, the equalizing action of market forces guaranteed that it did not. Thus, the most complete exercise of contractual freedom was consistent with the maintenance of freedom to trade.

But the germ of the eventual conflict between the two freedoms already existed. In a truly free system, contractual liberty should suffer no constraint even when its use may cause, or favor, the concentration of market power – that is to say, when it may violate freedom to trade by limiting horizontal competition. By way of example, think of industrial combinations. Absolute freedom of contract entails that firms be free to form, or join, a cartel or any other combination – indeed, the law should *protect* their right to do so as a legitimate expression of their contractual liberty. Monopolistic positions may

<sup>5</sup> In Herbert Spencer's version, the maxim read: "Every man has freedom to do all that he wills, provided he infringes not the equal freedom of any other man" (Spencer 1851, 103, original emphasis). Police power meant at the time the government's authority to protect public health, safety, morals and general welfare by appropriate laws. It was recognized that to perform their task these laws might even restrain individual rights of liberty, property and contract.

thus arise in the marketplace as spontaneous outcomes of bargaining freedom; moreover, they too are entitled to exemption from state interference. Such a way of exploiting freedom of contract is however in blatant violation of the other notion of free competition, namely, freedom to trade. It seems as if to defend the latter some limits need be put on the former. During the Gilded Age, the conflict was bound to explode. Economists, jurists and politicians struggled to reconcile the two freedoms – each, don't forget, based upon a specific economic notion of competition – in those all-too-frequent cases where they contradicted each other, namely, whenever market forces proved unable to dissolve monopolistic concentrations *against the prediction of classical economics* (see below).

The dichotomy between freedom to trade and freedom of contract, and the related one between horizontal and vertical competition, also entail a different perspective about the limits of state intervention in the economy.<sup>6</sup> The vertical view sees the market as a mechanism to facilitate the exchanges among self-interested individuals; by doing so, the Classics guaranteed, the market also promotes the public good. Analytically speaking, buyer/seller bargaining brings market price to its normal, or natural level, eliminating both excess profits and unsatisfied wants. Exchange activity takes place within a framework of private property and contractual freedom, where the only role for government and the law is to protect property rights and remove any impediment to their unfettered circulation. It follows that free competition is not only synonymous with freedom of contract, but also with *freedom from government interference*. A public commitment to promoting individual economic liberty emerges as an essential component of this perspective: government must limit itself to safeguarding everyone's freedom to manage their own business.

The horizontal view of competition also underlines freedom from government interference. Yet, a major ingredient of this view is the notion of an atomistic market. Such a configuration need be preserved, lest competition cease to work. As explained above, classical economists believed that, by default, any market power could only be short term because, in the absence of state privileges or constraints, supra-competitive profits would attract new firms into the market that would erode that power. Public intervention was not required and often counterproductive. Hence, with respect to government's role in the economy, both accounts of free competition, vertical and

<sup>6</sup> See Peritz 1996, Ch.1; Page 2008; Giocoli 2014, Ch.2.

horizontal, led classical economists to the same laissez faire conclusion. This conclusion stemmed from specific analytical arguments, not (or not exclusively) from political or moral ones.

But when in the last quarter of the 19<sup>th</sup> century, under the push of impressive technological and organizational progress, markets revealed a tendency towards increasing concentration, many economists of industrialized countries realized that, contrary to classical predictions, market power could well be a spontaneous and permanent feature of economic life. In the presence of significant market power, the freedom to trade of other economic agents diminishes, and possibly vanishes; the atomistic structure is altered; competition is no longer free. American economists soon realized that it was not just a matter of reconciling classical economics with the rising concentration of real world markets. The biggest problem was that in some cases market power could evolve – actually, had evolved – into *economic* power. While market power refers to the possibility of gaining supra-competitive profits, by economic power I mean power of a broader, more dangerous kind – power the sway of which extends beyond the boundaries of a single market and reaches the entire economy and, possibly, also the social, political and institutional spheres. A power that embodies the worst fears of the Founding Fathers (think of James Madison’s “fear of faction”), namely, the specter of the economically powerful thwarting democracy and vanquishing the Republic – what, in a word, Gilded Age Americans used to denounce as *plutocracy*.

By the end of the 19<sup>th</sup> century, the freedom-to-trade, horizontal view of competition became synonymous with *freedom from market* (and, therefore, economic) *power*. The latter kind of freedom replaced freedom from government interference as the necessary safeguard of free competition. Indeed, public intervention was expressly invoked. Many late 19<sup>th</sup>-century economists understood that the market mechanism alone was unable to dissolve market power – indeed, it might become the cradle that nurtured it. The only way to constrain, or eliminate, market power (thereby preventing its transformation into economic power) was through deliberate government action, in the form of either an antitrust law or direct market regulation. A public commitment to economic equality, in the sense of the preservation of equivalent opportunities in the marketplace, became the intellectual cornerstone of this approach. Accordingly, its supporters were ready to accept various kinds of legislative and executive intervention to uphold it.

Once again, it is easy to see how the two commitments to liberty and equality were destined to clash. The two views of free competition as freedom of contract and freedom to trade that could find reconciliation in the small-busi-

ness world envisaged by classical economists had morphed, by the end of the 19<sup>th</sup> century and in the wake of wholly different market conditions, into an opposition between freedom from government interference and freedom from market power. Pursuing the latter often meant negating the former, and vice versa. The “free from what?” question was about to occupy center stage in economic as well as jurisprudential and political debates.

### **3. The classical competitive mechanism**

The rest of this essay shows that, broadly speaking, American courts took a well-defined position in the controversy between the two freedoms. Scores of judges at all levels endorsed the freedom of contract perspective, with its commitment to individual liberty and hostility against state interference. Freedom of contract was the view that most properly fit the framework of the classical liberal Constitution as originally envisaged by the Founding Fathers. Given that it was also the feature of competition that classical economists most consistently remarked, contractual liberty emerged as the single most important thread connecting the American Constitution and the Gilded Age and Progressive Era courts with Adam Smith and his heirs.<sup>7</sup>

For classical economists, competition did not just represent a general organizing principle of social life. It also had a specific analytical function to perform in their models, namely, to bring market price to its normal, or natural level, via the elimination of both excess profits and unsatisfied wants.<sup>8</sup> Thus, classical economists conceived of competition as the process overseeing the functioning of the price mechanism. Competition and the price mechanism were two sides of the same coin, viz., what made markets work.

Within the classical model, competitive actions consisted first and foremost in the *vertical* intercourse between buyers and sellers. The Smithian slogan “buy cheap or sell dear”<sup>9</sup> captured the vertical essence of market activity

<sup>7</sup> On the classical liberal Constitution, see Epstein 2014.

<sup>8</sup> See McNulty 1967, 396. The natural price, which Smith also called “the price of free competition” (WN I.7.27), was the reference point for the theorist as well as for market participants.

<sup>9</sup> “The capital of a wholesale merchant, on the contrary, seems to have no fixed or necessary residence anywhere, but may wander about from place to place, according as it can either buy cheap or sell dear” (WN II.5.14).

and, therefore, of competition itself. Analytically speaking, competition was a *behavior* – a process, not a state (Blaug 1997, 66). It meant the actions and reactions of buyers and sellers bargaining in the marketplace, not a specific market structure or equilibrium position. For the competitive mechanism to work, the only analytical requirement was that these buyers and sellers be endowed with the utmost freedom during their vertical intercourse, i.e., that they be free to act and react.

In a world of vertical relations, where each bargaining party tries to gain the most from every transaction, the power to set and modify prices is the main competitive weapon. The rivalry between buyers and sellers – viz., their effort to “buy cheap or sell dear” – suffices for the price mechanism to work. Under free competition, open and costless access to all exchange opportunities guarantees that no party may eventually exploit another. In his best-selling treatise, *Political Economy*, American economist Francis Walker thus captured the gist of classical competition: “Competition signifies the operation of individual self-interest among the buyers and sellers of any article in any market. It implies that each man is acting for himself solely, by himself solely, in exchange, to get the most he can from others, and to give the least he must himself” (Walker 1892, 91-2).

The analytical view of competition perfectly matched the classical economists’ belief about the socially beneficial role of individual self-determination in directing the allocation of resources. A competitive agent was an agent free to act and react, i.e., to enter any contract she desired, unconstrained by either government or the law. Even when viewed from the theoretical, rather than institutional perspective, the essence of classical competition thus lay in the freedom of contract and, related to it, the freedom from government interference.

The vertical dimension of competition found counterpart in the common law notions of contract and contracting parties: a contract was nothing but a juridical crystallization of buyer/seller bargaining activity. Legally speaking, the requirement of freedom translated into the necessity of guaranteeing the parties complete liberty to enter, exit or mold contractual relations. Thus, freedom of contract lay at the core of the common law, itself the main component of the institutional framework classical economists took as a datum.

Competition was not only vertical, though. Competition among sellers or among buyers was the other necessary component of the price mechanism. Thus, classical competition also consisted of a *horizontal* struggle between agents operating at the same level in the same market – say, two or more producers of the same good. As it turns out, the centrality of the vertical dimen-

sion also affected the classical analysis of this kind of competition.

Horizontally speaking, the essence of, say, sellers' competition was to undersell rivals and lead customers to patronize one's own product. As George Stigler (1957, 1) put it, classical competition was a "rivalry in a race" to win the purchasers' favor. Setting and cutting prices were again the main competitive weapons. The analytical task of horizontal competition in the classical model was to prevent either the seller or the buyer from exploiting their vertical counterparts by coercing them into lopsided exchanges. Competition guaranteed that whenever the seller increased the price, the buyer would always have the opportunity to quit the relation and buy at a lower price from another seller. The assumption was that in a freely competitive system – now to be read as a system of unfettered freedom to trade – such an alternative seller *always existed* or, at least, was always ready to enter a market where profitable opportunities arose. In short, that potential competition was always an active and powerful force.

More generally, we may say that in the classical world of perfect economic liberty the number of horizontally-competing agents was *indefinite*. Any buyer had always the option of abandoning a seller charging an "excessive" price and turning to a cheaper provider of the same good, because, given actual and potential competition, such a cheaper provider was always assumed to exist. The same held for any seller facing a buyer who bid too low. To prevent entry and preserve its vertical relations, the original seller had to limit, or avoid, the price increase. Ditto for the original buyer. For classical economists, competition was first and foremost a force depriving market participants of the power to influence the price. Freedom of entry, itself an embodiment of freedom to trade, was therefore the only structural feature classical economists invoked in their market analytics. Granted complete entry freedom, even a one-to-one buyer/seller relation would suffice for the price mechanism to work, i.e., push price to its natural level. Other structural assumptions, like those about the large number and small size of competitors, were basically redundant in such a setup.

In a market with complete free entry, horizontal competition was necessarily impersonal. It could never be targeted at a specific rival because no firm could ever undertake a competitive action – say, selling below cost (so-called cutthroat competition: see below) – with the deliberate goal of harming *that* rival. The latter would always have the possibility to sell to someone else and survive the attack. In the classical view, unfair trading practices were therefore vain, if targeted horizontally, or simply suicidal, if implemented vertically (by, say, coercing one's own customers). Hence, it was not just in an ethical

sense that classical economists “t[ook] it for granted” that custom, i.e., fair trade, was “the only serious hindrance in the way of perfect competition” (where “perfect” should be read as “unbridled”).<sup>10</sup> The upshot was remarkable, judicially speaking.

Courts were often called to assess the effects of competition over someone’s property rights, like, say, the value of a business. For some judges, fair trade and obedience to custom carried with them a sort of “right to profit”, i.e., a businessman’s entitlement to a normal return on his own capital, provided he behaved correctly in the marketplace – something that could be taken for granted under classical competition, since no misbehavior was possible in the form of coercion over customers or rivals. Others courts held on the contrary that no such “right” existed, and that businessmen should always bear the consequences, however negative, of competition, conditional on the latter having been fairly and customarily practiced – which again was always the case, by definition. Crucially, the analytical paradigm of classical competition underlay the courts’ reasoning in both scenarios. The issue – which will be at the heart of all legal controversies about regulation and antitrust – was therefore theoretical as much as moral.

#### **4. Monopoly: special privilege or market outcome?**

Competition was not just an analytical engine for classical economists, but also – in fact, chiefly – an institutional assumption. An implication of conceiving of competition in the latter sense is that there is no need to define it analytically – viz., in terms of economic categories – nor to explain how competition is endogenously generated within the model. Joseph Schumpeter claimed that classical economists were “so firmly [...] convinced that the competitive case was the obvious thing, familiar to all, that they did not bother to analyze its logical content. In fact, the concept was usually not even defined. It just meant the absence of monopoly – which was considered as abnormal and was vigorously condemned, but was not properly defined either – and of public price fixing” (Schumpeter 1986 [1954], 545-6). As we explained in the previous section, this is not entirely correct, in that competition was indeed identified, albeit only indirectly, as freedom of contract. But even if we accept

<sup>10</sup> So recited the entry “Competition and Custom” in Inglis Palgrave’s influential *Dictionary of Political Economy* (Palgrave 1894, 377).

that classical competition “just meant the absence of monopoly”, the question arises as to what happened when competition could not work – that is to say, when monopoly existed.

#### 4.1. Monopoly as privilege

Classical economists substantially disregarded the issue of monopoly. According to Stigler (1982, 1), they followed the Smithian tradition of paying “no attention to the formal theory of monopoly”. The commonest explanation of this attitude underlines that, given the long-run orientation of their economics, their neglect of entry barriers, and the still limited role played in contemporary business affairs by sunk costs and technical progress, classical economists were bound to devote little attention to monopoly. Specific analytical factors play an important role in this account. In the classical model, monopoly was a short-term phenomenon that the forces of competition, if left free to operate without external interference, would quickly eliminate. The ubiquitous presence of potential competition, coupled with complete free entry, ensured that any supra-competitive profit would immediately attract new capital that, in turn, would bring profit back to its natural level. Here lay the gist of the *profit equalization theorem*, one of the core analytical principles of classical economics.

For Smith and his heirs, perfect resource mobility and entry freedom drove market prices to their natural level by equalizing the total advantages of alternative employments of labor and capital. Smith had claimed in the *Wealth of Nations* (WN) that an individual would invest in a resource, be it capital, land or labor, so as to earn the highest possible return on it. It followed that all uses of the resource should yield an equal rate of return (adjusted for the relative riskiness of each enterprise), lest reallocation of the resource to alternative uses result. Stigler (1976, 1201) called this argument the “most substantive proposition in all of economics”.<sup>11</sup>

Epitomized by the profit equalization theorem, the Smithian characterization of competition would be neither “amplified [nor] challenged in any significant respect for the next three-quarters of a century by any important member of the English school” of classical economics (Stigler 1982, 3). By way

<sup>11</sup> For the theorem to work, further conditions on knowledge and time are necessary: “The entrepreneur (or other agents) must know what returns are obtainable in various fields, [...] and he must be given time to make his presence felt in these fields” (Stigler 1982, 3). While rarely expressing them, classical economists held these conditions also to be true.



of example, in 1825 Ramsey McCulloch declared: “The inextinguishable passion for gain – the *auri sacra fames* – will always induce capitalists to employ their stocks in those branches of industry which yield, all things considered, the *highest rate of profit* [...] But the rate of profit in different employments has a natural tendency to equality; and it can never, when monopolies do not interpose, continue either permanently higher or lower in one than in the rest” (McCulloch 1825, 163, original emphasis). Three decades later, Nassau Senior could still write that the operation of competition “can be supposed to be perfect only if we suppose that there are no disturbing causes, that capital and labour can be at once transferred, and without loss, from one employment to another, and that every producer has full information of the profit to be derived from every mode of production” (Senior 1854, 102).

The theorem’s implication for monopoly theory was clear. Smith, like some of his predecessors and all of his heirs, “intensely disliked monopoly in all its forms” (Viner 1960, 65), but at the same time deemed permanent private monopoly impossible. This because classical economists extended to every industry and trade Smith’s argument that “the size of the task, the number of persons involved, and their dispersion over space, made the establishment of an enduring monopoly a practical impossibility” (ibid.). Easy and rapid entry of newcomers would always automatically discipline any firm, or group of firms, that attempted to charge monopoly prices. In short, free competition – actual and potential – made persistent monopoly unfeasible.

This thesis, which we may call the *non-persistence argument*, itself enjoyed persistence in the economic literature. Indeed, it long outlasted classical economics proper. Faith in potential competition remained strong in many of the early neoclassical authors. The best economic mind of the Gilded Age and Progressive Era, John Bates Clark, famously established potential competition as the most fundamental feature of the market mechanism – a sort of magic wand capable of dispelling any supra-competitive profit. “The competitor who is not now in the field, but who will enter it at once if prices are unduly raised, is the protector of the purchasing public against extortion”, he proclaimed. “The competition that is now latent, but is ready to spring into activity if very high prices are exacted, is even now efficient in preventing high prices” (Clark 1900, 407). As it turns out, the non-persistence argument (along with several other classical doctrines) was alive and kicking long after the “official”, text-book-wise ending of the classical era around 1870.

Never a spontaneous outcome of the market mechanism, permanent monopoly could only be an artificial result of external interference with that very mechanism. Such interference, the Classics believed, could take many forms,

but all of them could be subsumed under the single heading of *special privilege*. The latter term indicated any kind of legislation aimed at allocating resources in a way that denied equality of opportunity to all individuals, usually by constraining their freedom of contract. In other words, monopoly meant market power achieved through the only means that escaped the leveling effect of market forces, viz., government action arbitrarily obstructing competition.

It followed that monopoly was first and foremost a political, rather than theoretical problem, and that, accordingly, competition as freedom to trade (that is, absence of market power) received less attention in the classical system than competition as freedom of contract (that is, absence of government interference). Granted the latter, the former always subsisted and, what mattered more, could always police itself (because in a free market no monopolistic position could arise or survive). This explains why, as Stigler noted, these economists devoted little energy to the analysis of an issue, market power, they regarded as at most temporary and, in any case, fully within the sway of competitive forces.

## 4.2. Monopoly as market outcome

After the Civil War the structure of the American economy witnessed a dramatic change, pushed by an enormous acceleration in capital accumulation. Between 1869 and 1889 the average American factory doubled in size, capital invested per manufacturing worker almost trebled and total factory productivity grew at an exponential rate. Industry also introduced powerful new technologies. The outcome was impressive. Measured by per capita growth of income, tangible wealth, amount of savings, and investment funds, the 1880s was the decade that enjoyed the most rapid growth in US history to date.<sup>12</sup>

The rise of big business was the *leitmotif* of Gilded Age socio-economic discourse – and the toughest theoretical challenge for US economists. Big business brought to center stage the relation between firm size and competition. Size meant scale economies and increasing returns, in themselves positive phenomena, but which led to even greater concentration. The latter could take the form of either monopolization or various types of combination, like cartels, trusts, mergers, etc. The flip side of this revolution were more extreme competitive conditions and a greater fragility of financial and managerial structures, the evolution of which lagged behind that of capital and technology. Unable to sustain the competitive pressure of big firms, small business-

<sup>12</sup> For this information, see Sklar 1988, 44; Jensen 1993, 834; Perelman 2006, 70-71.

es in several industries had to shutdown and exit the market. Moreover, the American economy became exposed to sudden financial downturns, like the Panics of 1873 and 1893. The consequences from the socio-political point of view were as dramatic as those of the productivity boom. While the latter had long-term positive effects on aggregate welfare, the high costs associated with the now rapid obsolescence, if not abrupt annihilation, of human and physical capital generated substantial hardship and complaints.

American economists found themselves in the difficult situation of having to reconcile their classical view of competition with the multiple effects of rising business concentration. In the classical model, free competition meant either freedom to trade or freedom of contract. Yet, under the new industrial conditions of big business and huge fixed costs, the outcome of such freedoms was no longer a fluid and atomistic market structure of the kind envisaged by classical economists. Now *laissez faire* generated a clustering and persistence of economic power. The market mechanism, if left free to work, no longer guaranteed its beneficial effects in terms of power dispersion and, consequently, widespread justice and prosperity. On the contrary, leaving businesses free to compete in the presence of massive fixed costs led to ever more concentration.

Shockingly, American economists were discovering that market power could well be the natural product of competitive markets and that, under certain conditions, the inevitable outcome of free competition was monopoly, i.e., the demise of competition itself. This new awareness meant that what jurists called privilege, and economists monopoly, now had another source beyond state grants. What was worse, it was an endogenous source, the roots of which lay in the market mechanism itself.

Most US economists thus realized that a whole rethinking of the classical notion of competition was called for. A significant share of them – those gathered in the so-called “new school”<sup>13</sup> – became persuaded that in the new industrial landscape *laissez faire* favored the consolidation of market and economic power, as well as the rising inequality of wealth and the greater instability of the markets themselves. Eventually, these economists argued, the utmost liberty to compete played *against* individual freedom. Far from promoting social welfare and personal autonomy, *laissez faire* hampered them (McCann 2012, 223). In fact, competition itself could have disastrous consequences.

<sup>13</sup> This was the name of the group of young economists who founded the American Economic Association in 1885 in open opposition to classical economics. Fine 1956, Ch.7, is still a good introduction to the “new school”. For more modern accounts, see Yonay 1998, Chs.2-3, and Bernstein 2001, Ch.1.

## 5. Ruinous competition and freedom from contract

Already the leading pre-Civil War form of big business, railroads occupied center stage in the transformation of the American economy. Investments in the railway industry were huge: they surpassed aggregate investment in manufacturing for every decade between 1850 and 1890, while the book value of railroad capital exceeded aggregate capital for the entire industrial sector until the early 20<sup>th</sup> century. Railroads also provided the most significant example of the difficult reconciliation between the classical, idyllic view of socially beneficial competition and the reality of ruthless market rivalry, anti-competitive combinations and sudden financial collapses. The cheapest possible transportation could only be achieved by those railroads capable of fully exploiting their massive scale economies, but, as several authors recognized, this entailed a tendency to monopoly in the industry that was incompatible with classical competition.

Yale economist Arthur Hadley earned international fame for applying in his 1885 book *Railroad Transportation* the lessons of the railroad industry to the economy in general. Hadley observed that in the presence of significant sunk costs the classical idea that capital could be easily withdrawn from unprofitable uses and re-invested in more profitable ones was meaningless. This was obviously the case with railroads. But in the economic environment of the Gilded Age ever more industries looked like railroads, i.e., they were characterized by “a large permanent investment, which can be used for one narrowly defined purpose, and for no other. The capital, once invested, must remain. It is worth little for any other purpose” (Hadley 1885, 40).

Beyond the slow readjustment of bad investments, two further problems affected those industries. The adoption of new technologies and sheer scale effects combined to lower the production cost below current market price, generating supra-competitive profits and attracting new capital. As a consequence, industries characterized by huge fixed investments and increasing returns frequently experienced situations of excess capacity and overproduction. Yet, contrary to what classical economists maintained, capital in those industries was hardly fluid: as a matter of fact, it was very difficult to withdraw or redirect it. Hence, overproduction was a semi-permanent phenomenon in many heavily-capitalized sectors of the economy.

A big debate arose among American economists about the nature and permanence of the overproduction caused by irreversible investments in fixed capital – what Herbert Hovenkamp has called “the great fixed-cost contro-

versy” (Hovenkamp 1991, 311). Some scholars remained faithful to the classical thesis that, thanks to the working of competition, general gluts or enduring resource misallocations were impossible. Others, including Hadley, believed that the presence of enormous sunk costs undermined the classical mechanism: given the conditions of modern capitalism, overproduction emerged as an endogenous and chronic tendency of the market.

Economists in the latter group also shared a second, even more serious concern. Speaking of the railroads’ experience, but again only to instance a more general problem, Hadley noted: “In order to attract new capital into the business, [railway] rates must be high enough to pay not merely operating expenses, but fixed charges on both old and new capital. But, when capital is once invested, it can afford to make rates hardly above the level of operating expenses rather than lose a given piece of business” (Hadley 1886, 223). In modern terms: the price capable of attracting new investment in industries characterized by massive sunk costs must be above average *total* cost, but this is always higher – often much higher – than the price forcing existing firms to withdraw old investment, namely, a price below average *variable* cost. In Hadley’s smart synthesis: “the rate at which it pays [for capital] to come in is very much higher than the rate at which it pays to go out” (ibid.).

The situation turned problematic in view of the downward pressure on price that was unavoidable in any sector affected by persistent overproduction. The loss to be suffered (in terms of lost capital) by stopping production and exiting the market was so large that firms preferred to fight until the end and keep producing as long as price exceeded average variable cost, and maybe even below that. Such a low price would eventually drive those firms into bankruptcy. Contrary again to classical economics, in these industries the free entry mechanism stabilized neither the price nor the return on investment around their normal levels. In fact, no “normal” limit to competition existed in the presence of large sunk costs. Forced to compete by cutting prices, firms were inevitably destined to collapse. Competition was simply ruinous.

Several American economists shared Hadley’s opinion, claiming that competition was indeed destructive in industries with huge fixed costs. Cornell economist Jeremiah Jenks was exemplary in this regard. As he wrote: “competition will eventuate, not in the elimination of some few while the majority are still making profits, but rather in a depression of the entire business, so that only the very few most skilful or best situated will be making any profit at all, while the others still struggling along may be losing money for a long period before they finally yield. Indeed, the result may well be that for a considerable length of time all will be running at a loss” (Jenks 1900, 19). These

industries, Jenks and Hadley concluded, naturally gravitated away from competitive conditions. Competition often led to the survival of only one firm, i.e., a monopoly, and thus to the end of competition itself.

Monopoly could be either the spontaneous outcome of ruinous competition or the deliberate goal of so-called *cutthroat competition*. The latter was a strategy addressed to preserving a firm's invested capital by waging a price war: drastic price cuts were undertaken, aimed at forcing competitors to leave the market; once it remained the lone producer, the firm could exploit its plant fully again. Clark (1886, 120) so portrayed the strategy's result: "Easy and tolerant competition is the antithesis of monopoly; the cut-throat process is the father of it".

A third alternative existed, *combination*, which could take various forms (from loose deals to tight cartels to complex governance structures), but the essence of which always lay in an agreement between producers to act as a joint monopoly and set "fair" prices, i.e., non-competitive prices capable of guaranteeing an adequate return on capital to all participants. Such an artificial, rather than natural limit to competition, was an appealing alternative, free as it was of the wasteful consequences of ruinous or cutthroat competition. Yet, it remained that in all three cases a significant reduction in the number of actual competitors or, frequently, a fully-fledged monopoly would follow.

American economists could not escape the recognition that, far from being just the product of government interference in the marketplace, as the Classics believed, monopoly seemed ubiquitous in the new industrial era. It could emerge as the natural outcome in an industry characterized by enormous fixed costs, or as the product of a ferocious competitive struggle, or as the smooth escape from competition itself, in the form of a trust, a cartel or a merger. "New school" leader Richard Ely felt it legitimate to conclude that competition in the presence of large fixed costs was self-destructive and *inevitably* led to monopoly (Ely 1888, 121). In a similar vein, Jenks proclaimed: "Under a system of free competition industrial efficiency tends toward monopoly. The business genius whose industrial efficiency is greatest tends to overcome his rivals, and to take over a continually increasing proportion of the business, until he becomes a monopolist" (Jenks 1912, 349). Unlike the classical model, monopoly had therefore another possible source beyond state interference, a source dependent on the strict logic of competition paired with the concrete existence of giant fixed costs – a feature the Classics had little experience with. As it turned out, privilege was not only state-granted, but could have a pure economic origin. It could, in short, be *economic privilege*. Worse, this

second source entailed that monopoly was not only possible, but practically inevitable.

Given the mayhem caused by destructive or cutthroat competition in the context of heavily-capitalized businesses, many economists were not hostile to what looked like the least harmful solution. Neoclassical champion Irving Fisher argued that combination was a legitimate form of self-defense for a firm's investment: "The rise of trusts, pools, and rate agreements is largely due to the necessity of protection from competition, precisely analogous to the protection given by patents and copyrights" (Fisher 1997 [1912], 331). Even Clark recognized the private and social attractiveness of combinations.<sup>14</sup> As to Hadley, he took a fairly positive view of railroad combination, or pooling, which he saw as a market-based solution to the industry's peculiar problems – a solution, that is to say, preserving a role for market forces and voluntary contractual behavior with no outside coercion by the law (Hadley 1885, 76-7). Railroad pooling looked to him as the only way the stability of so crucial an industry could be reconciled with the true core of classical economic freedom, namely, freedom of contract and freedom from government coercion. The latter could thus survive and preserve the gist of free competition, classically understood, even when competition as freedom to trade had to give way, for objective industrial reasons, to concerted behavior.

Starting from the 1870s, destructive competition began to appear regularly as a courtroom defense raised by firms accused of having formed unlawful cartels or combinations.<sup>15</sup> Lawyers, as well as economists, argued that competition – or, more specifically, a low price – could be such a bad thing that firms could invoke their right to protect themselves against it. The novelty was not so much in the legal argument itself – the "ruinous competition" defense had been raised long before by participants to manifold contracts in restraint of trade, usually paired with appeals to custom and fairness. The new feature was that now the argument had found legitimization in economic analysis.

Beyond theoretical and judicial issues, the further dilemma arose as to what policy-makers should do with ruinous competition and inevitable monopoly. Fighting the new form of economic privilege meant interfering with the market mechanism, but letting it go undisturbed meant favoring the eventual demise of that very mechanism. Even worse, any business endowed with

<sup>14</sup> See e.g. Clark 1887, 55 and 57.

<sup>15</sup> See Hovenkamp 1991, 313. On the illegality of cartels and combinations, see below.

economic privilege – viz., monopoly power – had also the ability to interfere, one way or another, with other businesses' freedom to trade and contract. Significantly, this power to interfere often stemmed from the powerful firm's (or group of firms') contractual activity. Even those jurists and economists who identified *laissez faire* with freedom of contract, rather than with freedom to trade, had therefore to recognize that, in a world of rising concentration and business power, a central issue was how to guarantee freedom *from* contract.

By the latter expression is meant the opportunity for every market participant to pursue her economic activity – her freedom to trade and compete – without being coerced, either vertically or horizontally, by other firms' activity, which often takes the shape of formal or informal contracts.<sup>16</sup> Think, for example, of a firm whose freedom to trade is affected by its rivals' market power – more specifically, by the contractual restraints they can enforce. The typical case is that of a cartel, that is, a “contract” among some firms with the goal of restraining competition and, therefore, the freedom to trade of all non-participants: customers, competitors and suppliers alike. Classical economists believed that no such restraint could survive the competitive pressure brought by free entry and full resource mobility. Alas, no such equilibrating forces were available in industries impaired by ruinous competition and inevitable monopoly. State intervention could thus be invoked in such cases to preserve a firm's freedom to trade and, with it, the possibility of competition itself to survive. Remarkably, the plea did not exclusively, or did not necessarily, come from anti-business quarters. Even conservative supporters of freedom of contract might advocate government action for the simple reason that the constraints imposed by a powerful business affected not only a firm's freedom to trade but also, more directly, its freedom of contract (think for instance of the limitations suffered by a cartel's supplier).

By the last decade of the 19<sup>th</sup> century, *freedom from contract* became a catchword for everyone demanding that some fetters be placed upon unbridled competition. For many of those who appealed to it, it was synonymous with freedom to trade and compete. For others, it was a way to recall the importance of protecting the universal right of freedom of contract against any undue interference. For all, it meant urging government action to guarantee the

<sup>16</sup> Note that in modern economic jargon one would talk rather of freedom from market power (see above). In the absence of a recognized measure of market power (which will only emerge in the 1930s: see Giocoli 2012), late 19th-century scholars referred to such power's closest proxy, namely, the contracts in restraint of trade that only a powerful firm (or group of firms) could enforce.



persistence of competitive conditions in the new industrial landscape. The coercion of the law was invoked to block those particular business contracts or practices that could be used by powerful firms or groups to encroach someone else's economic freedom. Little surprise, then, that from these ranks emerged the voices in favor, first, of state and federal regulation; then, of a federal anti-trust statute; and, finally, of the latter's strictest enforcement.

## 6. The regulatory dilemma

Stephen Siegel explained long ago how the Gilded Age and Progressive era controversies about the power and limit of regulation stemmed from the transformation of the US economy from one based on small-scale, decentralized, individual businesses to one where large-scale, concentrated, corporate enterprises prevailed (Siegel 1984, 188). This revolution altered the "property v. privilege" dichotomy that characterized American law. Property used to denote only those valuables the acquisition of which was open to everybody, typically through work and saving. In this sense property stood in contrast to privilege, which was identified with "those assets that only specially designated individuals could acquire, characteristically through special governmental act" (ibid., 190).<sup>17</sup> Privilege endangered equality, harmony, freedom, and the overall economic and political liberty that property established. It corrupted government in particular, directing it from the pursuit of the general good to that of individual self-interest.

The conventional reading was questioned by the post-Civil War advent of large-scale business – a phenomenon that, as we just explained, took place thanks to, rather than despite, market forces. Now economic privilege – viz., monopoly – did not necessarily stem from government intervention. New economic conditions made market dominance possible through private contracts, rather than governmental grants. The dilemma arose as to what to do in the face of the free play of market forces. To many observers, especially outside the economics profession, the *laissez faire* recipe appeared counter-productive; in a shocking reversal, government regulation was now invoked by powerless individuals as a curb against rising monopoly power.

Aware of the multiple and complex sides of the regulatory dilemma, American jurists endeavored to handle it by redefining no less than the substance

<sup>17</sup> On the "property v. privilege" dichotomy, also see Siegel 1986-87, 55-75.

of constitutional law. Their privileged response, Siegel argued, was to *constitutionalize the free market* (ibid., 260). In this respect, *Lochner* was only one instance of a broader trend, which is best illustrated by the regulation of railroads.

As the most important American industry in terms of employees, capitalization and social influence, railroads were *the* issue as far as regulation was concerned. No surprise, then, that the first comprehensive regulatory measure passed by Congress, the 1887 Interstate Commerce Act, was designed to regulate the railroad industry. Indeed, by the time of its enactment, states had been regulating railroads for the previous half century and courts had been called to adjudicate law several times in regulatory conflicts between the states, the railroads and other parties with an interest in railway services.

Alas, railroad regulation was no easy task. During the Gilded Age the problem had at least two facets: how should railroads be regulated? Who, between the states or the federal government, should regulate them? The latter facet was as delicate as the former – possibly more so, in that it touched the constitutionally sensitive issue of the division of power between Washington and the states. The stakes were so high, both economically and constitutionally, that the jurisprudential answers given to both sides of the railroad regulation problem would eventually shape all other areas of regulatory activity in the US.

Admitting regulatory power over rates meant granting legislatures full control over the value of property. No surprise, then, that even when everybody recognized the legitimacy of such power, as in the case of railroads and other privileged businesses, many jurists still balked at the idea of placing vast amounts of private wealth, accumulated in the nation's most significant enterprises, at the mercy of state legislators. Still, for almost a decade the ruling precedent on regulatory matters was *Munn v. Illinois* (94 US 113, 1877). In sanctioning state regulation of a business (grain elevators) that enjoyed no legal privilege, Chief Justice Morrison Waite had expressly rejected the counsel's argument that the value of property – viz., its profitability – should receive the same degree of constitutional protection as its title and possession. The regulatory power of the state legislature, Waite's opinion held, was complete and exclusive, so much so that courts had no right to question the reasonableness of regulated rates. Protection against abuses of the regulatory power, he famously concluded, was to be found in "the polls, not [the] courts" (*Munn*, at 134).

As Siegel (1984, 210) explains, *Munn* was clearly anachronistic. By the late 19<sup>th</sup> century, non-landed wealth had achieved a significance it could not have held in the early decades of the Republic, when the main form of property was

land and the constitutional protection over title and possession sufficed to guarantee individual freedom from external interference. In modern economies, largely based upon immaterial wealth, value had to be recognized as an essential element of the constitutional conception of property. The Supreme Court recognized the problem and reneged on this side of *Munn*.

In the so-called *Railroad Commission Cases* (1886), Waite himself made a first move in this direction. “It is not to be inferred”, he averred, “that this power of limitation or regulation is itself without limit. *This power to regulate is not a power to destroy, and limitation is not the equivalent of confiscation.* Under pretense of regulating fares and freights, the state cannot require a railroad corporation to carry persons or property without reward; neither can it do that which in law amounts to a taking of private property for public use without *just compensation* or without due process of law”.<sup>18</sup> Three years later and in another railroad rate regulation case, the Court finally ruled that the Constitution mandated judicial review of the “reasonableness” of regulated rates. Justice Samuel Blatchford declared in *Chicago, Milwaukee & St. Paul Railway*: “The question of the reasonableness of a rate of charge for transportation by a railroad company, involving as it does the element of reasonableness both as regards the company and as regards the public, is eminently a question for judicial investigation, requiring due process of law for its determination”.<sup>19</sup>

The 1890 decision marked a turning point in American constitutional jurisprudence on regulatory matters. In a complete overturn of *Munn*, the Court now recognized that rate regulation was *not* beyond judicial review: the Constitution both allowed legislative control of the charges of privileged enterprises and required courts to substantially protect from spoliation the private wealth invested in them. The new doctrine did not just affect the (copious) case law on railroad rates. At the price of expanding the boundaries of judicial discretion, it established the reasonableness scrutiny over all applications of regulatory and, more broadly, police power. Needless to say, economic theory had a lot to say about how such scrutiny should be made.

In a series of papers,<sup>20</sup> I have reconstructed how at the turn of the 20<sup>th</sup> century federal courts and regulators applied classical and not-so-classical economic ideas (most notably, the notion of present value and reproduction

<sup>18</sup> *Stone v. Farmers' Loan & Trust Co.*, 116 U.S. 307 (1886), at 331, emphasis added. On this decision, see Smalley 1906, 27-8.

<sup>19</sup> *Chicago, Milwaukee & St. Paul Railway Co. v. Minnesota*, 134 U.S. 418 (1890), at 458. See again Smalley 1906, 32-8.

<sup>20</sup> See Giocoli 2016; 2017a; 2017b.

cost) to handle the manifold prongs of the regulatory conundrum. Economists in turn reacted to this jurisprudence, as well as to legislative proposals for tighter regulation, by investigating more thoroughly the complex subject of railroad pricing. Some of their discoveries still hold sway today, the theory of so-called Ramsey pricing being the most conspicuous example (see Hadley 1885, Appendix II; Taussig 1912). Still, regulation was not the only arena where economists and jurists were summoned to deploy their skills by the brand new world of Corporate America. Even more significant than regulation was the newborn subject of antitrust law.

## **7. Contracts in restraint of trade and American law**

This section sketches the principles followed by American courts when administering the common law of contracts in restraint of trade (CRTs henceforth), large parts of which had been developed by their English counterparts. The key insight is that, at least until the Sherman Act, a common pattern existed in the way judges administered the law of CRTs – which also encompassed combinations – in pursuit of antitrust goals. In particular, a steady shift of emphasis from freedom to trade to freedom of contract characterized Anglo-American jurisprudence throughout the 19<sup>th</sup> century.<sup>21</sup>

The common law of CRTs involves two contrary principles. The dichotomy – which mirrors that between the two freedoms – explains why over the last three centuries there has never been a monolithic attitude towards these contracts. Arguments over the distinction between good and bad restraints have been a perennial feature of common law scholarship and jurisprudence.

On the one hand, the common law may uphold a CRT in the name of contractual liberty. To own property means the right to freely dispose of property by contract; this freedom covers property in all its forms, including intangible property such as goodwill or trademarks, and the law should not interfere with the way a free and grown-up individual exercises her right. On the other hand, the common law may invalidate a CRT in the name of freedom to trade. The latter expression may mean either a given individual's liberty to exercise her own commerce or profession or a generic reference to everyone else's opportunity to enter a given trade. In the first, more traditional sense, CRTs are

<sup>21</sup> For a more detailed analysis, see Giocoli 2014, Ch.2, from which this section draws.

indicted at common law because they deprive an individual of the means to earn her livelihood. In the second meaning, which has gained increasing relevance in England since the 18<sup>th</sup> century, the law condemns CRTs because they deprive the public of the advantages of competition. We find in this conflict the germs of what would later become the key dilemma of antitrust enforcement, namely, the divergence between competition as freedom from market power and competition as freedom of contract, i.e., the liberty to exercise one's own property rights.

Writing for the Supreme Court in *Oregon Steam*, Justice Joseph Bradley formulated in 1873 what became the American legal orthodoxy on CRTs. "There are two principal grounds on which the doctrine is founded that a contract in restraint of trade is void as against public policy", Bradley averred. "One is the injury to the public by being deprived of the restricted party's industry; the other is the injury to the party himself by being precluded from pursuing this occupation and thus being prevented from supporting himself and his family. It is evident that both these evils occur when the contract is general not to pursue one's trade at all, or not to pursue it in the entire realm or country. The country suffers the loss in both cases, and the party is deprived of his occupation or is obliged to expatriate himself in order to follow it. A contract that is open to such grave objection is clearly against public policy. But if neither of these evils ensues and if the contract is founded on a valid consideration and a reasonable ground of benefit to the other party, it is free from objection and may be enforced".<sup>22</sup>

The Court believed that contracts restraining the freedom to trade were harmful both because they deprived an individual of his occupation and because they deprived society of that individual's activity. General CRTs suffered from both evils and were therefore void. But, the Court conceded, reasonable contracts did exist that could well be enforced, despite their restraining effect. This was especially true in the case of contracts pertaining to corporate life and activities. *Oregon Steam* thus provides a useful benchmark for understanding how, before the Sherman Act, American competition policy consisted of a combination of corporate law and the common law of CRTs.<sup>23</sup>

It is easy to realize the extent to which the classical view of competition shaped Bradley's doctrine. Classical competition emphasized contractual lib-

<sup>22</sup> *Oregon Steam Navigation Co. v. Winsor*, 87 U.S. 64 (1873), at 68. The doctrine mirrored the akin principle affirmed in English common law since the early 18th century.

<sup>23</sup> See Hovenkamp 1989, 1029.

erty and freedom from coercion, while giving no importance to a competitive market structure in the modern, neoclassical sense. So, for instance, a voluntary price-fixing agreement could never be “anticompetitive”, in the classical logic, if nobody’s freedom to act was artificially restrained. This explains why for the good part of the 19<sup>th</sup> century American law did not proscribe cartels and mergers. Price-fixing agreements could at most be unenforceable at common law, but they were not illegal, nor actionable by outsiders.

That the common law of CRTs did not aim at preserving, or establishing, a competitive market structure (neither in the US nor, for that matter, in Britain) was also clear from the way courts enforced it. Consistently with the classical model, American courts treated those CRTs that actually destroyed competitors exactly like those that did not. The key distinction lay elsewhere, namely, in the separation, of English origin, between general and partial restraints – the former being, under American common law, those covering for instance an entire state. Given this distinction, and consistently with Bradley’s doctrine, many CRTs were viewed unfavorably by US judges.

Hostility to CRTs went hand-in-hand with a relaxed attitude about cartels.<sup>24</sup> Within the classical model, a price-fixing conspiracy with no exclusionary practices directed at nonparticipants was not particularly offensive. Classical economists posited that monopoly prices could never persist *unless artificial restraints prevented new competition*. Apart from law-created obstacles, restraints to entry could take the form of either a contract, where the restricted business was a voluntary participant, or a combination, which was by definition directed at restraining other nonparticipant businesses. A mere agreement among sellers to fix prices was therefore of little concern, provided that neither the law nor the price-fixers prevented other firms from entering the market. If cartel members sought to charge monopoly prices, new competition would immediately frustrate their attempt.

Accordingly, the few common law cases that did condemn cartel agreements were founded on the defendants’ efforts to exclude, or coerce, outsiders. The latter, note well, never included customers. That buyers be required to pay a higher price was considered no coercion at all, because in the classical model patrons were always free to walk away and purchase elsewhere. In the revealing words of a late 19<sup>th</sup>-century New York jurist, the principle was that “the public is deprived of no legal right, unless some individual is deprived of a legal right; [...] What is this so-called right of ‘the public’, as to freedom of

<sup>24</sup> See Hovenkamp 1989, 1036-7.

competition? Who is there, that has the legal right – that two sellers of merchandise shall compete?” (Stickney 1897, 157).

Starting from these benchmark doctrines, American law gradually accommodated a more pro-business and pro-freedom-of-contract approach to CRTs. Fewer and fewer contracts were deemed unlawful at common law. Even general restraints, or contracts unlimited as to time and place, became reasonable in the eyes of Gilded Age courts. It was for instance recognized that those CRTs that just embodied simple noncompetition agreements accompanying the sale of a business were of no public concern in a competitive market. So, for instance, in an 1887 case the New York Court of Appeals concluded that a simple covenant not to compete could not create a monopoly and, therefore, had to be sanctioned; this in view of the fact that “the business is open to all others, and there is little danger that the public will suffer harm from lack of persons to engage in a profitable industry”. Fifteen years later, after the Sherman Act had been enacted and, above all, the jurisprudential transition towards a more lenient approach to CRTs had been completed, the same court was even more explicit: “Contracts between parties, which have for their object the removal of a rival and competitor in a business, are not to be regarded as contracts in restraint of trade. They do not close the field of competition, except to the particular party to be affected”.<sup>25</sup>

By the late 19<sup>th</sup> century, the blending of the common law of CRTs with traditional corporate law provided states with a set of powerful instruments to combat the proliferation of collective business agreements and the tremendous rise of industrial concentration. Corporate law already featured the possibility of enjoining corporate misbehaviors as *ultra vires*, i.e., beyond chartered powers. Moreover, before 1890 more than a dozen states had already included antimonopoly provisions in their constitutions or enacted their own antitrust legislation, often as an integral part of state corporation law. Violations were punished with the withdrawal of corporate charters or similar special privileges.<sup>26</sup> Some states had even forerun the Sherman Act by bringing action against some of the biggest national trusts, like those of sugar and oil. The states won all these litigations, leading to either annulment of defendants’ franchises and privileges or to forced severance of trust connections.

<sup>25</sup> *Diamond Match Co. v. Roeber*, 106 N.Y. 473 (1887), at 483; *Wood v. Whitehead Bros*, 59 N.Y. 545 (1901), at 551.

<sup>26</sup> On pre-Sherman Act antitrust activity at state level, see May 1987, esp. 499-501.

When combined with the common law of CRTs, state corporation law was considered as a powerful shield against the threatening economic transformations of the time. This view was shared by learned scholars and public opinion alike.<sup>27</sup> Legal treatises and surveys emphasized state statutes and decisions, giving them at least equal, and in some cases greater attention than federal developments. For its part, the public followed state antitrust cases with much interest, stimulated by their intriguing mix of economic, social, and political dimensions. The trust problem was headline news, with newspapers featuring coverage of law enforcement activity and litigation developments in the various states.

This was the environment in which most of the congressmen debating the Sherman Act expressed their confidence in state authority and the common law. Their main concern was how to strengthen what states were already doing. Hence, in voting for the new statute, they manifested a desire to merely supplement, and not hold back, state antitrust activity and the work of state courts. The Sherman Act simply aimed to sustain the antitrust efforts by the states by means of a federal statutory rendition of common law principles.

## **8. Senatorial competition**

The common law of CRTs was the main reference for the framers of the Sherman Act. The Act's wording, with the use of terms like "restraint of trade" and "attempt to monopolize", is self-explanatory in this regard. However, the common law roots of the 1890 statute do not tell the whole story of its enactment, nor suffice to answer the most basic question arising in the courts called to enforce it, namely, whether the Act embodied or superseded the common law. Should the offenses identified in the Act, namely, contracts, combinations and conspiracies in restraint of trade (as to section 1) and monopolizing attempts (as to section 2), be given the traditional common law explanation, or did the new statute redefine them in a substantive sense?

Congress turned to the trust problem in 1888. The electoral platforms of both major political parties contained statements indicating opposition to industrial monopolies and combinations. Initial attitudes in the 50<sup>th</sup> Congress were thus in favor of federal intervention because states seemingly lacked sufficient powers to check the growth of business concentration. However, as

<sup>27</sup> See May 1987, 504; Freyer 1992, 50-52.



we argued before, by the time the first legislative proposals came out, a score of successful cases had established the states' formal competence, and effective control, on the issue. This circumstance would be crucial in the process leading to the statute's approval.

Congressional debates on the Sherman Act once again validate our dichotomic characterization of free competition. Records show that both views, freedom of contract and freedom to trade, were openly on the table. The eventual outcome was an Act that featured elements of both. By adopting the common law language, it endorsed the freedom of contract approach that, inspired by classical political economy, had come to dominate the late 19<sup>th</sup>-century common law of CRTs (as always, the term also included combinations). But the Act also contained elements of the alternative approach, that is, of the idea that only court interventions could restore the proper working of competition in the face of excessive market power. In particular, by making violations actionable by either third parties or the government and by establishing civil and criminal sanctions against violations, the Act marked a procedural sea-change with regard to traditional common law.

The dichotomy between freedom of contract and freedom to trade thus provides a useful guide in interpreting the events leading to the statute's approval.<sup>28</sup> As is fairly well known, the enacted version of the Sherman Act was *not* authored by Ohio Senator John Sherman, but mainly by Vermont Senator, and prominent corporate lawyer, George Edmunds, chairman of the Senate Judiciary Committee.<sup>29</sup> It is not simply that the statute should more properly be called the "Edmunds Act". The real point is that the difference between Sherman's various proposals and Edmunds's final version exactly matches that separating the two views of free competition.

Sherman's original bill, titled "A Bill to declare unlawful trusts and combinations in restraint of trade and production", had been drafted on the assumption that the states were "unable to deal with the great evil that now threatens us", the rise of industrial concentration. The Ohio Republican warned his colleagues against the trust-builders' ability to evade the jurisdiction of state courts. It was up to Congress, therefore, to employ federal powers to dissolve combinations "extend[ing] to two or more States", while state officials should

<sup>28</sup> The whole debate is reproduced, with just a few lacunae, in *Bills and Debates in Congress Relating to Trusts, 1888-1902*, from which all the following quotations are drawn.

<sup>29</sup> For a more detailed reconstruction of the episode, see Letwin 1965, 93-5; McCurdy 1979, 325-6.

continue to police the charters of those active within a single jurisdiction. No doubt existed as to Congress's constitutional power in the field because the effect of trusts was to "restrain commerce, turn it from its natural courses, [and] increase the price of articles". All these activities, Sherman suggested, negatively affected interstate trade in the same way as a bridge obstructs interstate navigation or a state tax impedes the free circulation of goods – two realms where Congress power was undisputed.<sup>30</sup>

Clearly, Sherman had in mind a model of antitrust involving a direct control over market structures by federal and state governments. Much as a government could decide whether and where to build a bridge, so it should be allowed to determine the extent of competition in a given market. And exactly like unobstructed navigation or tax-free commerce, so free competition should be the default mode of operation. Accordingly, all of his proposals – he made three of them, between August 1888 and March 1890 – contained (a version of) the following statement: "That all arrangements, contracts, agreements, trusts, or combinations between persons or corporations made with a view, or which tend to *prevent full and free competition* in the production, manufacture, or sale of articles of domestic growth or production, [...] and all arrangements, contracts, agreements, trusts, or combinations between persons or corporations designed, or which tend, to *advance the cost to the consumer* of any of such articles, are hereby declared to be against public policy, unlawful, and void" (*Bill and Debates*, 8, emphasis added).<sup>31</sup> Regardless of Sherman's own political goals – on which much has been written<sup>32</sup> – this wording placed his bill firmly within the freedom to trade rhetoric. Both the notions of "full and free competition" and that of consumers paying the cost of anti-competitive behavior could find no place in the late 19<sup>th</sup>-century common law approach to CRTs. They were, on the contrary, consistent with a characterization of antitrust as committed to defend competition from market and, possibly, also economic power.

The enacted version of the statute would contain no such wording, including no mention at all of the term "competition". Senator Edmunds would draft it explicitly in the standard common law language of contracts and combinations in restraint of trade and monopolizing attempts, a phrasing justifying Justice Oliver W. Holmes's later remark that "the [Sherman] act says nothing

<sup>30</sup> See Sherman, *Bills and Debates*, 105.

<sup>31</sup> For further details on Sherman's bills, see Letwin 1965, 87-91, and Sklar 1988, 107-9.

<sup>32</sup> See for all Bradley 1990.

about competition”.<sup>33</sup> Were it not for its revolutionary procedural provisions, the Act could be plainly interpreted according to the common law (and, eventually, constitutional) doctrines of property rights and freedom of contract typical of English and American courts of the time.

The congressional debate triggered by Sherman's proposals illustrates the division between the two camps. The bill met a less-than-friendly reception in the Senate. Doubts arose about both the constitutionality and the necessity of so sweeping a reform. Several senators contended that Sherman's equation of the trusts' impact on interstate trade with other kinds of interference, like obstruction to navigation and state tax barriers, was legally unfounded, given the Supreme Court's consolidated distinction between transportation and manufacturing. Others asserted that, far from being helpless, states had recently demonstrated their legal capacity to handle structural problems caused by giant combinations. But the most interesting aspect of the debate turned on the two views of competition. A few glimpses will suffice.<sup>34</sup>

Sherman and other supporters of the original bill emphasized two negative consequences of combinations: harm to industrial liberty and harm to consumers. The former threat struck at the traditional image of a country of “small dealers and worthy men”, viz., of a fluid and atomistic economy of small businesses, none of which were endowed with significant market power. The latter struck at the possibility of consumers escaping from high prices by turning to alternative producers. Taken together, these two harms led to an even more dangerous threat against American society. Evoking Thomas Jefferson's republican ideal, Sherman emphasized how an economically independent citizenry was the cornerstone of representative government. Competitive equality in the marketplace was therefore crucial to preserve not only economic, but also political liberty and, ultimately, democracy itself. Failure to answer the citizens' demand for federal action against the combination ogre would risk opening the door to “the socialist, the communist, and the nihilist” (Sherman, *Bills and Debates*, 101).

Remarkably, the opposing camp was not manned by naïve believers in the power of *laissez faire* to automatically destroy monopoly. Most of those against Sherman's proposal actually believed – like “the younger economists of the country who have studied the question thoroughly” (von Halle 1896, 117) – that “full and free” competition could be even more dangerous than

<sup>33</sup> *Northern Securities Co. v. United States*, 193 US 197 (1904), at 403.

<sup>34</sup> For more details, see Letwin 1965, 85-99; Sklar 1988, 105-17; Peritz 1996, 14 ff.

combination and that private agreements could often mitigate the effects of destructive competition. They subscribed to the recent view of “new school” economists, who held that business concentration and monopoly were the inevitable products of competition in a modern economy and that such a natural, evolutionary outcome could not, and should not, be hindered by the government or the law. Many senators thus believed that, by preserving the most complete freedom of contract, the law could actually favor the birth of private agreements and combinations as a *safeguard* against the most destructive effects of competition. Like many economists of the time, they too thought that industrial liberty was not synonymous with unrestrained competition – that “unrestrained competition is not free competition”, as Rudolph Peritz (1996, 16) put it. Competition could only be termed “free” when market participants could exercise their most complete contractual freedom, including the freedom to voluntarily restrain one’s own market opportunities. The law should aim at preserving that freedom by proscribing only those contracts and practices that curtailed it – viz., that coerced an individual into adopting a behavior he would not voluntarily choose. But this was precisely the message of the late 19<sup>th</sup>-century common law of CRTs, itself inspired by classical political economy.

As it turns out, freedom of contract, that quintessential feature of classical thought and the pillar of Gilded Age jurisprudence, also provided the rationale for those who rejected any outright condemnation of business concentration and were ready to consider it as a natural, possibly beneficial evolution of market forces. So even a “new school” economist could see no contradiction in adopting classical principles, including the desirability of the utmost contractual liberty, while believing that new technological conditions had made the classical competitive order simply impossible. At the same time, it was perfectly possible to endorse freedom of contract as a constitutional principle, a common law doctrine and a milestone of classical political economy, while belonging to the opposing camp of those who believed the economic and political threats raised by business concentration outweighed its alleged benefits.

The extent of the common ground established by classical notions is revealed by what both camps of the Senate debate extolled as a key benefit of the market system, namely, the guarantee of a “fair price”. The true yardstick for assessing the desirability of a given price, fairness meant not only that consumers ought not to be exploited by trusts, but also – indeed, especially – that “every man in business” ought not to be deprived of his “legal and moral

right” to a “fair profit”.<sup>35</sup> All senators agreed that businessmen were entitled, as anyone else, to a fair price embodying the “just” reward for their honest work, where “just” was intended in the legal, and classical sense of respecting the property rights on the fruits of one’s own labor. “The true theory on these matters”, Senator Platt proclaimed, “is that prices should be just and reasonable and fair, that prices, no matter who is the producer or what the article, should be such as will render a fair return to all persons engaged in its production, a fair profit on capital, on labor, and on everything else that enters into its production” (Platt, *Bills and Debates*, 295).

Both sides of the debate would endorse these words. But only one side would assent to the rest of Platt’s statement. The two factions split again when it came to determining what the law and the government could do to guarantee the actual fairness of market prices. Sherman and his allies believed that a fair price could only be achieved when conditions of “full and free competition” prevailed in the marketplace. Hence, the law should strike at combinations like, say, price-fixing cartels, which generated an unjustly high reward for their members at the customers’ expense. The opposing view emphasized, as did also many economists of the time, the losses caused by ruinous competition, and saw combinations, including cartels, as voluntary contracts aimed at establishing fair prices and fair profits. The law should not prohibit these combinations, unless they undertook practices imposing unfair conditions upon someone else.<sup>36</sup>

Platt belonged to the latter camp. Unsurprisingly, he attacked “the theory of [Sherman’s] bill” for entailing that “no matter how much the price may have been depressed, no matter how losing the business may be, the parties engaged in it must have no understanding between themselves by which they will come together and say that they will obtain a fair and fairly remunerative price for the article which they produce. That is wicked, the bill says” (ibid., 295-6). For those on Platt’s side, no law should interfere with a businessman’s effort to defend his right to a fair profit. This was the logical consequence of considering that “right” as part of his property, i.e., itself a property right. As such, it should be in the owner’s full and free disposal. And as such, many courts would conclude, it was constitutionally protected and no federal statute should violate it.

<sup>35</sup> Connecticut Senator Orville Platt used these words while debating the final version of the bill: see *Bills and Debates*, 297.

<sup>36</sup> For example, a conspiracy to boycott a rival who refused to join the cartel should be declared unlawful because it coerced the rival’s freedom and caused him unfair losses.

The blending of contractual freedom with the property rhetoric made Platt's the winning position, and not only with respect to "fair price". Most congressmen saw in the combined action of state corporate law and common law an adequate weaponry to protect the American economy from both dangers – unrestrained combination *and* unrestrained competition. Accordingly, they relegated the Sherman Act to a supporting role. The phrasing of the first two sections in Edmunds's enacted version of the statute mirrored this view. The motion to commit the bill to the Judiciary Committee thus turned out to be the decisive event in the Act's legislative history: the Committee's recourse to familiar common law wording for defining statutory liabilities bears witness to its "reaction against the original bill's explicit and unmediated imposition of 'full and free competition' as the only natural and legitimate form of commerce" (Peritz 1996, 20).

In the end, the main outcome of the Senate debate over Sherman's proposal was to establish the constitutional *limitation* of any federal action against corporations and industrial concentrations. The Senate concluded that Congress "lacked the authority to vest federal officials with a roving license to 'bust' state-chartered corporations" (McCurdy 1979, 326). Still, in the debate about the final draft, no Senator attempted to specify what the federal authorities' role should be in fighting the trusts. Presenting the Committee's work, Edmunds simply explained that the bill incorporated traditional common law categories and that Congress ought to leave courts free to determine the categories' applicability on a case-by-case basis (Edmunds, *Bills and Debates*, 315). An unexceptional piece of legislation, in short, which senators could safely vote into law.

A few additional details only were provided in the discussion about the ambiguous notion of "monopolizing". CRTs necessarily involved an agreement between two or more parties, but the Judiciary Committee had added a provision extending section 1's prohibitions to target the restraints to trade caused by a single individual or firm. The problem with section 2 was the word "monopolize", which had no obvious correspondence in the common law. Edmunds was aware of the vagueness of the new provision, which could be read as a prohibition of all loosely-defined monopolies, and so as a radical attack against any big business endowed with significant market power. Hence, he tried to reassure perplexed senators that section 2 dealt with just another form of restraint of trade and would never be applied against any individual who had conquered a whole business by his own "superior skill and intelligence". The Committee, Edmunds insisted, had read "monopoly" as a technical term at common law, meaning "the sole engrossing to a man's

self by means which prevent other men from engaging in fair competition with him". In this reading, even "monopolizing" fell within the common law boundaries of "coercion", viewed as the limitation of someone else's freedom. It followed that no market position, no matter how large, could ever be called a monopoly if gained by "superior skill and intelligence".<sup>37</sup>

Thus reassured, an almost unanimous Senate voted the bill as proposed by the Committee, with a new title: "A Bill to protect trade and commerce against unlawful restraints and monopolies". As a courtesy to Senator Sherman, the Act retained his name, although its substantive content had entirely changed.<sup>38</sup>

Unsurprisingly, most American economists were less than excited about the new statute, which they deemed useless at best and dangerous at worst. Critical voices were hardly homogenous, though, in that they focused on different sides of the trust problem. Three main positions emerged.

As we know, "new school" economists believed that scale effects, better managerial techniques and massive capital investments increased productive efficiency so much that even full-blown monopolies could further social welfare. A few admitted just one proviso for the beneficial effect of size, namely, that the magic wand of potential competition be allowed to work properly. These economists thus viewed the Sherman Act as at best a misguided attempt that, in the hands of theoretically untrained courts, could twist or obstruct the natural evolution of markets towards higher concentration. Combination should be encouraged, not thwarted, as the most virtuous way to cope with the inevitability of monopoly.<sup>39</sup>

A second group remained faithful to the classical view of competition. These traditional economists singled out contractual liberty and freedom from government interferences as the key requirements for the competitive process. Faith in the power of the invisible hand led them to accept whatever outcome the spontaneous play of market forces might produce. Judicial in-

<sup>37</sup> For the lively exchange involving Edmunds and other senators on this specific issue, see *Bills and Debates*, 322-4. It was the very last point debated by the Senate before the final vote.

<sup>38</sup> And despite Sherman's own complaint to the *New York Times* that the final version was "totally ineffective in dealing with combinations and trusts. All corporations can ride through or over it without fear of punishment or detection". See Letwin 1965, 94.

<sup>39</sup> That in its early years the Sherman Act was enforced almost exclusively as an anti-cartel, rather than anti-trust statute and that, as a consequence, American firms were pushed towards tighter forms of concentration (the slogan being: "merge, don't collude") was therefore an outcome these economists could welcome.

tervention was only justified when someone's freedom had been coerced, but in general no grounds existed for interfering with voluntary contracts. Even this group thus concluded that no real need existed for an antitrust statute, if not as a way to federalize the established common law of CRTs. Any use of the new act beyond this narrow boundary would represent an illegitimate interference with economic freedom.

The rhetoric of “small dealers and worthy men”, which had played so big a part during senatorial debates, became the slogan for a third group. Populist commentators praised in principle the outcomes of the free play of market forces, but unlike members of the other two groups, they ascribed economic evils to the abandonment of old-style “fair” competition among roughly equal agents. In the new industrial era, the spread of monopoly power had disrupted the natural and socially beneficial harmony of free markets. Crucially, these populist writers – most of whom were not professional economists – believed that it was size itself, rather than any specific practice, that should be blamed for the disruption. Concern for this undesirable outcome went beyond economic effects. American democracy itself was threatened by the rising plutocracy and trustification.

While this group could invoke no solid theoretical argument in support of its diagnosis, its policy recipe was clear. It was up to the federal government to restore the markets' natural harmony by fighting, and possibly dissolving, big business. A rigorous antitrust law was instrumental to this task. Not surprisingly, the populists were the only declared supporters of Senator Sherman's original proposals. Though the enacted version of the statute did not fully satisfy their expectations, American courts soon offered them some grounds for hope.

## **9. The classical economics of early antitrust jurisprudence**

The so-called formative era of US antitrust law conventionally spans almost 25 years, from 1890 to 1914. This is the most intensively studied period in the history of American competition policy. The reason is simple: the courts of the time, first among them the Supreme Court, handed down a series of landmark decisions the doctrines of which would shape antitrust law for several decades and, in a few instances, even survive into the new millennium. Some of these decisions are almost legendary, if only because they involved the biggest names in American capitalism – J.P. Morgan in the *Northern Securities* case,



J.D. Rockefeller in *Standard Oil*, or J.B. Duke in *American Tobacco*. These cases dominated the news of the day. Presidential elections were won and lost in their shadow. A mythical aura surrounds the Justices who left their mark in this jurisprudence.

It is simply impossible to investigate this extraordinary period in a few pages.<sup>40</sup> My goal is therefore just to pinpoint the key turning points of the Supreme Court's jurisprudence, with a view to understanding how events of the period were shaped by the "freedom to trade v. freedom of contract" dichotomy. To this aim, the customary division of the formative era into three phases – the common law phase, the literalist phase and the rule of reason phase – is quite convenient. Each phase saw the Supreme Court's majority entertain a different attitude towards the Sherman Act. Each attitude was, in turn, the outcome of the conflict within the Court between the two freedoms.

The Justices split in two factions, usually labeled Literalists and Rule-of-Reasonists. In broad terms, the former endorsed the idea that the goal of antitrust should be freedom from market power, as epitomized by Senator Sherman's image of "full and free competition" among roughly equal businesses. The latter viewed the principle of contractual freedom as the guiding light for antitrust enforcement and believed that reasonable CRTs could often be the best method of guaranteeing fairness of price and profit instead of destructive competition.

### 9.1. It's all in common law

The landmark case in the first phase is *In re Greene*, an 1892 decision about the whisky trust by Ohio Circuit judge, and future Supreme Court Justice, Howell Jackson.<sup>41</sup> The key doctrine established by Jackson vindicated Senator Edmunds's interpretation of "monopolizing". Jackson declared that "a business might be a monopoly in fact, yet not a monopoly at law, if it had become the sole producer of a commodity without having done anything illegal to achieve or retain its control" (Letwin 1965, 147). What made the decision truly important was the way he reached the conclusion that the whisky trust had done nothing illegal "to achieve or retain" its dominant position.

Jackson argued that no trust could violate the Sherman Act by its mere existence. A monopoly at common law consisted of two necessary elements:

<sup>40</sup> Good general references are Letwin 1965, Chs.5-6-7; Sklar 1988, Ch.3; Freyer 1992, Ch.4; and Peritz 1996, Ch.1.

<sup>41</sup> *In re Greene*, 52 F. 104 (C.C.S.D. Ohio, 1892).

an exclusive right enjoyed by the monopolist and some restrictions imposed on other agents' rights or liberty in order to prevent them from infringing that privilege (*In re Greene*, at 115). Unfortunately, the plaintiff, the US Department of Justice, had proved neither. First, the whisky trust had no exclusive control of the market: competitors, actual or potential, did exist. Moreover, the indictment contained no allegation that the trust had undertaken any particular action to curtail its competitors' business freedom (*ibid.*, at 116-18). In case a restraint did exist, it should be evaluated according to the common law test of reasonableness. In particular, Jackson pointed at the version of the test developed by the House of Lords in a recent cartel case:<sup>42</sup> a restraint would be unreasonable only if it were "more injurious to the public than is required to afford a fair protection to the party in whose favor it is secured" (at 118). The choice of an English common law precedent was crucial because it impressed a well-defined stamp on the first phase of the formative era. Enforcing the Sherman Act meant applying the ongoing common law of CRTs - in other words, it meant defending freedom of contract as well as the intangibility of property rights.

Classical competition was for Jackson the rationale for constructing the Sherman Act as a statute addressed to protecting contractual freedom and property rights. Nothing prevented the whisky trust's competitors from offering larger rebates to their dealers, he argued. Actual and potential competitors would always be around, ready to steal business from the trust if it ever charged excessive prices. Hence, a trust having no exclusive control of the market or doing nothing wrongful to exclude its rivals could never be guilty of monopolizing.

If Jackson's decision captured the gist of the conventional economists' viewpoint, "new school" ideas found application in the first two trials in the *Trans-Missouri Freight Association (TMFA)* case.<sup>43</sup> Both the district court and the Court of Appeals for the Eighth Circuit affirmed the lawfulness of an agreement to fix uniform rates and terms of freight carriage entered by a number of railroad companies for the purpose of avoiding destructive competition. Both courts agreed that the Sherman Act should follow the common law and that the latter did not aim at preserving competition as such. As if he were rebuking Senator Sherman himself, the trial judge wrote: "the public is not

<sup>42</sup> *Mogul Steamship Co v. McGregor, Gow & Co and others*, All ER Rep 263 (1891-94), at 280. On this case, see Giocoli 2014, Ch.2.

<sup>43</sup> *United States v. Trans-Missouri Freight Association*, 53 F. 440 (C.C.D. Kansas 1892), affirmed 58 F. 58 (8th Circuit, 1893). See Letwin 1965, 152-4.

entitled to free and unrestricted competition, but what it is entitled to is fair and healthy competition". A restraint eliminating excessive competition and setting reasonable prices ought to be reasonable. In fact, not only should it be lawful, it would actually be beneficial to society.

Casting doubts on the unlimited merits of competition may once again seem miles away from the classical logic. The distance gets shorter, however, if the Act is read with the lenses of common law, and in particular of its guiding light, contractual freedom. Liberty of contract policed competition in both cases: according to Judge Jackson, freedom of contract, if unrestrained by the trust's unlawful practices, would guarantee competition via free entry and the "rivalry in service" towards customers; according to the *TMFA* lower courts, freedom of contract included the freedom to voluntarily enter a combination to preserve "fair and healthy" competition from ruinous outcomes.

## 9.2. Free competition, literally

The Supreme Court's 5-to-4 decision to reverse the lower courts' acquittals in *TMFA* marked the transition between the first and the second phase of the formative era.<sup>44</sup> The 1897 verdict was shocking. Writing for the Court, Justice Rufus W. Peckham – the future author of the *Lochner* opinion – declared that with the Sherman Act the Congress had aimed at *canceling and replacing* the common law of CRTs, especially the distinction between reasonable and unreasonable restraints. Moreover, his decision featured the Court's earliest formulation of what we may call the *competition principle*, that is, the idea that the only "fair" prices are competitive prices.

Peckham's main thesis was based on a plain, literal reading of the Act. "By the simple use of the term 'contract in restraint of trade'", he wrote, "all contracts of that nature, whether valid or otherwise, would be included, and not alone that kind of contract which was invalid and unenforceable as being in unreasonable restraint of trade" (*TMFA*, at 328). It followed that, for instance, all price-fixing agreements violated the Act, regardless of the reasonableness of the established prices. But Peckham went beyond that. He deemed it impossible to determine whether any agreed price be reasonable. Only competitive prices were reasonable: "Competition, free and unrestricted, is the general rule which governs all the ordinary business pursuits and transactions of life" (*ibid.*, at 337).

His statement about the absolute merits of "free and unrestricted competition" vindicated, after almost a decade, Sherman's original bill. Absent

<sup>44</sup> *United States v. Trans-Missouri Freight Association*, 166 US 290 (1897).

from the Act's final text, the competition principle found its earliest formulation in *TMFA* and has accompanied American antitrust law to the present day. Beyond faith in competition, the principle also expressed the literalist majority's dread of the socio-political consequences of an excessive concentration of economic power. Like Sherman and his faction, Peckham viewed rivalry among roughly equal firms of relatively small size as the bedrock of individual liberty. Powerful combinations in restraint of trade could drive out of business "the small dealers and worthy men whose lives have been spent therein", turning each of them from "an independent businessman, the head of his establishment, small though it might be, into a mere servant or agent of a corporation" (at 323-4). Only "free and unrestrained competition" could effectively rein in economic power. The principle was therefore instrumental to preserve economic democracy and, in the long run, democracy *tout court*. Still, it is important to underline that the competition principle has lived a life of its own in antitrust law: even when divested of Peckham's socio-political reading, the principle has found support in the *neoclassical* image of a competitive market structure.

The principle also helped dismiss the habitual ruinous competition defense used by the railroads' counsel. Peckham countered that different opinions existed as to the ultimate effects of combination, even in the railroad industry. It was therefore impossible to determine whether the prices fixed by the combination were reasonable (at 338-9). But, he added, this was precisely the motive Congress had *not* distinguished between reasonable and unreasonable restrains. The Act should be read literally, that is, as a statute directed against *all* CRTs.

*TMFA* is also important for Justice Edward White's dissenting opinion, which contained the first statement of the doctrine later known as the "rule of reason". The problem, as White saw it, was the permanent industrial warfare that would inevitably follow from a system of "free and unrestrained competition". Worse, Peckham's approach, if truly taken *literally*, would construe the Sherman Act as prohibiting almost every trade, because most trades involving the sale of property have some restraining conditions attached to them. The result would be paradoxical: "The plain intention of the law was to protect the liberty of contract and the freedom of trade", White wrote. "Will this intention not be frustrated by a construction which, if it does not destroy at least gravely impairs, both the liberty of the individual to contract and the freedom of trade?" The solution lay in the rule of reason, viewed as the best safeguard of the common law standard of "utmost liberty of contracting" (at 355).

As it turns out, by 1897 the conflict between the two visions of antitrust

that had animated the Senate debate had reignited at the highest judicial level. One view had prevailed, but the Court's 5-to-4 split testified how precarious the new equilibrium might be. Peckham himself recognized this: in a series of later decisions, though never reneging his *TMFA* opinion, he nevertheless endeavored to soothe its most irking implications.<sup>45</sup>

### 9.3. Enter the rule of reason

Flash forward fourteen years. The rule of reason was formally introduced into antitrust law in 1911, with the *Standard Oil* decision.<sup>46</sup> The case was a big deal, because it involved the emblem of all trusts, Rockefeller's mammoth petroleum business. At the same time, it was a difficult case, because, despite the massive trial record the Justices were presented with, the decision turned on a single, thorny issue: could a giant trust like Standard Oil ever be lawful under the Sherman Act? Or, even more bluntly, could trusts themselves be lawful as a form of organizing corporate affairs?

The Department of Justice had prepared the case with great care. Sensing the difficulty of winning the verdict, it had alleged that Standard Oil had violated not only the Act's section 1, but also section 2. Ample evidence justified invoking the latter provision. Defendants had notoriously engaged in several practices aimed at securing a substantial portion of the interstate petroleum market, practices such as taking preferential rebates from shippers, employing commercial spies, or setting cutthroat prices. These conducts would satisfy, so the government argued, the *In re Greene* doctrine that "monopolizing" meant forcibly excluding competitors. Even the defendants indeed conceded that Standard Oil had used these practices. The crux of the matter was whether they represented illegal monopolization, i.e., an unlawful way to get control of a large share of the market.

Circuit judges had argued that unlawfulness followed from the very fact that, by forming the trusts, the defendants had joined in a combination in restraint of trade. This, as established by the *TMFA* doctrine, was per se illegal under section 1. Such interpretation, which entailed that every trust was per se unlawful, was rejected by the *Standard Oil* Court, with a decision authored by the new Chief Justice White, the same Justice who had dissented in *TMFA*.

Two issues were at stake. The first was whether trusts as big as Standard Oil should be declared per se illegal. The second was whether in that specific

<sup>45</sup> On these decisions, see Letwin 1965, 178-81. Closer scrutiny indeed reveals that Peckham's true reading of the Act was far less "literal" than usually believed.

<sup>46</sup> *Standard Oil Co. of New Jersey v. United States*, 221 US 1 (1911).

case the defendants had violated the Sherman Act. White managed to keep them separate: the lawfulness of trusts ought to be decided case by case, by applying the rule of reason rather than any per se prohibition; yet, in that specific case, Standard Oil deserved dissolution. The trick for splitting the issues, while at the same time avoiding marking too large a distance from previous jurisprudence,<sup>47</sup> consisted of “imagining the trust problem as a question of individual conduct rather than an instance of combination”. By so doing, “the Court could reconcile competition policy as articulated in the cartel cases with common-law private property rights and with the constitutionalized liberty of contract” (Peritz 1996, 50).

In practical terms, White’s rule of reason stood for the implicit proposition that there were good trusts and bad trusts, and that the former could be perfectly lawful monopolies even under the Sherman Act. The point was, as always, that competition was not necessarily beneficial. Competition could well be destructive; hence, combinations could be reasonable. Trusts could be good if and when they avoided ruinous competition. Remarkably, White rejected the universal validity of the principle of “free and unrestrained” competition without invoking anymore a businessman’s right to fair price or fair profit. The Constitution protected freedom of contract and individual property rights against government interference, while the common law guaranteed that neither trusts nor monopolies could be illegal just because of their existence or size, but only “because of their restriction upon individual freedom of contract and their injury to the public” (*Standard Oil*, at 54).

In an illuminating passage, White revealed the classical inspiration behind his argument. The Sherman Act, he wrote, “indicates a consciousness that the freedom of the individual right to contract, when not unduly or improperly exercised, was the most efficient means for the prevention of monopoly, since the operation of the centrifugal and centripetal forces resulting from the right to freely contract was *the means by which monopoly would be inevitably prevented* if no extraneous or sovereign power imposed it and no right to make unlawful contracts having a monopolistic tendency were permitted” (*ibid.*, at 62, emphasis added). A classical economist could not have said it better. Freedom of contract, undisturbed by government interference, was the key liberty fostering competition and its beneficial effects, including the dissolution of monopolies.

<sup>47</sup> White was very successful at reconciling the opposing factions. The decision was almost unanimous, with only Justice Harlan concurring in the verdict but dissenting from the opinion.

The 1911 decision against Standard Oil, and the analogous one against the American Tobacco trust,<sup>48</sup> were epoch making. Under White's rule of reason, freedom of contract and the protection of property rights became the Supreme Court's guiding light for antitrust enforcement until the New Deal. While the competition principle remained formally valid, it would not drive adjudication. First and foremost, contractual freedom and property rights had to be protected against government intrusions into private agreements. As in the first phase of the formative era, the Court was back with a property rights language, though now devoid of any reference to the fairness of profits and prices. The rule of reason aimed at affirming freedom of contract, while prohibiting only unreasonable exercises of that freedom. The goal was to *broaden* contractual freedom, rather than limiting it.

White's construction of the rule of reason also contained the rationale for condemning Standard Oil and, a few days later, American Tobacco. Liberty of contract was never absolute. The common law set its limit in respect of anyone else's contractual freedom – viz., in the classical *sic utere tuo* doctrine (see sect. 1). A trust was unreasonable when and if it engaged in conduct that unduly restricted rivals' possibility to compete. It was precisely the infringement of other market participants' liberty that doomed the two giant trusts.

The Court argued that contractual freedom could be exercised in an “unnatural” way, by making use of acts or agreements of an “unusual and wrongful character” (*American Tobacco*, at 181), such as predatory pricing, boycotting or cartelization. Hence, the problem was never the trust itself. Reasonable restraints of trade, like the good trusts, did not violate the Sherman Act, no matter how general or how powerful. Unlawfulness only emerged when the trust “abused” its contractual freedom to curb other market participants' freedom. In the classical view of competition, this behavior would represent an obstacle to the natural working of the competitive process, and so would warrant condemnation. Standard Oil and American Tobacco had done exactly that. By acting badly against their competitors, they had trespassed the common law boundaries of reasonableness, i.e., they had unreasonably restrained trade.

That with only one exception the rest of the Justices undersigned White's *Standard Oil* opinion is noteworthy. The ample majority signaled a substantial agreement with respect to the illegality of the trust's practices. It also indicated that the Justices did not view White's opinion as so radical a change with respect to previous jurisprudence. White took care to clarify that, following

<sup>48</sup> *United States v. American Tobacco Co.*, 221 U.S. 106 (1911).

TMFA, cartels would remain per se unlawful even under the rule of reason. A few unilateral practices that everybody deemed unreasonably exclusionary (most famously, predatory pricing) would be per se illegal too. For the rest, free competition should still be policed in the name of (classical) freedom of contract, rather than (neoclassical) freedom from market power.

## 10. Conclusion: the shadow of old debates

The paradox of *Standard Oil* is well-known. A verdict that had condemned the mother-of-all-trusts was read by many contemporary observers<sup>49</sup> as establishing a very lenient approach towards big business. This eventually pushed Congress to approve new antitrust legislation. Even more significantly for my aims here, it led a prominent economist to rethink his views about competition and actively campaign for legislative reform.

One of this essay's goals was to show how the classical dichotomy between the two freedoms did not just shape the Gilded Age and Progressive Era discourse about regulation and antitrust, but also the way these themes are handled today. Under several respects, we still live in the shadow of these old debates. For final evidence about such present-day relevance, let us consider the economic rationale underlying two further pieces of US antitrust legislation that are still with us, more than one hundred years after their enactment.

The main intellectual source for the 1914 Clayton and FTC Acts was John Bates Clark. While this is, again, fairly well-known,<sup>50</sup> what is seldom recognized is that the great American economist formulated the proposals that would eventually be incorporated in the two statutes only after a profound reconsideration of the "freedom to trade v. freedom of contract" dichotomy. A comparison of the two editions of *The Control of Trusts* – the first, published in 1901, written by Bates Clark alone; the second, which appeared in 1912, co-authored with his son John Maurice – is enlightening in this regard.

The declared intent of both editions was to influence the public debate about trusts and other combinations. The policy advocated by Clark in 1901 was very simple, and very classical: rely upon the power of actual and, above all, potential competition and do whatever is needed to help market forces

<sup>49</sup> And not only observers. Stock indexes rose after the decision, signaling widespread business satisfaction.

<sup>50</sup> See e.g. Fiorito 2012a.



exercise their restraining action on trusts. In his words: "It is the policy that relies wholly on competition as the regulator of prices and wages and as the general protector of the interests of the public. It welcomes centralization, but aims to destroy monopoly, and to do this by keeping the field open to all independent producers who may choose to enter it" (Clark 1901, v). Clearly, at that time he still believed in the self-regulative ability of freely competitive markets.

However, Clark was also aware that actual and potential competition could fail in the presence of exclusionary practices undertaken by big firms and combinations against independent rivals. Blocking the efficiency-based selection operated by market forces paved the way for external intervention: "When economy in production no longer saves [the independent competitors], it is time for the state to intervene; and it needs to do this, if it would carry out the very end for which it was originally established, — *the protection of property itself, by the suppression of refined forms of robbery*" (ibid., 65, emphasis added). The italicized words clarify that intervention should not aim at directly protecting consumers or social welfare. At stake was a specific *right*, viz., the independent competitors' right to thrive from their market efforts without suffering undue interference from bigger rivals. The classical premise still held that if agents enjoyed the utmost freedom to carry on their business, then the whole society would inevitably prosper – producers, workers and consumers alike. No ideal image of a specific market structure – like, say, neoclassical perfect competition – influenced Clark's reasoning. Exclusionary practices operated by big businesses simply robbed some firms of their fair chance to win a hard-earned market prize. Safeguarding this chance – and, more generally, the competitive process, actual and potential – was a legal matter of defending a well-defined, though abstract, property right. Those practices should therefore be expressly outlawed.

No new antitrust statute was required, though, but just the list of proscribed practices already identified by the common law. Clark still believed in 1901 that the latter would provide the ultimate weapon. "Statutes are not our sole reliance", he wrote, "the most efficient action that has thus far been taken in curbing the power of trusts has been taken under the common law" (ibid., 69). In short, both pillars of Clark's 1901 view of antitrust policy were thoroughly classical, namely, potential competition and the common law. Courts should just recognize, and protect, the role of potential competition.

More than a decade later, and now writing with his son, Clark would abandon the optimism of the first edition and invoke a wider program of antitrust reform. Benefiting from the experience of another decade of antitrust

enforcement, the 1912 edition contained a mature assessment of the pros and cons of alternative proposals. The authors' blueprint for regulating competition via the proscription of a specific list of business behaviors and the break-up of excessive concentrations of economic power marks the 1912 work as a very modern contribution to antitrust policy, ahead of its times in several respects. Yet, the volume was also a child of its time, its publication owing much to the *Standard Oil* decision and ensuing paradox.

The Clarks' disillusion with respect to the actual power of potential competition was now strong and, absent explicit legislative intervention, definitive. Pessimism seeped already in the Preface: "When the first edition of this work was issued, so called potential competition had shown its power to control prices". Unfortunately, combinations had "soon discovered their own power to crush audacious rivals when they appeared. In a number of ways, which are now well known and are discussed in this volume, they could club a competitor whenever he should show himself in an active way. They so often did this that their evident power to do it had its effect in advance, and deterred competitors from appearing. The potentiality of unfair attacks by the trust tended to destroy the potentiality of competition" (Clark & Clark 1914 [1912], vi-vii). And later in the book: "The real difficulty is that the influence of this latent competition cannot be trusted as it could in earlier days. [...] the possible competitor does not become an actual one as promptly as he should. The trouble is that he has not a fair chance for his life when he appears on the scene. He is in danger of being crushed by the trust, and that too, not in any natural way, but by certain entirely abnormal things that the trust is able to do" (ibid., 27). As it turns out, by 1912 the Clarks had abandoned the classical and common law ideal that what they called the "honest race for cheapness" (ibid., 28) could represent the only kind of horizontal competition between rival firms.

Still, the wand of potential competition had not entirely lost its magic, if only the curse inhibiting it could be removed: "If the great company could not do these things, the competitor would be comparatively safe, and in many departments of industry he would appear promptly whenever profits should become high enough to make his presence desirable" (ibid., 27). Much of the 1912 book was dedicated to explaining the nature of the curse and to devising a counter-spell that would restore the power of the magic wand.

The abracadabra was congressional intervention, in the form of a new statute. The statute should declare per se illegal a list of exclusionary practices (such as price discrimination, selective discounts and predatory pricing), regardless of the nature of the firm undertaking them or of their eventual outcome: "The situation demands that such acts be made illegal in themselves, whether or not they have

been carried so far as to result in monopoly” (ibid., 100).<sup>51</sup> In short, a business’s behavior – a.k.a., its contractual freedom – should be statutory constrained. The shift towards a freedom *from* contract perspective was a sea-change with respect to Bates Clark’s previous views, which only admitted the auxiliary role of common law courts, and a significant breach in the classical edifice.

Political vagaries notwithstanding, the gist of the Clarks’ proposal would feature in the Clayton and FTC Acts.<sup>52</sup> Yet, their 1912 blueprint for reform contained a second horn, complementary to the list of proscribed practices. Fearing troubles with courtroom enforcement, the Clarks advocated an administrative, rather than judicial, approach. Accordingly, Congress should create a new administrative body, an “Interstate Commerce Commission” as they called it (ibid., 117-18). Guided by economic experts, the Commission would take most regulatory power away from courts; in particular, it would guarantee a prompt reaction against any newly invented form of exclusionary practice.

Both horns of the Clarks’ proposal would find application in the 1914 statutes. From the viewpoint of the history of economic thought, this was a remarkable achievement. Contrary to what had happened with the Sherman Act, this time the economists’ reflections on the two meanings of free competition, and how to reconcile them under the new industrial conditions, had a direct and explicit (and not just indirect and implicit) impact on the legislative process. One century later, a big part of the antitrust enterprise is still driven by the outcomes of those very reflections. It is up to historians of economic thought to keep this memory alive.

## References

- Bernstein M. A. (2001), *A Perilous Progress. Economists and Public Purpose in Twentieth-Century America*. Princeton: Princeton University Press.
- Blaug M. (1997), “Competition as an End-state and Competition as a Process”. In *Not Only an Economist: Recent Essays by Mark Blaug* (pp. 66-86). Cheltenham: Edward Elgar.
- Bradley R. L. Jr. (1990), “On the Origins of the Sherman Antitrust Act”. *Cato Journal*, 9 (3), 737-742.

<sup>51</sup> Anticipating the legislative difficulty of providing an exhaustive list of undesirable practices, the Clarks also proposed that the new law should contain a broader, and quite modern, definition of exclusionary behavior: “unfair competition’ comes to mean, virtually, any practice whose natural result is to make survival depend on other qualities than industrial efficiency” (Clark & Clark 1912, 103).

<sup>52</sup> For the details, see Winerman 2003; Fiorito 2012b.

- Clark J. B. (1887), "The Limits of Competition". *Political Science Quarterly*, 2 (1), 45-61.
- Clark J. B. (1900), "The Necessity of Suppressing Monopolies While Retaining Trusts".  
In *Chicago Conference on Trusts* (pp. 404-409). Chicago: Civic Federation of Chicago.
- Clark J. B. (1901), *The Control of Trusts*. New York: Macmillan.
- Clark J. B., Clark J. M. (1914) [1912], *The Control of Trusts. Rewritten and Enlarged*. New York: Macmillan.
- Ely R. T. (1888), *Problems of To-Day. A Discussion of Protective Tariffs, Taxation, and Monopolies*. New York: Thomas Y. Crowell & Co.
- Epstein R. A. (2014), *The Classical Liberal Constitution. The Uncertain Quest for Limited Government*. Cambridge Mass.: Harvard University Press.
- Fine S. (1956), *Laissez Faire and the General-Welfare State. A Study of Conflict in American Thought, 1865-1901*. Ann Arbor: University of Michigan Press.
- Fiorito L. (2012a), "When Economics Faces the Economy: John Bates Clark and the 1914 Antitrust Legislation". *Review of Political Economy*, 25 (1), 139-163.
- Fiorito L. (2012b), "The Influence of American Economists on the Clayton and Federal Trade Commission Acts". *Research in the History of Economic Thought and Methodology*, 30, 1-58.
- Fisher I. (1997)[1912], *Elementary Principles of Economics*. In *The Works of Irving Fisher*, edited by W. J. Barber, Vol. 5. London: Pickering & Chatto.
- Freyer T. A. (1992), *Regulating Big Business. Antitrust in Great Britain and America 1880-1990*. Cambridge, UK: Cambridge University Press.
- Giocoli N. (2012), "Who Invented the Lerner Index?". *Review of Industrial Organization*, 41 (3), 181-191.
- Giocoli N. (2014), *Predatory Pricing in Antitrust Law and Economics. A Historical Perspective*, Routledge, London & New York, 2014
- Giocoli N. (2016), "When law and economics was a dangerous subject: the controversy over railroad regulation before the Hepburn Act". *Research in the History of Economic Thought and Methodology* 34A: 293-338.
- Giocoli N. (2017a), "The (Rail)Road to Lochner: Reproduction Cost and the Gilded Age Controversy over Rate Regulation". *History of Political Economy* 49 (1): 31-58.
- Giocoli N. (2017b), "Value is not a fact: reproduction cost and the transition from classical to neoclassical regulation in Gilded Age America". *Journal of the History of Economic Thought*, forthcoming.
- Hadley A. T. (1885), *Railroad Transportation. Its History and Its Laws*. New York: Putnam's Sons.
- Hadley A. T. (1886), "How Far Have Modern Improvements in Production and Transportation Changed the Principle That Men Should Be Left Free to Make Their Own Bargains? I". *Science*, 7 (161), 221-225.
- Hovenkamp H. (1989), "The Sherman Act and the Classical Theory of Competition". *Iowa Law Review*, 74, 1019-1065.
- Hovenkamp H. (1991), *Enterprise and American Law, 1836-1937*. Cambridge, Mass.: Harvard University Press.
- Jenks J. W. (1900), *The Trust Problem*. New York: McLure, Phillips & Co.
- Jenks, J. W. (1912), "Economic Aspect of the Recent Decisions of the United States Supreme Court on Trusts". *Journal of Political Economy*, 20 (4), 346-357.
- Jensen, M. C. (1993), "The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems". *Journal of Finance, Papers and Proceedings*, 48 (3), 831-880.

- Letwin W. (1965), *Law and Economic Policy in America: The Evolution of the Sherman Antitrust Act*. Chicago: University of Chicago Press.
- May J. (1987), "Antitrust Practice and Procedure in the Formative Era: The Constitutional and Conceptual Reach of State Antitrust Law 1880-1918". *University of Pennsylvania Law Review*, 135 (3), 495-593.
- McCann Ch. R. Jr. (2012), *Order and Control in American Socio-Economic Thought. Social Scientists and Progressive Era Reform*. London: Routledge.
- McCulloch J. R. (1825), *The Principles of Political Economy: with a Sketch of the Rise and Progress of the Science*. Edinburgh: William and Charles Tait.
- McCurdy C. W. (1979), "The Knight Sugar Decision of 1895 and the Modernization of American Corporation Law, 1869-1903". *Business History Review*, 53 (3), 304-342.
- McNulty P. J. (1967), "A Note on the History of Perfect Competition". *Journal of Political Economy*, 75 (4), 395-399.
- Mill J. S. (1909 [1848]), *Principles of Political Economy with some of their Applications to Social Philosophy* (7<sup>th</sup> ed.). London: Longmans, Green and Co.
- Page W. H. (2008), "The Ideological Origins and Evolution of U.S. Antitrust Law." *Issues In Competition Law and Policy* (ABA Section of Antitrust Law), 1: 1-17.
- Palgrave R. H. Inglis, ed. 1894. *Dictionary of Political Economy*. Vol. I. London: Macmillan.
- Perelman M. (2006), *Railroading Economics. The Creation of the Free Market Mythology*. New York: Monthly Review Press.
- Peritz R. J. R. (1996), *Competition Policy in America. History, Rhetoric, Law*. Oxford: Oxford University Press.
- Schumpeter J. A. (1986 [1954]), *History of Economic Analysis*. London: Routledge.
- Senior N. W. (1854), *Political Economy* (3<sup>rd</sup> ed.). London: Richard Griffin and Co.
- Siegel S. A. (1984), "Understanding the Lochner Era: Lessons from the Controversy over Railroad and Utility Rate Regulation". *Virginia Law Review*, 70 (2), 187-263.
- Siegel S. A. (1986-87), "Understanding the Nineteenth Century Contract Clause: the Role of the Property-Privilege Distinction and 'Takings' Clause Jurisprudence". *Southern California Law Review* 60 (1): 1-108.
- Sklar M. J. (1988), *The Corporate Reconstruction of American Capitalism, 1890-1916. The Market, the Law, and Politics*. Cambridge, UK: Cambridge University Press.
- Smalley H. S. (1906), "Railroad Rate Control in Its Legal Aspects: A Study of the Effect of Judicial Decisions Upon Public Regulation of Railroad Rates". *Publications of the American Economic Association, 3rd Series*, 7 (2), 4-147.
- Smith A. (1904 [1776]), *An Inquiry into the Nature and Causes of the Wealth of Nations* (5<sup>th</sup> ed., E. Cannan, ed.). London: Methuen & Co.
- Spencer H. (1851), *Social Statics; or, The Conditions essential to Happiness specified, and the First of them Developed*. London: John Chapman.
- Stickney A. (1897), *State Control of Trade and Commerce by National or State Authority*. New York: Baker.
- Stigler G. J. (1957), "Perfect Competition, Historically Contemplated". *Journal of Political Economy*, 65 (1): 1-17.
- Stigler G. J. (1976), "The Successes and Failures of Professor Smith". *Journal of Political Economy*, 84 (6), 1199-1213.
- Stigler G. J. (1982), "The Economists and the Problem of Monopoly". *American Economic Review*, 72 (2), 1-11.

- Viner J. (1960), "The Intellectual History of Laissez Faire". *Journal of Law and Economics*, 3 (October), 45-69.
- von Halle E. (1896), *Trusts or Industrial Combinations and Coalitions in the United States*. New York: MacMillan.
- Walker F. A. (1892), *Political Economy, Third Edition*. London: Macmillan.
- Winerman M. (2003), "The Origins of the FTC: Concentration, Cooperation, Control, and Competition". *Antitrust Law Journal*, 71, 1-97.
- Yonay Y. P. (1998), *The Struggle over the Soul of Economics. Institutional and Neoclassical Economists in America between the Wars*. Princeton, NJ: Princeton University Press

# Specific investments and residual control rights: a reassessment\*

MASSIMILIANO VATIERO

Itaque rem suscipit et a Sequanis impetrat,  
ut per fines suos Helvetios ire patiantur,  
obsidesque uti inter sese dent perficit:  
Sequani, ne itinere Helvetios prohibeant,  
Helvetii, ut sine maleficio et iniuria transeant.  
Giulio Cesare, *De Bello Gallico*

## 1. The issue

With complete contracts, all uses of resources will be specified in all possible eventualities and, therefore, observable by a third party (the Enforcer). In reality, contracts are not comprehensive and, therefore, for a third party can be costly (or impossible) to observe if one party's conduct in a certain contingency is consistent with contractual clauses. In Hart's (1995:39) words, "the contract will contain gaps and missing provisions." In general terms, each contract is incomplete; namely it is affected by positive transaction costs. Unlike the Coase Theorem, transaction costs are marginally significant in incomplete contract theory (see also Pagano & Vatiéro 2015).

There is a further important distinction between the Coase Theorem (and, more generally, Coasian argument) and incomplete contract theory. This dis-

\*The author acknowledges comments from Luciano Fanti, Nicola Giocoli, Nicola Meccheri, Mario Morroni, Giuseppe Niglia, Ugo Pagano, Neri Salvadori, Rodolfo Signorino, Giuseppe Vittucci and participants of the workshop "Oligopolio, Istituzioni e Performance delle Imprese" (26-27 January 2017, University of Pisa).

tion concerns the role of outside options. As affirmed by Coase in explaining the externality between the farmer and the rancher, his attention is limited to a context where outside options are given and exogenous:

[i]f the crop was obviously sold in conditions of perfect competition, marginal cost was equal to price for the amount of planting undertaken and any expansion would have reduced the profits of the farmer. [...] Of course, if cattle-raising commonly involved the destruction of crops, the coming into existence of a cattle-raising industry might raise the price of the crops involved and farmers would then extend their planting. But I wish to confine my attention to the individual farmer. (Coase 1960:3-4)

In other words, the externality between the two parties (e.g. the farmer and the rancher) and its solution do not affect the market prices of outside options' resources (crops, lands and meat). Coase (1960) finds as justification the fact that the other relationships (namely, between alternative farmers and alternative ranchers) are in perfect competition; therefore, it is not surprising that the first formulation of the Coase Theorem is "under perfect competition, private and social costs will be equal" (Stigler 1966:113). However, what matters here is not that Coase assumes a certain type of competition rather than another type of competition, but that in doing so Coase (1960) limits his argument to a bilateral relationship, neglecting the role played in a transaction by outside options (such as alternative farmers and ranchers).<sup>1</sup>

While they do not matter to the Coase Theorem, outside options have a prominent role in the incomplete contract theory; in particular, in the emergence of the hold-up problem and in solution of that problem.

From the perspective of Grossman & Hart (1986), Hart & Moore (1990), and Hart (1995) (hereinafter GHM), to be credible the opportunistic behaviours (i.e. hold-up) of the "Non-Investor", he/she should be *credibly* able to switch on outside options (e.g. the spot market of the "widget"). This is a commonality among contributions of incomplete contract theory: The credibility of a hold-up threat depends on the availability of outside options. A second commonality is that, because without outside options the hold-up threat is not credible, the solution to the problem of hold-up should limit or prevent the switching on outside options. The main distinction in the incomplete contract literature is *how* to limit or prevent outside options.

<sup>1</sup> Coase confined his examples to two-party disputes, not only the farmer and the rancher, but also the railway emitting sparks and the owner of an adjacent woodlot, the keeper of coney-burrows and the farmer, and the parties in the nineteenth-century nuisances cases (see Merrill & Smith 2001:371).



For GHM, the assignment of residual control rights (RCRs) to the investor fosters investments. The reason is that, after the integration, it is very hard for the non-investor to switch to the outside market. In fact, after the integration, the non-investor's physical assets are under the control of the investor. According to Williamson, instead, owning specific investments may have a commitment role which makes it very costly for the counterparty to switch on outside options. Williamson argues that specific investments may reduce the field of available alternatives (i.e. outside options) from a large number, i.e. the ex-ante bargaining situation, to a small number, i.e. the ex-post bidding bilateral monopoly; and such a *transformation* should decrease the risk of opportunistic behaviours. Furthermore, Williamson describes this process by using the notion of Fundamental Transformation (FT) (e.g., Williamson 1975, 1979, 1985, 1996, 2005, 2010; see also Klein 1996): a process through which a market configuration characterized by large numbers bidding at the outset will evolve, thanks to the parties' durable investments in transaction-specific assets, into an ex-post market configuration characterized by a bilateral monopoly. In this respect, investments may affect the counterparties' outside options and, by-product, they may limit or prevent the counterparty's potential hold up.

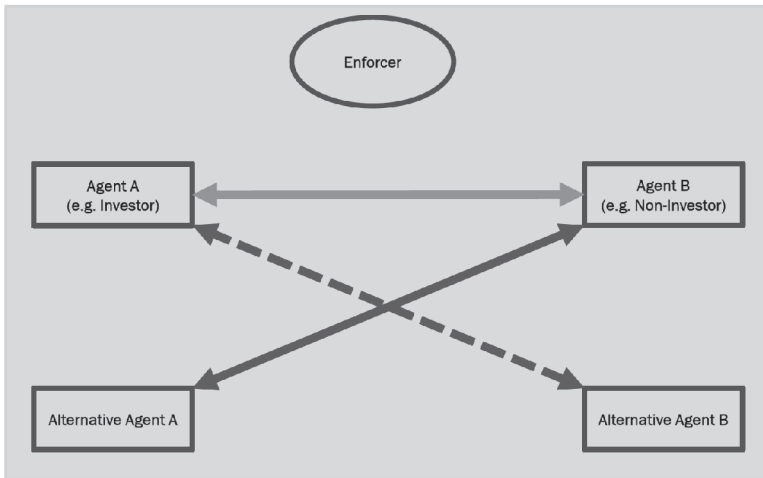
In this paper, I argue that, though sometimes one hears the claim that GHM formalized Williamson (e.g. Gibbons 2005), in reality GHM and Williamson offer different theories, which are essentially orthogonal. In GHM, specific investments are the origin of the problem of the hold-up and the assignment of RCRs is a second best<sup>2</sup> solution. In the Williamsonian FT, because each transaction among parties creates a *transaction residual* that favours a continued bilateral trading relationship, specific investments can mitigate the hold-up phenomenon, even without the allocation of RCRs. In particular, I argue that if we consider the role of specific investments in transforming the market configuration (as in FT), the allocation of RCRs may not be necessary or may even be undesirable.

This paper is structured as follows: In Section 2, I illustrate the transaction in Hart (1995) and the Williamsonian FT. Section 3 gives some examples of the Williamsonian FT, while Section 4 discusses the potential divergences between the Williamsonian FT and the Hartian RCRs. Finally, Section 5 is dedicated to my results.

<sup>2</sup> According to GHM, property rights represent a safeguard against hold-up. However, while it encourages owner's incentives, the assignment of property rights will decrease non-owners' incentives to invest. Then, only second best outcomes are possible.

## 2. Transaction à la Hart and Williamson

Consider a standard setting with an incomplete contract and a specific relationship between the Agent A (the Investor) and the Agent B (the Non-Investor). For the specificity of investments—namely investments that have little or no value outside a relationship but great value inside—it is very costly for the Investor to switch on outside options (e.g. the Alternative Agent B). In Figure 1, the dotted arrow indicates that there are high costs for the Investor when abandoning the specific relationship. On the contrary, the Non-Investor can threaten to switch on the spot market because there are available alternatives (and she does not invest in asset-specificity). Here, the asymmetry in terms of outside options is clear: The Investor is locked in a relationship; namely, there are not outside options, while the Non-Investor has available alternatives. It implies that, because the Non-Investor can credibly threaten to renegotiate the agreement, the Investor has reasons to underinvest in asset specificity. This is a common result of Hart and Williamson. The distinctions are apparent in the remedies.



**Figure 1: The role of outside options in the emergence of the hold-up problem**

In accordance with GHM, an integration (i.e. the Investor acquires the Non-Investor's assets) may mitigate the hold-up problem because after the vertical

integration, the Non-Investor's assets are under the control of the Investor; consequently, for the Non-Investor it is very difficult to switch on alternative parties.

For example, once General Motors owns Fisher Body and Fisher becomes an employee of General Motors, Fisher can not longer take advantage of an imperfect contract to threaten nonperformance because the ultimate power to make most important investment and management decisions, such as the location of the auto-body production plant, is legally transferred to the employer/owner [...]. The reduction in the ability of employees to threaten nonperformance can be seen clearly if we consider the possibility of bribes or side payments by third parties to influence an employee's plant-location decision. While an independent contractor has the right (and even the duty to its shareholders) to consider such payments in making decisions, employees cannot do so. (Klein & Murphy 1997:418)

Hence, according to GHM, ownership matters. Additionally, we argue that ownership matters because it has a negative impact on the non-owner's outside options.

Since outside options have a prominent role in the emergence of the hold-up problem, it is not surprising that outside options have a role in mitigating the problem of hold-up. Indeed, an ex-ante context, where Non-Investor's outside options make credible a renegotiation, is transformed by the assignment of RCRs in an ex-post context, where the Non-Investor's renegotiation is very costly (see Figure 2).

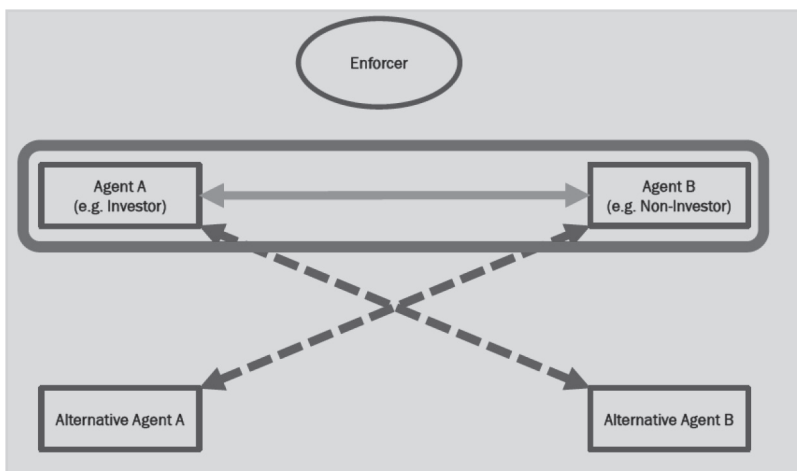
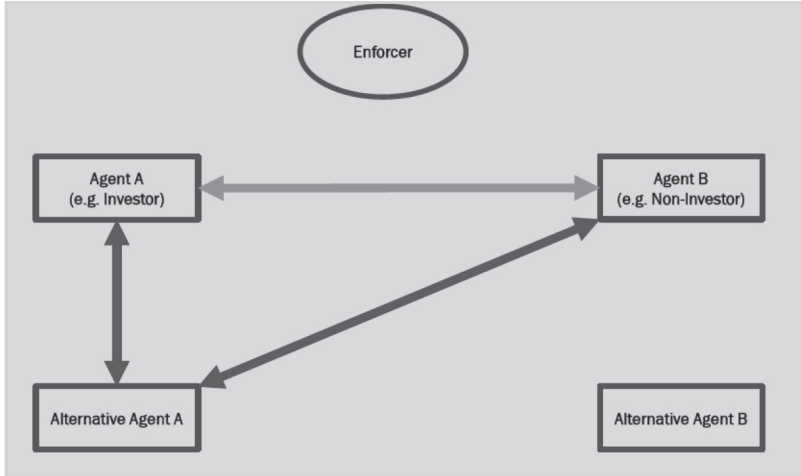


Figure 2: Outside options after the allocation of residual control rights

Williamson proposes a different solution, which is based on the FT process. Specific investments may diminish the Non-Investor's outside options and, therefore, it may diminish the credibility of the hold-up threat, as noted in Figure 3.



**Figure 3: Investments can affect counterparties' outside options**

The idea of Williamson is that specific investments may attain the transformation of the transactional environment from an ex-ante competitive market to a bilateral monopoly, as in Figure 4.<sup>3</sup>

In the case of a bilateral monopoly, the Non-Investor's opportunistic behaviour is extremely costly because there are no alternatives to the specific relationship with the Investor. Hence, in an ex-post market configuration characterized by bilateral monopoly, the threat of opportunistic behaviour is not credible; meanwhile, there is a credible commitment by both parties.

<sup>3</sup> Stating it differently, the FT relies on three periods:

- Pre-Transformation: Parties in an ex-ante competitive market agree to realize some transaction that involves specific investments.
- Transforming the setting: Specific investments are performed and transaction residual is created. An anonymous market exchange is transformed in a bilateral monopoly with non-anonymous parties.
- The transformed setting: A post-opportunistic behaviour losses transaction's residual gains.

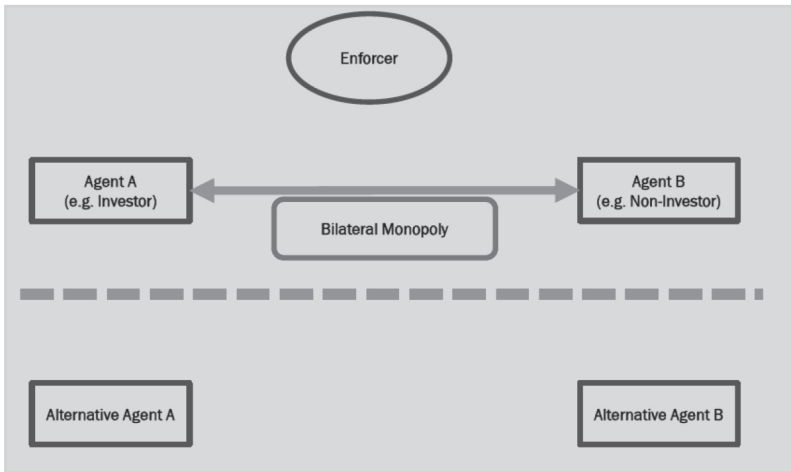


Figure 4: Bilateral monopoly

### 3. Fundamental Transformation: Some examples

Before showing the potential trade-offs between FT and RCRs, it may help to give two examples of the FT process. The first example is the so-called “Hostage model,” that is a commitment by a voluntary hostage positing in the sense of pledging a bond. This mechanism may favour cooperation even without repeated encounters. This basic idea is sharply illustrated in Schelling (1960:300). The example concerns a cooperation problem between a kidnapper who got ‘cold feet’ and his prisoner.

Both the kidnapper [...] and the prisoner may search desperately for a way to commit the latter against informing on his captor [...]. If the victim has committed an act whose disclosure could lead to blackmail, he may confess it; if not, he might commit one in the presence of his captor, to create the bond that will ensure his silence (Schelling 1960, fn. 43).

In other words,

[t]he essence of these tactics is some voluntary but irreversible sacrifice of freedom of choice. They rest on the paradox that the power to constrain an adversary may depend on the power to bind oneself; that, in bargaining, weakness is often strength, freedom may be freedom to capitulate, and to burn bridges behind one may suffice to undo an opponent (Schelling 1960:22).

Williamson (1983) develops this insight as follows: Assume that efficiency requires that a supplier invests in specific investments but that incompleteness

of contracts undermines the credibility of the buyer's promises to maintain purchase volume as a way of compensating the supplier for its investment. According to the hold-up theory, the supplier is likely to underinvest in specific assets. However, efficient levels of investments, thought in a context with incomplete contracts, can be sustained if the buyer posts/pledges a "hostage," namely something that is sacrificed in the event of premature termination of the contract. By posting a hostage, an actor incurs a commitment that serves as a safeguard for the partner to cooperate. Because the hostage is sacrificed in the case of termination of the relationship, this hostage has no value outside the relationship; therefore, the hostage is substantially a specific investment. In other words, the hostage model states that in order to foster specific investments of one party, the other party should invest in a specific relationship. A form of hostage model has been applied in a variety of settings, such as retail estate (Raub & Keren 1993), automobiles trade (Anderson & Bensaou 1999), fisheries (Koss 1999) and other domains (cf. Raub & Keren 1993, Anderson & Bensaou 1999).

A second example of the FT process—which is provided in Williamson 1983 and developed by, among others, Vatiero 2009, Nicita & Vatiero 2014 (see also Meccheri 2011)—relates with the literature on deterrence entry (e.g. Bain 1956, Sylos-Labini 1969, Dixit 1980, and Allen et al. 2000). According to this literature, sunk investments may represent a strategic device adopted by incumbents to deter entry. Namely, the incumbent may adversely affect the entrant's expected profits by increasing the level of her own sunk investments. A neglected consequence of this literature is that, if specific investments may potentially deter competitors, then specific investments may affect counterparties' outside options and, by-product, reduce counterparties' opportunistic behaviours. For instance, investments in capacity represent the standard example of specific investments (cf. the *General Motors vs. Fisher Body* case) as well as a threat/commitment to a war price that may affect the choices of (new-entrant) firms like in Dixit (1980) and Allen et al. (2000). Hence, specific investments (in capacity, at least) can solve the hold-up problem by the fact that such investments have a deterring effect.

#### **4. Fundamental Transformation vs. Residual Control Rights**

Because it diminishes the field of available alternatives in the outside market, the FT reduces the risk of hold-up. As a result, specific investments can be a safeguard from, rather than the cause of, opportunistic behaviours. If so, the allocation of RCRs may be unnecessary or even undesirable. In sum, specific investments have a double role:

- They are the origin of the hold-up problem. The standard literature shows that the specificity of investments is needed in order to create the conditions for post-opportunistic behaviour.
- They are the origin of the mitigation of hold-up. Each transaction among parties creates a “transaction residual” that favours a continued bilateral trading relationship. Indeed, specific investments transform the ex-ante competitive bargaining into an ex-post bilateral monopoly and such an ex-post market configuration of a bilateral relationship may provide a certain safeguard for their own specific investments by preventing parties from freely switching on alternatives.

This means that the specific investment, on the one hand, generates the conditions for the hold-up threat and, on the other hand, produces a credible commitment for the so-called transactional residual. This double role of specific investments is recalled also by Williamson. He writes,

[c]redible commitments and credible threats share this common attribute: Both appear in conjunction with irreversible, specialized investments. But whereas credible commitments are undertaken in support of alliance and to promote exchange, credible threats appear in the context of conflict and rivalry (Williamson 1985: 167)

If one takes seriously into account this double role of specific investment, then parties may be led to invest or over-invest in asset specificity to create a market structure that can make opportunistic behavior extremely costly (cf. Nicita & Vatiello 2014). That is, agents may select specific investments in order to reduce outside options, to worsen counterparties’ ex-post contractual power, and to minimize the probability of a counterparty’s hold-up: it implies that specific investments, when they are able to reduce relevant outside options, may protect one’s own specific investments from opportunistic behaviours. One further consequence is that the allocation of RCRs may be unnecessary if specific investments lead to a bilateral monopoly.

In addition, the allocation of RCRs may distort the FT process: Due to the fact that they impact on outside options, RCRs may discourage investments in transforming these outside options. In other terms, while vertical integration spurs specific investments under the conditions analysed by GHM, it is very likely that it also reduces the competition in the markets as well as incentives in investing in the outside options. This potential trade-off between incentives in the specific relationships and RCRs (which is largely neglected by the GHM model) is particularly relevant for antitrust implications concerning contractual remedies (e.g., exclusive dealings) or mergers,

which tend to diminish incentives deriving from market competition. Indeed, each antitrust case concerning complementary assets faces a trade-off between the protection of investments and the protection of competition. Both concerns should be analyzed when selecting the optimal institutional arrangement.

## 5. Conclusions

Agents may use the outside market as a discipline device to sanction (and prevent) post-contractual opportunism. It implies, on the one hand, that the allocation of RCRs, as proposed by the GHM approach, can be unnecessary or even undesirable. On the other hand, it means that the definition of RCRs should not only concern the relations between actual parties but also the importance of potential alternatives (i.e. outside options) in terms of incentives to invest in transforming the transaction.

More generally, the allocation of RCRs should take into account the FT process and a broader notion of transaction in which parties' investments as well as RCRs can affect market configuration (see also Nicita & Vatiere 2007).

## References

- Allen B., Deneckere R., Faith T., Kovenock D., "Capacity precommitment as a barrier to entry: a Bertrand-Edgeworth approach". *Economic Theory*, 15(3): 501-530.
- Anderson E., Bensaou M. (1999), "Buyer-supplier relations in industrial markets: When do buyers risk making idiosyncratic investments?" *Organizational Science*, 10:460-481.
- Bain J. S. (1956), *Barriers to new competition*. Cambridge: Harvard University Press.
- Coase R.H. (1960), "The Problem of Social Costs". *Journal of Law and Economics*, 3:1-44.
- Dixit A. (1980), "The role of investment in entry deterrence". *Economic Journal*, 90:95-106.
- Grossman S., Hart O. (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94: 691-719.
- Hart O. (1995). *Firms, Contracts and Financial Structure*. Oxford: Oxford University Press.
- Hart O., J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98: 1119-1158.
- Klein P.G. (1996), "Economic Calculation and the Limits of Organization," *The Review of Austrian Economics*, 9(2): 3-28.
- Klein B., Murphy K.M. (1997), "Vertical integration as a self-enforcing contractual arrangement," *American Economic Review*, 87(2):415-420.



## Specific investments and residual control rights: a reassessment\*

- Koss P. (1999), "Self-enforcing transactions: Reciprocal exposure in fisheries," *Journal of Law and Economics*, 43:105-141.
- Meccheri N. (2011), "Power, Law and Economics, and institutions [a review article of Understanding Power by M. Vatiéro], *History of Economic Ideas*, 19(2):201-212.
- Merrill T.W., Smith H.E. (2001), "What Happened to Property in Law and Economics", *Yale Law Journal*, 111: 357-398.
- Nicita A., Vatiéro M. (2007), "The contract and the market", *Studi e Note di Economia*, 1:7-22.
- Nicita A., Vatiéro M. (2014), "Dixit vs. Williamson: the 'Fundamental Transformation' Reconsidered", *European Journal of Law and Economics*, 37: 439-453.
- Pagano U., Vatiéro M. (2015), "Costly Institutions as Substitutes: Novelty and Limits of the Coasian Approach", *Journal of Institutional Economics*, 11(2, Ronald H. Coase memorial issue): 265-281.
- Raub W., G. Keren G. (1993), "Hostages as a commitment devices. A game-theoretic model and an empirical test of some scenarios," *Journal of Economic Behavior and Organization*, 21:43-67.
- Schelling T. (1960), *The strategy of conflict*, London: Oxford University Press.
- Stigler G. (1966), *The Theory of Price*. New York: Macmillan.
- Sylos-Labini P. (1969), *Oligopoly and Technical Progress*, Cambridge: Harvard University Press.
- Vatiéro M. (2009). *Understanding power: a 'Law and Economics' approach*. Saarbrücken: VDM Verlag.
- Williamson O.E. (1975), *Markets and Hierarchies, Analysis and Antitrust Implications*, New York: Free Press.
- Williamson O.E. (1979), "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics*, 2: 233-261.
- Williamson O.E. (1983), "Credible commitments: Using hostages to support exchange," *American Economic Review*, 73(4):519-540.
- Williamson O.E. (1985), *The Economic Institutions of Capitalism*. New York: Free Press.
- Williamson O.E. (1996), *The Mechanisms of Governance*. Oxford: Oxford University Press.
- Williamson O.E. (2005), "The Economics of Governance," *American Economic Review*, 95(2): 1-18.
- Williamson O.E. (2010), "Transaction Cost Economics: An Overview". In Klein P.G. and M.E. Sykuta, *The Elgar Companion to Transaction Cost Economics*, pp. 8-26. Edward Elgar.



# **Un database storico su imprese e settori dell'economia italiana, 1979–2015: nota metodologica\***

ALBERTO BIANCHI, CARLO BRAMBILLA, GIUSEPPE CONTI, FABIO LAVISTA,  
VALERIA PINCHERA

I saggi di Alberto Bianchi, Carlo Brambilla, Giuseppe Conti, Fabio Lavista, Valeria Pinchera e Raffaello Seri, pubblicati nel volume rappresentano il primo risultato di un progetto di ricerca sulle trasformazioni che hanno investito il sistema produttivo italiano negli ultimi cinquant'anni. Essi si avvalgono di un nuovo dataset realizzato attraverso l'acquisizione sistematica della totalità delle informazioni fornite dagli annuari di Mediobanca e di una loro organizzazione al fine di renderle facilmente accessibili e utilizzabili per varie finalità di ricerca.

Le raccolte di informazioni contabili ed extracontabili pubblicate annualmente da Mediobanca costituiscono ormai da tempo una delle principali fonti storiche sulla struttura produttiva italiana, una raccolta che consente di scendere nei dettagli delle situazioni aziendali e settoriali. Tutto ciò permette non solo un'accurata rappresentazione statica, ma anche dinamica lungo un percorso evolutivo che può essere, con gli opportuni interventi, ricostruito per ogni società e per comparti operativi. Per l'Italia – ma non solo per questo paese – tali annuari sono rari esempi di diffusione e organizzazione di informazione economica.

In particolare, Mediobanca pubblica raccolte annuali di informazioni sulle società italiane dall'inizio degli anni 1960 e dagli anni 1970 con maggior

\* Il progetto è nato da una collaborazione tra il Dipartimento di Economia dell'Università degli Studi dell'Insubria e il Dipartimento di Economia e management dell'Università degli Studi di Pisa. Per contatti: [carlo.brambilla@uninsubria.it](mailto:carlo.brambilla@uninsubria.it) e [giuseppe.conti@unipi.it](mailto:giuseppe.conti@unipi.it)

sistematicità. Le informazioni contabili sono alla base de *Le principali società italiane (Ps)*, pubblicato la prima volta nel 1966 e *Dati cumulativi (Dc)*, dal 1962. Il primo annuario riclassifica una selezione di dati di bilancio secondo uno schema omogeneo e comune a tutte le società considerate per gli ultimi due esercizi (quelli immediatamente precedenti all'uscita della pubblicazione). I conti riguardano lo stato patrimoniale e il conto economico, più alcune indicazioni extracontabili: settore produttivo prevalente, gruppo di appartenenza (dal 1995 in poi), numero di addetti, la posizione in graduatoria per volume di fatturato (o per raccolta nel caso delle banche), alcuni *ratios* finanziari e altro ancora. Nei *Dc* le singole poste contabili delle imprese sono aggregate per settore, in alcuni casi (non segnalati) facendo ricorso a consolidamenti di bilancio anche per anni nei quali i bilanci consolidati non erano richiesti. I due volumi si integrano, pur con alcune discrepanze significative. Gli annuari *Dc* presentano un ventaglio di poste contabili più accurato per lo stato patrimoniale e per il conto economico e forniscono in più un conto di variazioni patrimoniali e informazioni relative a una scomposizione dell'autofinanziamento. Gli annuari *Ps* forniscono una gamma di voci contabili più ristretta e negli annuari più recenti (dal 2002 in poi) senza più un bilanciamento tra attività e passività o tra costi e ricavi.

Le finalità "commerciali" degli annuari giustificano molte di queste scelte che possono rendere più difficili e meno accurati i confronti intertemporali. Lo scopo principale di tali pubblicazioni è infatti quello di offrire un calepino di consultazione per agevolare confronti su una sommaria struttura patrimoniale e sulla performance di ciascuna impresa nell'ultimo biennio, nel caso delle *Ps*. Le riclassificazioni in base a uno stesso schema contabile permette quasi a colpo d'occhio di individuare caratteristiche, elementi virtuosi e di criticità tra imprese diverse per prodotto e settore, ordinate per dimensione decrescente. Sebbene entrambi gli annuari raccolgano un campione piuttosto ampio delle maggiori società italiane (il cui numero nel caso dei *Dc* è specificato fin dal titolo dell'annuario), tuttavia i due campioni sono diversi e non raccolgono lo stesso numero di imprese. Benché in *Dc* manchino spesso indicazioni circa il numero di imprese comprese in ogni settore, esso raccoglie una serie decennale di dati contabili che permette un importante confronto temporale sugli andamenti settoriali e complessivi e su scomposizioni (anche in questo caso a numerosità non specificata) del campione in imprese pubbliche e private, industrie manifatturiere, imprese con risultato d'esercizio positivo, in pareggio o negativo ed altro ancora. Commenti di sintesi ai dati sono forniti nel rapporto che apre l'annuario. Investitori, azionisti e altri operatori economici hanno potuto quindi avvalersi di alcuni strumenti di base per avere

un'informazione utile, essenziale e preliminare al fine di orientare consapevolmente le proprie scelte.

Gli annuari non sono stati evidentemente concepiti per costruire serie storiche immediatamente utilizzabili per finalità analitiche e scientifiche su periodi lunghi. Per poterli utilizzare allo scopo di studiare nel lungo andare l'economia italiana è stato necessario, da un lato, tener conto dei criteri che ne hanno informato la compilazione, dall'altro, procedere a e/o prevedere una serie di interventi correttivi. I criteri alla base della compilazione di *Ps* infatti sono cambiati nel tempo. Ad esempio, la soglia d'ingresso per le imprese manifatturiere era stata inizialmente fissata a 10 miliardi di lire di fatturato, ma è stata via via innalzata a 20 miliardi nell'edizione 1986, a 25 in quella 1988. Dall'edizione 1996 le soglie sono stabilite in 50 miliardi per la singola impresa e 100 miliardi di lire per i gruppi (dal 1999 arrotondate rispettivamente a 25 e a 50 milioni di euro). Dall'annuario 2001, la soglia per le imprese industriali e di servizi (non finanziari) è stata quindi innalzata a 50 milioni di euro (sia imprese, sia gruppi). Nel caso delle aziende di credito, sin dall'inizio la soglia di ingresso fissata era pari a 50 miliardi di lire di depositi (o 100 miliardi di depositi interbancari). Analogamente ai dati relativi all'industria, dal 1999 la soglia d'ingresso è fissata rispettivamente a 25 e a 50 milioni di euro. Dal 2006 infine, la soglia viene portata a 50 milioni di attivo netto tangibile. Un'ulteriore discontinuità riguarda i cambiamenti di criteri contabili nel corso degli anni dipendenti dalle normative societarie ma anche dalle aggregazioni compiute per rendere confrontabili i conti nei singoli annuari. Una prima forte cesura nei criteri di riclassificazione tra gli annuari *Ps* 1980 e 1981 ha sinora impedito di estendere la raccolta e l'utilizzo dei dati pregressi e anteriori all'esercizio 1979 (annuario 1981 appunto). Altre rilevanti cesure si possono notare soprattutto dal 1991 con l'introduzione dello schema IAS (secondo lo schema di bilancio fissato nel d.l. 9 aprile 1991, n. 127), dal 1994 quando vengono introdotti bilanci consolidati a livello di gruppo, e nel periodo più recente, dal 2002 in avanti quando, oltre all'introduzione di nuovi criteri, gli annuari presentano un ventaglio molto più ridotto di voci contabili. Ciò è forse dovuto a ragioni di spazio editoriale a fronte di un aumento del numero di imprese non finanziarie e finanziarie censite. Di tali passaggi non sono fornite informazioni esauritive, tali cioè da facilitare il raccordo tra "nuovi" e "vecchi" dati (che come accennato non rientrava nelle finalità delle raccolte). Infine, occorre segnalare che le aggregazioni in settori presenti in *Ps* non collimano con quelle ricavabili da *Dc*, probabilmente a causa delle elaborazioni introdotte dai curatori in fase di aggregazione e consolidamento, dell'esclusione da *Ps* di imprese i cui conti sono invece inclusi in *Dc*, nonché probabilmente a causa della diversa strut-

tura dei due repertori: *Ps* infatti include un insieme variabile di imprese che superano una certa soglia dimensionale, laddove invece il campione proposto in *Dc* in ogni annuario è un insieme chiuso. Le imprese nel corso degli anni cambiano in vari casi e per varie ragioni il settore di appartenenza. Sui nuovi ingressi e sulle uscite dal campione mancano spesso informazioni. Negli annuari si dà comunque conto di cambiamenti di denominazione sociale, di fusioni e incorporazioni tra società.

Tali annuari non sono disponibili in formato digitale editabile, né - ovviamente - sono organizzati in time series (se non per gli ultimi anni). È stato quindi necessario raccogliere i dati dai volumi pubblicati da Mediobanca (in versione cartacea o pdf), e procedere alla immissione dei dati in fogli elettronici, anzitutto attraverso l'utilizzo di software a riconoscimento ottico dei caratteri (OCR). Soprattutto per gli annuari precedenti la seconda metà degli anni 1990, però, si è reso necessario un ricorso all'imputazione manuale dei dati (o alla correzione sistematica, quando necessario). Allo stato attuale, il dataset così ottenuto consta di una media annua di 1944 società industriali e 1105 società finanziarie, per un numero di campi di dati contabili di 2.479.164 (quasi 5 milioni considerando che ogni annuario mette a confronto 2 esercizi).

Sebbene la revisione del dataset, necessaria per verifiche e correzioni preliminari dovute a errori materiali e di stampa, di digitalizzazione, così come a cambiamenti nei criteri di classificazione e aggregazione contabile nel corso degli anni, sia già quasi del tutto ultimata, è tuttavia necessario procedere ad alcune rilevanti integrazioni. Da un lato, è necessaria (e già in fase di implementazione) l'introduzione di chiavi primarie che permettano l'identificazione univoca delle imprese, in modo che sia possibile seguirne l'evoluzione nel tempo. Parallelamente, e a completamento delle informazioni contabili e quantitative già raccolte, si prevede di integrare il dataset con le informazioni relative ai cambiamenti - a livello di singola impresa - in ordine a denominazione sociale, ristrutturazioni/ricostituzioni patrimoniali, fusioni e acquisizioni, ecc., tutte informazioni presenti negli annuari, ma che devono essere raccolte "manualmente". Anche per questo è indispensabile procedere anzitutto all'identificazione univoca delle imprese. È infine necessario cercare di integrare i dati per il periodo precedente il 1981 e, soprattutto, successivo il 2002, per il quale ultimo - come s'è detto - le *Ps* sono particolarmente poche di informazioni.

Per un'analisi storica accurata sotto ogni rispetto occorrerà un'ulteriore affinamento della critica della fonte da incrociare con altre e consentire così un'analisi sistematica dei bilanci e una ricostruzione coerente di serie storiche (per voci di stato patrimoniale e conto economico, per informazioni extracon-

Un database storico su imprese e settori dell'economia italiana, 1979-2015:

tabili) in sintonia con le informazioni macroeconomiche e settoriali di fonti ufficiali (Istat, Banca d'Italia e altre). La definitiva sistemazione della continuità delle singole società (ragioni sociali che cambiano nel tempo, fusioni, acquisizioni, ristrutturazioni aziendali in strutture di gruppo, ecc.) e la revisione della suddivisione per settori presente nella fonte e riclassificazione attraverso sistemi standard (Ateco, Isic) consentirà finalmente un accesso rapido alla mole di informazioni raccolte nel database.





# Settore bancario e frontiere di efficienza

CARLO BRAMBILLA E RAFFAELLO SERI\*

## 1. Introduzione

L'ultimo ventennio del Novecento è stato una fase di profondi cambiamenti per il sistema bancario italiano. Alla progressiva deregolamentazione del sistema, culminata con la legge bancaria del 1993, hanno fatto seguito dapprima un rapido processo di privatizzazione, quindi – dalla seconda metà degli anni 1990 – un'ondata di fusioni ed acquisizioni che hanno coinvolto pressoché l'intera industria bancaria. Tutto ciò, insieme alle modalità con le quali quei processi si sono sviluppati, ha avuto effetti sia sulla struttura del sistema nel suo complesso, sia sui singoli intermediari. Al primo e più immediatamente evidente di tali effetti, un piuttosto rapido processo di consolidamento e concentrazione del settore, possono essere ascritte due importanti conseguenze. Sebbene analoghi processi di consolidamento dei sistemi bancari nazionali si siano infatti verificati in numerosi altri paesi nel corso degli ultimi quarant'anni, la quasi scomparsa del comparto degli intermediari non-for-profit – oggi rappresentato solo da un pulviscolo di piccole banche di credito cooperativo – si configura come una peculiarità italiana. Inoltre, sostanzialmente nessuna delle grandi banche storiche più prestigiose e rilevanti nella storia economica italiana è sopravvissuta al rimescolamento prodottosi nel sistema in seguito alle privatizzazioni e alle fusioni degli anni 1990 (Brambilla, Conti, 2016).

Quelle profonde trasformazioni hanno generato una vasta letteratura che è andata arricchendosi dall'inizio del secolo, via via che la direzione e gli effetti dei cambiamenti in atto divenivano più chiari. Secondo molti studi, la deregolamentazione e le liberalizzazioni avrebbero esposto un sistema bancario impreparato

\* Il saggio è stato concepito congiuntamente dai due autori. In particolare, tuttavia, i paragrafi 1, 2, 4 e 5 sono da attribuire a Carlo Brambilla, l'analisi dei dati e il paragrafo 3 a Raffaello Seri.

al rischio di essere fagocitato dalla concorrenza estera. L'ondata di fusioni ed acquisizioni che fecero seguito alle privatizzazioni, quindi, sarebbero state allo stesso tempo una strategia difensiva e la premessa necessaria per incentivare la ricerca del profitto. La trasformazione delle aziende di credito e degli enti creditizi in imprese a scopo di lucro, in concorrenza tra loro su un mercato liberalizzato, avrebbe stimolato l'efficienza e la produttività, sia delle singole imprese, sia del settore nel suo complesso. Fusioni e acquisizioni dunque sarebbero state il mezzo attraverso il quale realizzare ristrutturazioni produttive e perseguire processi di integrazione verticale e orizzontale nell'intera industria finanziaria. La miglior performance e nuovi e più solidi gruppi bancari avrebbero infine condotto a un sistema più stabile ed efficiente (Ciocca, 2005; Giordano, 2007).

Obiettivo di questo lavoro è studiare l'impatto che i cambiamenti nell'architettura del sistema e le riforme del quadro istituzionale degli anni 1990 hanno avuto sull'efficienza dell'industria bancaria italiana, e tentare quindi una verifica empirica di come i processi di deregolamentazione e di liberalizzazione del mercato del credito, nonché la concorrenza tra intermediari, abbiano condotto a un incremento di efficienza del sistema e di singole banche. Adottando una prospettiva di lungo periodo, lo studio utilizza un dataset che raccoglie i dati di bilancio di fine esercizio di un elevato numero di intermediari – in termini di asset il campione copre la quasi totalità del sistema – per il periodo 1979-2000. L'analisi è basata sulla stima nonparametrica dell'insieme produttivo attraverso la Data Envelopment Analysis (DEA).

Il saggio è articolato in quattro paragrafi oltre l'introduzione. Nel secondo paragrafo si richiama il contesto e si discutono le riforme e i cambiamenti introdotti nel sistema bancario dagli anni 1980, fornendo una breve sintesi degli esiti dei processi di privatizzazione e consolidamento del sistema. Segue una breve descrizione del dataset e la discussione della metodologia utilizzata, nel paragrafo 3. Il quarto paragrafo illustra sinteticamente e commenta i risultati dell'analisi. Infine, alcune prime conclusioni e proposte per ulteriori sviluppi della ricerca sono oggetto del quinto paragrafo.

## **2. Ristrutturazione e trasformazione dell'industria bancaria**

Nel corso degli anni 1970, l'architettura del sistema bancario e finanziario italiano, introdotta quarant'anni prima ed implementata nel corso del dopoguerra, cominciò a mostrare i propri limiti e a rivelarsi sempre meno adeguata ai

cambiamenti in atto. Quell'impianto regolativo aveva come obiettivo primario la stabilità del sistema ed era quindi fondato sulla segmentazione del mercato creditizio e sulla specializzazione funzionale degli intermediari, consentiva una concorrenza tra banche molto limitata e imponeva uno stretto controllo sugli intermediari da parte delle autorità monetarie e di vigilanza, in particolare per contenerne le dimensioni e le quote di mercato (specie delle grandi banche).

I mutamenti nel quadro macroeconomico e istituzionale, con il crollo del sistema monetario internazionale di Bretton Woods, seguito da una forte accelerazione della dinamica dei prezzi, e la progressiva integrazione dei mercati internazionali colpirono le banche italiane, non diversamente da quanto avveniva in altri paesi industrializzati. Mentre la riduzione dei tassi di interesse reali ne assottigliava i margini di profitto, la concorrenza delle banche estere – favorita dallo sviluppo dei mercati internazionali dei capitali e dall'incipiente rivoluzione nelle tecnologie dell'informazione e della comunicazione – minacciava di ridurne ulteriormente la redditività. Gli effetti della concorrenza erano rafforzati dalle politiche promosse in sede comunitaria volte alla progressiva integrazione e liberalizzazione dei mercati finanziari e creditizi nazionali, anche attraverso l'insediamento delle banche in altri paesi membri che – rendendo ancor più drammatico lo svantaggio competitivo di intermediari come quelli italiani, costretti entro i limiti di una normativa molto rigida – avrebbe favorito la “convergenza istituzionale” tra i paesi europei (Dermine, 2003).

D'altro canto, la segmentazione dei mercati creditizi aveva finito per incidere negativamente sull'efficienza degli intermediari italiani, molti dei quali dagli anni 1970 erano divenuti inefficienti ed avevano perso capacità allocativa, mentre non infrequenti erano i fenomeni di ingerenze partitiche, clientelismo e corruzione, alcuni dei quali culminati in scandali e fallimenti eclatanti (Spero, 1980; Bellavite Pellegrini, 2002). Il coinvolgimento degli Istituti di credito speciale nella crisi dell'industria chimica, che ne causò il dissesto dissipando risorse pubbliche, è un'ulteriore evidenza della crisi nella quale era caduto e si dibatteva il settore (Barca, Trento, 1997). Per rivitalizzare un sistema che appariva ingessato e incapace di risollevarsi, decisi cambiamenti nella struttura dei mercati creditizi apparivano necessari. Si optò quindi, spinti anche dalle necessità imposte dalle mutate condizioni internazionali e dalla direzione che andavano prendendo l'innovazione tecnologica e quella finanziaria, per una soluzione che, permettendo la crescita nella dimensione media degli intermediari e liberalizzando l'offerta di servizi finanziari, stimolasse la concorrenza, riducesse i costi operativi e desse impulso all'incremento sia della profittabilità, sia dei mezzi propri delle banche.

Nel corso del decennio successivo dunque, liberalizzazione finanziaria e integrazione europea procedettero sostenendosi vicendevolmente. Mentre la Banca d'Italia anticipò il legislatore, sostenendo e promuovendo il cambiamento attraverso i propri poteri regolativi e la raccomandazione che gli enti pubblici creditizi dovessero operare in un ambiente competitivo, essere liberi di definire la propria offerta e affidarsi al mercato per gli aumenti di capitale (Banca d'Italia, 1981)<sup>1</sup>, lentamente anche le direttive europee di coordinamento bancario furono tradotte in legge. Le raccomandazioni della banca centrale furono recepite dalla legge 23/1981, mentre poco dopo la legge 74/1985 riconosceva agli intermediari creditizi il carattere di imprese a scopo di lucro e la libertà di stabilirsi in qualsiasi paese della Comunità. Nel 1989 furono rimosse le ultime limitazioni al libero insediamento delle banche estere in Italia, mentre l'anno successivo furono abolite le restrizioni valutarie e sulla apertura di sportelli bancari. Dopo il recepimento della seconda direttiva comunitaria di coordinamento bancario (1992), fu infine promulgata la nuova legge bancaria (1993). Con quest'ultima furono rimosse le ultime restrizioni volte alla segmentazione e specializzazione dell'offerta di credito, di fatto reintroducendo la banca universale nel sistema italiano; l'erogazione del credito cessava di essere considerata un servizio pubblico; venivano cancellate le forme societarie ibride imponendo alle banche – ormai considerate imprese dedite ad attività imprenditoriali a tutti gli effetti – di assumere la forma di società di capitali (per azioni o cooperative) (Galanti et al., 2012).

All'inizio degli anni 1990, il paese si dibatteva in gravi difficoltà: il cumularsi di annosi problemi quali il pesante deterioramento della finanza pubblica e la crisi, ormai più che decennale, dell'industria pubblica e del sistema delle Partecipazioni statali, con quella del sistema politico-istituzionale e lo scoppio di una violenta crisi valutaria, rendeva non più procrastinabili interventi e riforme radicali. Una – se non la principale – risposta a tale situazione fu la decisione di avviare un ampio e rapido processo di privatizzazione dell'impresa pubblica, con l'obiettivo primario di alleviare le tensioni sul debito e sulla finanza pubblica. D'altro canto, le privatizzazioni erano la premessa necessaria all'effettiva liberalizzazione dei mercati, avviata sin dal decennio precedente. Così facendo, inoltre, seppur con ritardo rispetto agli altri paesi

<sup>1</sup> Le “raccomandazioni” della Banca d'Italia assumono particolare rilevanza se si considera che, dopo i salvataggi e le riforme del sistema introdotte dagli anni 1930, più di due terzi del sistema bancario era divenuto direttamente o indirettamente di proprietà e/o a controllo pubblico, e che ancora nel 1994 lo stato possedeva circa il 62% del sistema, in termini di attivo (Messori, 2009; Brambilla, Conti, 2011).

europei, si avviava un processo di trasformazione del sistema economico italiano in linea con i trattati e le politiche comunitarie cui l'Italia aderiva. Nel giro di pochi anni, il quadro istituzionale cambiò radicalmente. Per il sistema bancario, la svolta fu la trasformazione delle banche di proprietà o a controllo pubblico in società di capitali (l. 218/1990), che pose le premesse all'apertura del mercato per il controllo societario nell'industria bancaria. Tra 1992 e 1994, una serie di interventi normativi prepararono l'avvio delle dismissioni dell'industria pubblica: nel 1992 gli enti economici pubblici (Iri, Eni, Ina, Enel) furono trasformati in società per azioni, furono definiti piani di riorganizzazione e obiettivi strategici, nonché le norme per la loro dismissione sul mercato; l'anno successivo fu abolito il ministero delle Partecipazioni statali; infine nel 1994, la legge 474 definì il quadro normativo generale e le regole con le quali procedere alle privatizzazioni. Dal 1992 al 2000, la maggior parte delle imprese industriali e delle banche possedute dallo stato furono privatizzate: in termini sia di rapidità, sia di dimensioni si trattò del più ampio processo di privatizzazione mai verificatosi nei paesi occidentali (Bemporad, Reviglio, 2001; Inzerillo, Messori, 2000).

Nel settore bancario alle privatizzazioni seguì immediatamente un rapido processo di consolidamento, che coinvolse quasi l'intero settore a partire dalla metà del decennio con effetti sull'intera industria bancaria. Tra 1990 e 2000 mutarono radicalmente la geografia del sistema e la struttura dimensionale degli intermediari: il numero di intermediari si ridusse del 30 per cento circa; gran parte dei marchi storici sparirono assorbiti all'interno di nuovi grandi gruppi bancari; si produsse una polarizzazione dimensionale piuttosto forte, con una caduta nel numero di intermediari di medie dimensioni e la crescita delle dimensioni medie dei gruppi maggiori. Meno chiaro è l'esito che tali fenomeni di consolidamento hanno avuto sulla performance delle imprese e dell'industria. Alcuni studi hanno infatti sottolineato come vi sia stato un incremento della concorrenza nel settore a partire dal processo di consolidamento, ovvero che la concentrazione del sistema non abbia portato a un aumento del potere di mercato, bensì a una riduzione dei costi operativi e al miglioramento della qualità degli attivi bancari (Angelini, Cetorelli, 2003; Focarelli et al., 2002). Tuttavia in altri contributi i risultati sembrano più controversi<sup>2</sup>. In particolare, Fiordelisi (2009) ha analizzato gli effetti delle fusioni ed acquisizioni

<sup>2</sup> Si vedano, tra gli altri, Bonaccorsi di Patti, Gobbi (2001); Sapienza (2002); Focarelli, Panetta (2003); Messori et al. (2003).

sull'efficienza delle banche nei quattro maggiori paesi europei, tra 1995 e 2002. Secondo questo studio le banche italiane hanno sperimentato una riduzione dell'efficienza di costo nei cinque anni seguenti l'operazione di fusione/acquisizione, a causa delle rigidità nel mercato del lavoro che hanno reso l'integrazione e la riorganizzazione post fusione, e quindi il conseguimento di economie di scala e il rilancio dell'operatività su linee di business, prodotti e servizi più remunerativi, più arduo. Messori (2009), pur riscontrando un incremento di efficienza e competitività nei mercati creditizi, ha sottolineato come i modelli di *governance* e di mercato per il controllo societario prevalenti, sostanzialmente immutati rispetto al passato, abbiano impedito di migliorare l'efficienza complessiva del sistema. In questo lavoro si adotta una prospettiva di più lungo periodo rispetto alla gran parte degli studi sull'argomento e si tenta, a partire da una base di dati ampia e sinora poco utilizzata, di verificare empiricamente l'evoluzione nell'efficienza degli intermediari bancari nell'ultimo trentennio del Novecento.

### **3. Dataset e metodologia**

#### **3.1. Il dataset**

Parte di un più ampio progetto di ricerca sulla trasformazione del sistema produttivo italiano a partire dagli anni 1970, questo lavoro si basa anch'esso su di un dataset costruito a partire dagli Annuari pubblicati da Mediobanca, *Le principali società italiane* e relativo all'intero periodo 1979-2000. Dati e grafici presenti nel testo (ove non indicato diversamente) sono nostre elaborazioni sul dataset. Il campione comprende le banche con depositi pari a o maggiori di 50 miliardi di lire (dal 1999, 25 milioni di euro) o raccolta interbancaria pari o superiore ai 100 miliardi di lire (dal 1999, 50 milioni di euro). In termini di numero di imprese, il dataset copre il 40 per cento circa dell'intero sistema bancario all'inizio degli anni 1980 e via via quote maggiori sino al 76 per cento nel 2000, con una copertura, in termini di raccolta e di totale attivo, nell'ordine dell'80-90 per cento del sistema.

#### **3.2. La metodologia**

Per spiegare nel dettaglio l'analisi che svolgeremo nel seguito, è necessario introdurre alcune definizioni di teoria della produzione. Siano  $x$  e  $y$  due vettori nonnegativi di dimensioni  $p$  e  $q$  che rappresentano, rispettivamente, le quan-

tà di input ed output del processo produttivo. L'insieme produttivo o tecnologico si definisce come:

$$\Psi = \{(x, y): x \text{ è in grado di produrre } y\}$$

La *frontiera di produzione* o di *efficienza* si definisce come la frontiera di questo insieme. In genere si suppone che  $\Psi$  rispetti alcune assunzioni. La prima, detta *free disposability*, richiede che, se gli input  $x$  sono in grado di produrre gli output  $y$ , ossia  $(x, y) \in \Psi$ , allora passando ad  $x'$  e  $y'$  con  $x' \geq x$  e  $y' \leq y$ , si ha  $(x', y') \in \Psi$ . Ciò significa che, se è possibile produrre gli output  $y$  con gli input  $x$ , è anche possibile produrre una quantità minore o uguale di  $y$  con una quantità maggiore o uguale di  $x$ . Esistono nella letteratura delle eccezioni a questa ipotesi, ma sono rare. La seconda assunzione, detta *convessità*, richiede che se  $(x, y) \in \Psi$  e  $(x', y') \in \Psi$  allora anche tutte le combinazioni lineari convesse di  $(x, y)$  e  $(x', y')$  appartengono a  $\Psi$ . A differenza della precedente, questa ipotesi non è universalmente accettata.

Ogni unità produttiva corrisponde ad un punto  $(x, y)$  appartenente a  $\Psi$ , e la distanza di questo punto rispetto alla frontiera di  $\Psi$  è una misura dell'efficienza nell'uso degli input. La ragione è che, quando un punto  $(x, y)$  è lontano dalla frontiera di  $\Psi$ , gli input  $x$  possono essere ridotti senza alterare gli output  $y$ . Sebbene nella letteratura esistano vari concetti alternativi di efficienza, noi ci concentreremo sull'*efficienza tecnica*, ossia l'efficienza con la quale un vettore di input  $x$  viene utilizzato per produrre un livello di output  $y$ . La definizione che useremo (normalmente attribuita a Farrell, 1957, sebbene fosse già stata introdotta da Debreu, 1951, sotto il nome di *coefficient of resource utilization*) è basata sugli input ad output fissi<sup>3</sup>:

$$\theta(x, y) = \inf \{\theta: (\theta x, y) \in \Psi\}$$

Questa misura di efficienza tecnica è quindi definita come il minimo numero compreso tra 0 e 1 per il quale si può moltiplicare il vettore di input  $x$  pur restando in grado di produrre lo stesso livello di output  $y$ . Un valore prossimo a 1 indica efficienza, e man mano che i valori si avvicinano a 0 l'inefficienza

<sup>3</sup> È possibile definire misure simmetriche, che mantengano fissi gli input e varino gli output, ma non le discutiamo perché veicolano più o meno la stessa informazione di quelle basate sugli input ma, mentre queste ultime sono limitate tra 0 e 1, quelle basate sugli output sono uguali a 0 o maggiori di 1. La loro interpretazione è perciò più difficoltosa.

umenta. Un'impresa (o più generalmente un'unità produttiva) si dice tecnicamente efficiente (secondo Farrell) se non è possibile ridurre proporzionalmente tutti gli input nella stessa misura pur continuando a produrre lo stesso livello di output. Le osservazioni sulla frontiera sono tecnicamente efficienti poiché hanno  $\theta(x, y)$  pari a 1.

In genere l'insieme produttivo  $\mathcal{P}$  non è noto, ma può essere ricostruito a partire da un campione di  $n$  unità produttive delle quali venga osservata la combinazione di input ed output,  $(x_i, y_i)$  per  $i = 1, \dots, n$ . Il problema della stima di  $\mathcal{P}$  dalle osservazioni è stato affrontato nella letteratura con numerose tecniche. In questo lavoro considereremo una delle tecniche che richiedono meno ipotesi sulla forma della frontiera di  $\mathcal{P}$ .

La *Data Envelopment Analysis* (DEA) è una metodologia nonparametrica di stima dell'insieme produttivo e della frontiera di produzione (Bogetoft, Otto, 2011). Originariamente sviluppata in ricerca operativa (Charnes, Cooper, Rhodes, 1978, sulla base di Farrell, 1957) senza fare esplicito riferimento ad un processo probabilistico di generazione dei dati (Data Generating Process, DGP), la DEA è stata più recentemente riformulata come tecnica di stima dell'insieme produttivo  $\mathcal{P}$  quando si suppone che i dati siano stati generati da un modello definito in termini statistici (Daraio, Simar, 2007). Questo ha reso disponibili una serie di strumenti (come il bootstrap) che hanno consentito di ampliare il campo delle analisi statistiche realizzabili e delle ipotesi testabili su un dataset. Ciò ha reso possibile non solo la stima dell'efficienza di una unità produttiva, ma anche la correzione del bias di questo stimatore (cfr. sotto per maggiori dettagli), e lo svolgimento di test di ipotesi sull'efficienza di una singola unità o sulle assunzioni usate nella stima stessa.

Esistono in realtà più varianti dello stimatore DEA, ognuna delle quali si caratterizza per il fatto di richiedere che l'insieme produttivo stimato rispetti una assunzione circa la forma dei rendimenti di scala dell'industria che è oggetto dello studio. Il più semplice degli stimatori DEA, detto DEA-VRS (*Varying Returns to Scale*, rendimenti di scala variabili), stima l'insieme produttivo come il più piccolo insieme  $\hat{\mathcal{P}}$  che contiene i dati e rispetta le due proprietà di convessità e *free disposability*<sup>4</sup>. Altri stimatori aggiungono alle due precedenti una ulteriore assunzione circa il fatto che i rendimenti di scala siano costanti (CRS, *Constant Returns to Scale*), crescenti (IRS, *Increasing Returns to Scale*) oppure decrescenti (DRS, *Decreasing Returns to Scale*). In questa sede non entreremo nei dettagli computazionali delle diverse varianti. L'ipotesi di rendimenti decre-

<sup>4</sup> Per una trattazione si rimanda a Bogetoft, Otto (2011).



scenti di scala è adatta quando è naturale supporre che non esistano vantaggi tecnologici derivanti dall'aumento della scala di produzione, mentre l'ipotesi di rendimenti crescenti di scala è accettabile quando gli effetti negativi dell'aumento della scala sono più che controbilanciati dagli effetti positivi. La scelta di una particolare assunzione circa i rendimenti di scala è meno rilevante quando il numero di osservazioni è elevato, perché ciò consente una descrizione della frontiera accurata anche in assenza di assunzioni aggiuntive. Quindi, sia perché il numero di osservazioni su cui lavoreremo sembra essere sufficiente, sia perché non sembra ragionevole fare ipotesi sui rendimenti di scala, nel seguito ci limiteremo a considerare lo stimatore DEA-VRS.

Lo stimatore DEA dell'insieme produttivo  $\widehat{\Psi}$  consente di ottenere un stimatore dell'efficienza tecnica  $\widehat{\theta}(x, y)$  per una generica combinazione di input ed output  $(x, y)$ , semplicemente sostituendo  $\widehat{\Psi}$  a  $\Psi$  nella definizione di  $\theta(x, y)$ :

$$\widehat{\theta}(x, y) = \inf \{ \theta : (\theta x, y) \in \widehat{\Psi} \}$$

Il caso più interessante è quello in cui  $(x, y) = (x_i, y_i)$  per un certo indice  $i$ , perché fornisce l'efficienza tecnica di una unità produttiva del campione. I dati osservati giacciono sempre all'interno della frontiera di produzione reale, o al limite si posizionano su di essa. Pertanto, se le assunzioni di base del metodo di stima sono verificate, l'insieme di produzione stimato cade anch'esso all'interno dell'insieme di produzione reale. Ciò implica che le efficienze tecniche stimate siano sempre maggiori rispetto a quelle reali. Questa distorsione sistematica può essere corretta usando un algoritmo di bootstrap (Daraio, Simar, 2007; Simar, Wilson, 1998).

In questo lavoro utilizzeremo la DEA-VRS per il calcolo delle efficienze tecniche. Data la presenza di distorsioni sistematiche nelle efficienze stimate, correggeremo tale distorsione usando il bootstrap. Prima di procedere è necessario definire gli input e gli output del processo produttivo.

### 3.3. La scelta di input e output

Solitamente nello studio dell'efficienza di imprese, comparti o interi settori, la definizione e la scelta degli input e degli output è piuttosto semplice ed immediata. Ciò non è vero nel caso dell'industria bancaria, poiché a seconda di come si intende definire l'attività degli intermediari che si stanno studiando, così come a seconda del tipo di domanda cui si vuole rispondere, possono essere considerati input e output diversi. Nella letteratura sull'efficienza del sistema bancario, si possono identificare cinque approcci principali all'identificazione degli input e degli output del processo produttivo

di un'impresa bancaria (Favero, Papi, 1995). Occorre notare che i confini tra questi approcci sono spesso labili e scelte simili di input ed output possono derivare da approcci diversi.

Nell'*intermediation approach* (Colwell, Davis, 1992, p. S113) le banche sono considerate intermediari puri tra agenti economici in surplus ed agenti in deficit di risorse finanziarie. Pertanto, prestiti ed investimenti finanziari sono gli output del processo produttivo i cui input sono lavoro e capitale. I depositi possono essere considerati di volta in volta come input o output. L'aspetto negativo di questo approccio è che misura l'attività bancaria attraverso quantità di stock e non di flusso. L'*asset approach* (Berger, Humphrey, 1992, p. 247) può essere considerato una variante dell'approccio precedente, nella quale gli output vengono scelti tra le voci dell'attivo della banca, il che elimina i depositi dalla lista dei possibili output. Il principale difetto di questo approccio è che gran parte dell'attività delle banche, ossia la produzione di depositi, non viene considerata in alcun modo, né dagli input né dagli output. Nel *production approach* (Colwell, Davis, 1992, p. S113) le banche sono imprese che usano lavoro e capitale (input) per produrre operazioni su prestiti e depositi (output). In genere gli output vengono misurati non come ammontare monetario, ma attraverso il numero di operazioni e/o il numero di conti bancari, dati spesso non disponibili pubblicamente e di assai difficile reperimento, specie per periodi più lontani. Ciò rende questo approccio particolarmente utile per confrontare diverse filiali all'interno della stessa banca, ma di difficile applicazione per il confronto tra banche diverse e per studi di carattere storico.

Lo *user-cost approach* determina l'appartenenza dei prodotti finanziari agli input o agli output sulla base del contributo netto al fatturato della banca. La difficoltà di misurare accuratamente prezzi e costi delle varie voci contabili e la sensibilità della classificazione di input e output rispetto a piccole variazioni degli stessi ne fanno un approccio difficile da applicare in pratica. Infine, secondo il *value-added approach* tutte le voci di attivo e passivo possono essere considerate come output del processo produttivo. La classificazione delle voci contabili come output avviene sulla base del contributo di ciascuna voce al valore aggiunto. Le voci che contribuiscono maggiormente al valore aggiunto vengono identificate come output, le altre possono essere prodotti intermedi o input. Ciò porta in genere all'identificazione di capitale, lavoro, raccolta interbancaria e altre forme di raccolta esclusi i depositi (*purchased funds*) come input, e di impieghi diretti, depositi e redditi da commissioni e investimenti come output. Si noti che benché i depositi e le altre forme di raccolta assolvano la stessa funzione all'interno della banca, essi vengono considerati in modo diverso a causa della diversa quantità di input fisici che richiedono (Berger, Humphrey, 1992, pp. 248-249).

Il *value-added approach* è generalmente considerato in letteratura il più adatto allo studio del cambiamento tecnologico nel tempo. Dato che obiettivo di questo lavoro è cercare di definire l'impatto dei cambiamenti, delle riforme e della riorganizzazione del sistema sull'efficienza dell'industria stessa e delle sue componenti, si è scelto di adottare questo tipo di approccio

Nel seguito, come input utilizzeremo lavoro e capitale, misurati rispettivamente dal numero di impiegati e dagli immobilizzi tecnici (immobili, mobili e impianti netti). Come output utilizzeremo gli impieghi diretti, i depositi o raccolta da clientela, ed il margine da servizi. Le variabili sono simili a quelle che Resti (1997), per esempio, usa all'interno del *value-added approach*, ma la loro scelta è giustificabile anche nel quadro concettuale fornito dall'*intermediation approach*. Le variabili sono riassunte nella Tabella 3.1<sup>5</sup>.

Tabella 3.1. Definizione delle variabili		
Variabile	Nome	Definizione della variabile
$y_1$	Impieghi diretti	Impieghi verso clientela
$y_2$	Depositi	Raccolta da clientela
$y_3$	Margine da servizi	Saldo commissioni attive e passive; utili e perdite da operazioni finanziarie
$x_1$	Lavoro	Numero di dipendenti
$x_2$	Capitale fisico	Immobili, mobili e impianti netti

## 4. Analisi d'efficienza, 1979–2000

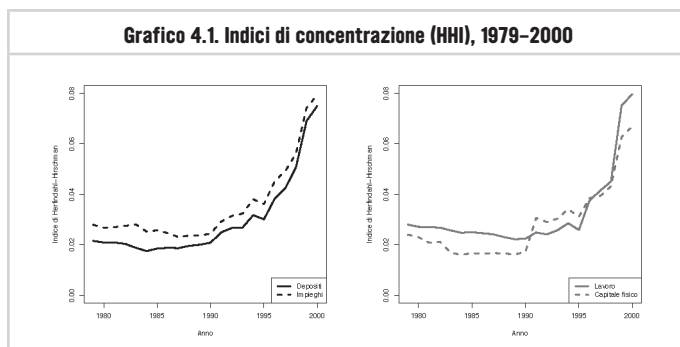
Come s'è detto, obiettivo di questo lavoro è l'analisi degli effetti delle riforme e dei profondi cambiamenti nell'architettura istituzionale e nel quadro normativo, introdotti nell'industria bancaria a partire dai primi gradualisti interventi di liberalizzazione degli anni 1980 e culminati nel decennio successivo nella promulgazione del Testo unico bancario (1993) e nella privatizzazione del sistema. Ciò permise e diede il via al processo di consolidamento delle banche italiane attraverso quell'ondata di *mergers & acquisitions* che caratterizzò gli anni 1990 e i primi anni 2000. Prima di illustrare i risultati dell'analisi di effi-

<sup>5</sup> Dall'analisi è esclusa, per tutto il periodo, la Cassa di sovvenzioni e risparmio tra il personale della Banca d'Italia, poiché date le sue grandi dimensioni e il bassissimo numero di addetti in relazione al livello degli output, la sua inclusione distorcerebbe pesantemente il calcolo dei livelli di efficienza dell'intero campione.

cienza, e dei cambiamenti che sembrano emergere dai dati a questo proposito, è bene quindi soffermarsi brevemente sugli effetti che ciò ebbe sulla struttura dell'industria bancaria.

Il processo di consolidamento del sistema portò a una caduta di quasi il 30 per cento nel numero di intermediari operanti in Italia, tra 1990 e 2000 (Banca d'Italia, 2001). L'effetto fu una trasformazione della struttura del mercato creditizio nel senso di un rapido aumento della concentrazione, come risulta chiaramente dal Grafico 4.1, che riporta gli indici di Herfindahl-Hirschman calcolati sull'intero campione per il periodo 1979-2000.

Il grafico di sinistra riporta gli indici di concentrazione rispetto alla raccolta e agli impieghi diretti, due tradizionali indicatori dimensionali dell'impresa bancaria. Il livello di concentrazione del sistema si mantiene basso sino a tutti gli anni 1980, evidenziando anche una flessione nella prima metà del decennio. Dalla fine degli anni 1980 gli indici ritornano crescenti, una crescita che si accentua visibilmente dal 1990, prima ancora che abbiano inizio le privatizzazioni, per poi impennarsi dal 1995 sino a fine periodo<sup>6</sup>.



Il grafico di destra riporta gli indici di concentrazione del sistema in termini di forza lavoro e di capitale fisico (cfr. Tabella 3.1). I due indici mostrano un andamento simile e coerente con quelli riportati nel grafico di sinistra, con un andamento che tuttavia rimane sostanzialmente stabile sino alla fine degli anni 1980. Essi sembrano fornire qualche indicazione aggiuntiva alle trasfor-

<sup>6</sup> Presumibilmente ciò è dovuto alle operazioni di ristrutturazione e consolidamento all'interno di alcuni gruppi di proprietà pubblica in vista, appunto, del lancio della privatizzazione. Un esempio è il caso Banco di Roma-Banco di Santo Spirito-Cassa di risparmio di Roma (Brambilla, Conti, 2016).

mazioni strutturali in atto all'epoca. Infatti l'indice relativo al capitale fisico risulta sensibilmente più basso di quello relativo agli addetti sino al 1990, presentando quindi un "salto" tra 1990 e 1992 che lo porta a assestarsi su un livello stabile e più elevato rispetto all'altro fino alla metà del decennio. Ciò è probabilmente dovuto ai rigidi vincoli all'apertura di dipendenze e forse anche a un ritardo nell'adozione e/o rinnovo delle infrastrutture informatiche. Il "salto" all'inizio del decennio 1990 registrerebbe in questo senso, da un lato la completa liberalizzazione nell'apertura degli sportelli, dall'altro interventi di ristrutturazione e riorganizzazione in vista delle privatizzazioni<sup>7</sup>. L'indice relativo agli addetti rimane sostanzialmente costante sino al 1995. Dal 1995, entrambi presentano una forte crescita, che ricalca quella degli indici relativi a depositi e impieghi. La ragione del disallineamento dei due indici negli anni finali del periodo, con una più marcata concentrazione del sistema in termini di addetti, rispetto agli immobilizzi tecnici, è forse dovuta alla difficoltà per i nuovi gruppi frutto delle fusioni di operare rapidamente riorganizzazioni produttive in presenza di rigidità nel mercato del lavoro (Fiordelisi, 2009).

I grafici dal 4.2 al 4.6 mostrano i risultati dell'analisi di efficienza condotta attraverso la DEA. L'analisi è stata condotta per ognuno degli anni del periodo, tuttavia, per ragioni di spazio, qui se ne mostrano solo alcuni. Di seguito quindi sono riportati i risultati relativi a cinque anni benchmark significativi. Il Grafico 4.2 è relativo all'inizio del periodo (1980), segue il 4.3 che mostra la situazione a metà anni 1980, quando furono introdotte le prime graduali riforme. Il Grafico 4.4 fotografa la situazione alla vigilia del processo di privatizzazione, e all'indomani di alcuni rilevanti interventi normativi di liberalizzazione (1990), mentre il 4.5 registra i cambiamenti alla metà del decennio 1990, quando le maggiori banche sono state privatizzate ed è iniziato il processo di consolidamento del settore. L'ultimo, il Grafico 4.6, fotografa la situazione a fine periodo (2000).

Ciascun grafico riporta due *scatterplots*: il primo, a sinistra, rappresenta l'efficienza tecnica rispetto alla dimensione degli intermediari (depositi); il secondo, a destra, mostra la proporzione degli input – capitale fisico e lavoro – impiegati nella produzione da ciascuna banca<sup>8</sup>. Ne *Le principali società italiane* le banche sono suddivise in tipologie che rispecchiano la normativa vigente all'epoca della pubblicazione dei singoli annuali, e dunque la segmentazione

<sup>7</sup> Contemporaneamente alla progressiva diminuzione del numero di intermediari bancari, nel corso degli anni 1990 crebbe sensibilmente il numero complessivo di sportelli. Cfr. Banca d'Italia, *Relazioni Annuali*, vari anni.

<sup>8</sup> Gli assi sono in scala logaritmica; le quantità monetarie sono espresse in euro 2000.

istituzionale del settore. La classificazione muta nel tempo, così come la consistenza dei diversi gruppi di intermediari, con il procedere dei processi di liberalizzazione del settore. Si è scelto di mantenere – pur semplificandole in parte – quelle classificazioni, e di renderle visibili anche nei grafici, poiché esse forniscono informazioni rilevanti sull'evoluzione e trasformazione del settore, nonché – indirettamente – indicazioni circa le dimensioni, i modelli di proprietà e controllo societario prevalenti e l'operatività degli intermediari che ne fanno parte. I colori utilizzati negli *scatterplots* rappresentano l'appartenenza delle singole banche all'una o all'altra di esse. In arancione sono indicate le casse rurali e artigiane, poi banche di credito cooperativo; in giallo le banche popolari; in rosso le casse di risparmio; in verde le banche commerciali<sup>9</sup>; infine in blu gli istituti di credito speciale, poi banche a medio e lungo termine. I punti cerchiati in nero indicano i gruppi bancari (Grafico 4.6). Nei grafici relativi all'analisi di efficienza sono stati inoltre evidenziati alcuni tra i maggiori intermediari dei due sottoperiodi, pre e post privatizzazioni e fusioni<sup>10</sup>.

Nei grafici di destra, relativi alla proporzione di input impiegata dalle banche nel processo di produzione, si può notare come le banche si distribuiscano quasi esattamente lungo una retta nel corso dell'intero periodo, suggerendo il fatto che l'industria bancaria impieghi una funzione di produzione a coefficienti fissi (*à la* Leontief). Dall'inizio del decennio 1990, la curva si sposta verso l'alto e diviene più ripida, quindi nella seconda metà del decennio la “nuvola” si sfarina, con uno spostamento verso l'angolo superiore destro. Ciò sembra essere coerente, dapprima con un effetto traino della crescita delle grandi banche sull'intera industria, quindi più direttamente con la rapida concentrazione del sistema a fine anni 1990. L'ulteriore spostamento della curva verso destra negli ultimi anni considerati appare correlato sia al crescente numero di fusioni ed acquisizioni (Banca d'Italia, 1998), sia agli effetti dell'innovazione tecnologica.

Nei grafici di sinistra, come s'è detto, sono evidenziati i risultati della DEA. L'osservazione dei grafici mostra come le banche tendano a distribuirsi – via via in modo più evidente nel corso del periodo – lungo una curva a forma di U. Tra 1979 e 1985, le banche minori, in larga misura casse rurali

<sup>9</sup> Istituti di diritto pubblico, banche di interesse nazionale, società ordinarie di credito.

<sup>10</sup> Legenda dei maggiori intermediari evidenziati nei grafici: BCI, Banca commerciale italiana; BI, Banca Intesa; BNL, Banca nazionale del lavoro; BPER, Banca popolare dell'Emilia Romagna; BPB, Banca popolare di Bergamo; BPM, Banca popolare di Milano; CI, Credito italiano; MB, Mediobanca; MPS, Monte dei paschi di Siena; SP, Istituto San Paolo; UC, Unicredito italiano.

e banche cooperative, ma anche qualche banca ordinaria, emergono come le più efficienti dell'intero sistema, in particolare delle grandi banche. Gran parte delle banche considerate tuttavia si attesta su livelli di efficienza tecnica al di sotto di 0,5. Nella seconda metà del decennio, il campione tende a distribuirsi lungo una V: solo poche banche minori e alcune delle maggiori si avvicinano alla frontiera, mentre gran parte del campione precipita su livelli molto bassi di efficienza.

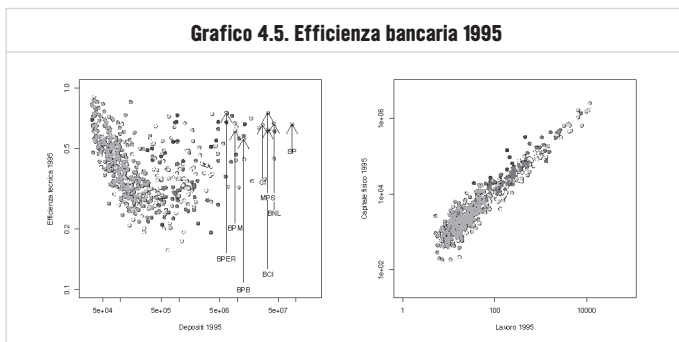
Nella prima metà degli anni 1990, i sommovimenti e le trasformazioni in atto sono verosimilmente alla base di un recupero di efficienza nei livelli inferiori e dell'ispessimento della fascia dimensionale media: crescita dimensionale media degli intermediari e "rimescolamento" del sistema appaiono abbastanza chiaramente. Nella seconda metà del decennio, il processo di concentrazione del sistema emerge con chiarezza. Si riduce visibilmente il numero di intermediari e si accentua la polarizzazione dimensionale. I maggiori gruppi si attestano su livelli di efficienza abbastanza elevati, sebbene inferiori rispetto a un piccolo numero di banche minori (banche di credito cooperativo).

Nel complesso sembrano delinearsi tre *patterns* distinti. Il primo relativo alle banche di credito cooperativo, per le quali è confermata lungo tutto il periodo una relazione inversa tra efficienza e dimensione d'impresa. Il secondo, relativo ai grandi gruppi, i quali sembrano – dopo numerosi cambiamenti dovuti al processo di concentrazione – attestarsi su livelli di efficienza non dissimili rispetto all'inizio del periodo; l'effetto maggiore della riorganizzazione del sistema, da questo punto di vista, sembra essere un aumento della dimensione media dei grandi intermediari e di alcuni di quelli medi e medio-grandi. Il terzo riguarda le banche di medie dimensioni (in gran parte popolari e casse di risparmio): i livelli di efficienza di queste, mediamente bassi nel corso degli anni 1980, e con una tendenza alla ripresa nella prima metà del decennio successivo, tendono a peggiorare dalla seconda metà degli anni 1990 a fine periodo; contestualmente, per effetto di fusioni e acquisizioni, questo gruppo si assottiglia notevolmente rispetto agli altri due. Da esso emergono alcune banche medio-grandi a fine periodo, che tuttavia non sembrano essere state in grado di riorganizzarsi in seguito ai processi di crescita esterna sperimentati, con effetti negativi sulla loro efficienza. Un discorso a parte meriterebbe il caso delle banche a lungo termine/banche di investimento. Esse infatti hanno caratteristiche proprie e in particolare adottano un mix di input diverso dalle altre banche: in presenza di proporzioni quasi fisse tra gli input, questo spiega come esse appaiano in media meno efficienti delle altre.

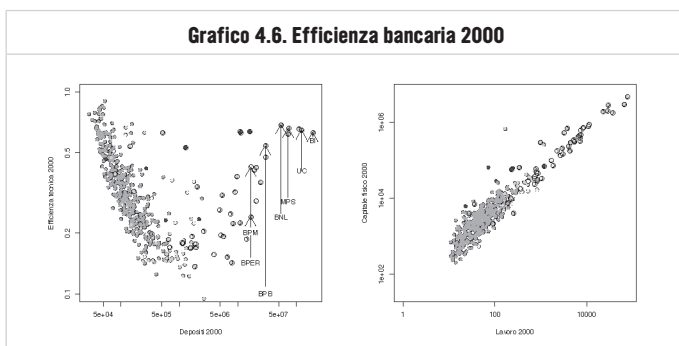




**Grafico 4.5. Efficienza bancaria 1995**



**Grafico 4.6. Efficienza bancaria 2000**



## 5. Conclusioni e agenda

Questo lavoro è un primo tentativo di studiare nel lungo periodo gli effetti sull'efficienza delle banche, e dell'industria bancaria nel suo complesso, dei processi di liberalizzazione, privatizzazione e concentrazione che hanno interessato il sistema italiano, ridisegnandone l'architettura istituzionale e modificandone la struttura nell'ultimo ventennio del secolo scorso.

L'analisi ha evidenziato il prodursi di una forte polarizzazione dimensionale, con una crescita della dimensione media degli intermediari maggiori accompagnata dalla quasi sparizione delle banche di dimensione media e medio piccola. Secondo quanto emerso dall'analisi, le banche minori si confermerebbero le più efficienti lungo tutto il periodo, nonostante una certa capacità di ripresa di almeno alcuni dei grandi gruppi nati dal processo di consolidamento del sistema. Molto negativi appaiono invece i risultati relativi alle banche medie, sia perché risultano sempre posizionate molto lontane dalla frontiera produttiva, sia perché nel com-

plesso i livelli di efficienza tendono a ridursi nel corso del periodo. Ciò è probabilmente dovuto alla difficoltà di implementare riorganizzazioni e ristrutturazioni, rese necessarie dalle mutate condizioni del mercato del credito dalla fine degli anni 1990, e con esse all'incapacità di queste banche di raggiungere dimensioni di scala sufficienti. Le banche minori sembrano essere invece state in grado di inserirsi e/o consolidare la propria posizione su mercati di nicchia, sfruttare appieno vantaggi quali il radicamento nel proprio territorio e la conoscenza delle economie locali.

Rimane tuttavia ancora molto da fare. Nella sua forma attuale, il dataset contiene anche informazioni relative ai prezzi degli input. Questo consentirebbe di calcolare altre misure di efficienza, le cosiddette efficienze di costo ed allocative, che permetterebbero di completare l'analisi presentata in precedenza, mostrando non solo quanto le banche italiane sappiano utilizzare efficientemente gli input ma anche quanto le combinazioni di input che usano consentano di minimizzare il costo di produzione. Inoltre anche la struttura dei rendimenti di scala del sistema bancario italiano merita di essere studiata in maggior dettaglio, per esempio calcolando le efficienze di scala e testando formalmente l'ipotesi che i rendimenti di scala siano o meno costanti. Occorre però sottolineare che al momento attuale le efficienze summenzionate possono essere calcolate nell'ambito della DEA, ma gli autori non sono a conoscenza di nessun algoritmo per la correzione delle loro distorsioni sistematiche. Pertanto, all'analisi empirica si dovrà affiancare lo studio teorico delle proprietà di queste misure di efficienza.

Altre analisi richiederanno invece un'estensione del dataset attuale. Da un lato è necessario migliorare il dataset, in particolare introducendo chiavi primarie che permettano di seguire ciascuna banca nel tempo; dall'altro, aggiornare il dataset stesso – recuperando i dati mancanti nelle edizioni più recenti dell'annuario Mediobanca – almeno al 2008. Si prevede inoltre di raccogliere anche le informazioni extra contabili contenute (sotto forma di note) nella fonte, in modo da poter ricostruire la storia di ciascuna impresa, in particolare relativamente ai processi di fusione e acquisizione. La ricostruzione di questi avvenimenti consentirebbe di studiare dal punto di vista dinamico le variazioni di efficienza nell'intorno degli eventi stessi.

## **Riferimenti bibliografici**

- Angelini P., Cetorelli N. (2003), *The Effects of Regulatory Reform on Competition in the Banking Industry*. *Journal of Money, Credit, and Banking*, 35, 663–684.  
Banca d'Italia (1981), *Ordinamento degli enti pubblici creditizi*.  
Banca d'Italia (1998), *Relazione Annuale*.

- Banca d'Italia (2001), *Relazione Annuale*.
- Barca F., Trento S. (1997), *La parabola delle partecipazioni statali; una missione tradita*. In *Storia del capitalismo italiano: dal dopoguerra a oggi* a cura di F. Barca. Donzelli, 185–236.
- Bellavite Pellegrini C. (2002), *Storia del Banco Ambrosiano: fondazione, ascesa e dissesto: 1896-1982*. Laterza.
- Bemporad S., Reviglio E. (2001), *Le privatizzazioni in Italia: 1992-2000*. Edindustria.
- Berger A.N., Humphrey D.B. (1992), *Measurement and efficiency issues in commercial banking*. In *Output Measurement in the Service Sectors* a cura di Z. Griliches. University of Chicago Press, 245–300.
- Bogetoft P., Otto, L. (2011), *Benchmarking with DEA, SFA, and R*. Springer.
- Bonaccorsi di Patti E., Gobbi G. (2001), *The Effects of Bank Consolidation and Market Entry on Small Business Lending*. Banca d'Italia, *Temi di discussione*, n. 404.
- Brambilla C., Conti G. (2011), *Ownership Structure and Control, Regulation and Performance in Italian Banking. A Long-term Perspective*. In *Corporate Governance in Financial Institutions. Historical Developments and Current Problems* a cura di I. Elferink, J. Mooji, A. Perlinge. EABH, 128–145.
- Brambilla C., Conti G. (2016), *New bottles, old wine? Concentrations and performances in Italian banking, 1970s-2000s*. In *Immortal Banks. Strategies, Structures and Performances of major Banks* a cura di M. Lescure. Droz, 111–137.
- Charnes A., Cooper WW, Rhodes E. (1978), *Measuring the efficiency of decision units*. *European Journal of Operational Research*, 2, 429–444.
- Ciocca P. (2005), *The Italian financial system remodelled*. Palgrave Macmillan,
- Colwell R.J., Davis E.P. (1992), *Output and productivity in banking*. *The Scandinavian Journal of Economics*, 94, S111–S129.
- Daraio C., Simar L. (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. Springer.
- Debreu G. (1951), *The coefficient of resource utilization*. *Econometrica*, 19, 273–292.
- Dermine, J. (2003), *The New World of Euro Banking*. In *Handbook of International Banking* a cura di A.W. Mullineux, V. Murinde. Edward Elgar, 105–129.
- Farrell M.J. (1957), *The measurement of productive efficiency*. *Journal of the Royal Statistical Society. Series A (General)*, 120, 253–281.
- Favero C.A., Papi, L. (1995), *Technical efficiency and scale efficiency in the Italian banking sector: a non-parametric approach*. *Applied Economics*, 27, 385–395.
- Fiordelisi F. (2009), *Mergers and Acquisitions in European Banking*. Palgrave Macmillan.
- Focarelli D., Panetta F., Salleo C. (2002), *Why do Banks Merge?* *Journal of Money, Credit, and Banking*, 34, 1047–1066.
- Focarelli D., Panetta F. (2003), *Are Mergers Beneficial to Consumers? Evidence from the Market for Bank Deposits*. *American Economic Review*, 93, 1152–1172.
- Galanti, E., D'Ambrosio R., Guccione A.V. (2012), *Storia della legislazione bancaria finanziaria e assicurativa. Dall'Unità d'Italia al 2011*. Marsilio.
- Giordano F. (2007), *Storia del sistema bancario italiano*. Donzelli.
- Inzerillo U., Messori M. (2000), *Le privatizzazioni bancarie in Italia*. In *Le privatizzazioni italiane* a cura di S. de Nardis. Il Mulino, 119–190.
- Messori M. (2009), *Consolidation, Ownership Structure and Efficiency in the Italian Banking System*. In *The Banks and the Italian Economy* a cura di D.B. Silipo. Springer Physica-Verlag, 211–243.

- Messori M., Tamborini R., Zazzaro A. (2003), *Il sistema bancario italiano: le occasioni degli anni novanta e le sfide dell'euro*. Carocci.
- Resti A. (1997), *Evaluating the cost-efficiency of the Italian banking system: What can be learned from the joint application of parametric and non-parametric techniques*. *Journal of Banking and Finance*, 21, 221-250.
- Sapienza P. (2002), *The Effects of Banking Mergers on Loan Contracts*. *Journal of Finance*, 57, 329-368.
- Simar L., Wilson P.W. (1998), *Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models*. *Management Science*, 44, 49-61.
- Spero J.E. (1980), *The failure of the Franklin National Bank: challenge to the international banking system*. Columbia University Press.

# Società e gruppi nei settori tessile e abbigliamento dal 1979 al 2009

GIUSEPPE CONTI E VALERIA PINCHERA

## 1. Introduzione

Negli ultimi trent'anni l'economia italiana ha dovuto adattarsi e reagire a cambiamenti profondi intervenuti a livello internazionale e riconducibili, soprattutto, a tre aspetti particolarmente rilevanti per la struttura produttiva, per la concorrenza interna ai settori, per i profitti delle singole imprese e per le scelte di investimento. Il primo aspetto riguarda un crescente grado di apertura e di integrazione dei mercati dei beni, dei capitali e del lavoro; ad esso si aggiunge l'avvento di nuove tecnologie e, infine, il progressivo ridimensionamento delle capacità di indirizzo e di protezione da parte delle politiche industriali nazionali. Non è questo il luogo per trattare compiutamente l'impatto di fenomeni così complessi sulla struttura industriale e sull'economia italiana nel suo complesso. La letteratura al riguardo ha generalmente sottolineato due conseguenze, di per sé contraddittorie e problematiche, la prima relativa a un declino economico complessivo registrato a livello di alcune delle principali grandezze macroeconomiche e macrosettoriali e, per contro, elementi di vitalità imprenditoriale riscontrabili in alcuni settori industriali legati principalmente a quello che viene detto il "made in Italy", i distretti industriali a produzione flessibile, anche se tali etichette non sono sempre riferibili agli stessi fenomeni, con caratteristiche omogenee, né sono tra loro intercambiabili. Nel presente lavoro, l'analisi è focalizzata su due comparti dell'industria "leggera", legati tra loro da affinità merceologiche e integrazioni produttive come quelli del tessile e dell'abbigliamento, e all'interno di essi su alcune delle principali società al fine di fornire primi elementi di riflessione sulle ristrutturazioni avvenute nel trentennio, sulle dinamiche interne e relative alle prestazioni

di settore e di alcune singole imprese o gruppi societari. Settori e mercati di riferimento sono in questi anni sempre meno identificabili in maniera precisa per effetto dei cambiamenti nelle tecnologie e nei metodi di produzione, ma anche per fenomeni di diversificazione per motivi strategici e produttivi o finanziari. Le interconnessioni finanziarie tra imprese anche di dimensioni medie si è accompagnata alla diversificazione produttiva. La crescita dei gruppi societari non è recente, ma dagli anni '90 trova riferimenti normativi più precisi e una diffusione su più ampia scala per tenere sotto controllo strategie di *outsourcing* e di gestione di servizi o produzioni intermedie o collaterali. I gruppi hanno avuto una parte importante nelle ristrutturazioni produttive, sulle forme concorrenziali, sulla diversificazione di prodotto e sulla finanza delle società controllate. Le conglomerate rendono meno evidenti i confini dei mercati di riferimento e accrescono la mobilità orizzontale degli investimenti da un settore all'altro. Anche in termini di distribuzione di profitti e perdite, i gruppi hanno svolto un ruolo rilevante consentendo passaggi di fondi da una società all'altra, come pure migrazioni interne di addetti e di interi comparti produttivi. Le singole imprese, specialmente quelle di dimensioni medio-grandi, sono così sempre meno identificabili sulla base di una o più semplici linee produttive e singoli marchi di fabbrica.

Gli studi sul declino riguardano un ampio ventaglio di fattori di fragilità strutturali e dinamiche dell'economia italiana rispetto a quella di altri paesi europei e di economie emergenti (Toniolo e Visco, 2004; Nardozi, 2004; Berta, 2004)<sup>1</sup>.

Il declino rimane, comunque, un tema controverso e generalmente individuato a livello di sistema. Le analisi settoriali sottolineano in vari casi luci e ombre con dinamiche molto differenziate riconducibili alla crisi dell'industria fordista e al rafforzamento, specialmente fino a tutti gli anni '90, di realtà produttive più flessibili localizzate in distretti industriali o al centro di reti di imprese. Tuttavia, di recente anche alcuni tessuti distrettuali di successo hanno mostrato segni di cedimento. Varie economie locali hanno imboccato una parabola discendente, ma incerta a livello aziendale e produttivo per effetto di ricalizzazioni delle attività, abbandono di processi e ristrutturazioni complessive per fronteggiare una concorrenza globale dagli esiti incerti. Realtà così differenziate pongono problemi di composizione dai livelli micro di singola impresa produttiva a quelli macro di settore. In particolare, – secondo alcune tesi – la

<sup>1</sup> Alla luce dei problemi recenti gli storici economici hanno rivisto il senso delle «sorti progressive» che andavano “dalla periferia al centro” per avanzare dubbi e timori di una più recente inversione di tendenza e di riapertura di divergenze che sembravano colmate.

debole crescita dell'economia non sarebbe da imputare a un declino né di quote di mercato né di dinamica dei comparti manifatturieri. Le industrie manifatturiere avrebbero, invece, rafforzato la propria competitività internazionale nonostante le condizioni avverse dovute a una domanda interna depressa dal peso eccessivo del debito pubblico e dalle conseguenti politiche di austerità e di elevata pressione fiscale senza riscontro con i servizi resi da una burocrazia farraginosa e inefficiente (Fortis, 2016). Le capacità innovative di settori come la meccanica, "made in Italy", e altri sono attribuite al mantenimento di standard qualitativi indipendentemente dalle dimensioni aziendali.

In questa sede l'obiettivo è di tentare una prima verifica delle tesi del declino o della capacità di adattamento e di crescita da parte di una fascia "alta" di società in due comparti tradizionali dell'industria "leggera" con possibili legami di integrazioni produttive. Le società considerate non sono dunque imprese tipo, né possono essere considerate, se non in qualche misura, rappresentative in settori tradizionalmente densi di piccole imprese. Sono però imprese influenti, o anche molto influenti, per indotto, leader nelle subforniture, per *outsourcing*, imitazione di prodotto, di schemi manageriali e altro ancora. La scelta è stata dettata - indipendentemente dalla dimensione (es. le "prime 4 o 8") - dalla maggiore permanenza nel drappello delle "principali". Nella prossima sezione sono esaminate succintamente alcuni dei risultati analitici offerti in letteratura. Non sempre sono confrontabili con quelli qui presentati perché si riferiscono a periodi diversi e, soprattutto, fanno generalmente riferimento ad alcune variabili, ma non - come nel nostro caso - a fonti contabili più complete e organizzate. Le peculiarità del campione di società "grandi" - 11, di cui 2 organizzate in forma di gruppo - sono esaminate nel § 3 con un inquadramento nel contesto storico-evolutivo di riferimento. Nella parte successiva vengono analizzati, sulla base di dati di bilancio, i principali risultati e andamenti societari in confronto con quelli di settore e dell'insieme delle società non finanziarie. Le conclusioni sono ancora esplorative, ma utili a circoscrivere meglio alcuni problemi e sviluppi di ricerca.

## **2. Il puzzle delle industrie "leggere"**

Ai distretti italiani si attribuisce il ruolo di motori trainanti delle esportazioni e di pilastri della crescita dell'economia italiana (Fortis, 2016). L'apertura sull'estero e la specializzazione caratterizzano specialmente i settori del cosiddetto "made in Italy". L'inserimento nel commercio internazionale avviene su produzioni su larga scala (tessuti, mobilio, e altro) o su prodotti di nicchia

(bottoni, apparecchi per la vinificazione ecc.). La struttura produttiva dispersa in piccole unità assicura una concorrenza interna vivace tra imprese, un'integrazione continua tra strategie rivali e di collaborazione, consentendo un accumulo di conoscenze che ne favorisce radicamento e sviluppo (Bottazzi et al., 2002) attraverso un ricambio frequente nelle posizioni di punta. Nel lungo periodo più che la performance della singola impresa conta quella della "specie" distrettuale grazie alla flessibilità del sistema nel suo insieme (Fortis, 2016, 163 e 165). I distretti non sarebbero settori tradizionali con scarsa capacità di adattamento e forti resistenze alla specializzazione ma «innovatively traditional», secondo l'ossimoro di Fortis (2016, 85; e Rullani, 2002, 40).

Il comparto della moda, interno alla filiera tessile e abbigliamento, sarebbe una delle "4F" dell'eccellenza produttiva nazionale<sup>2</sup>. I prodotti dell'industria tessile e abbigliamento italiana hanno occupato tradizionalmente il primo posto nell'esportazione su scala mondiale con 5,7 miliardi di dollari nel 1979, all'incirca uguale alla quota di tutte le economie in via di sviluppo (Underhill, 1998, 42). Alcuni prodotti specifici conseguono posizioni alte nella graduatoria del contributo alle esportazioni (per il 1996 v. la tab. in Bottazzi et al., 2002, 187-8).

Le imprese, o società, qui considerate non sono propriamente rappresentative di imprese di distretti o tipiche di sistemi locali, anche se molte hanno unità produttive insediate al centro o in prossimità di una, o più di una, realtà distrettuale. Già questo disegno a geografie multiple le esclude dal rango. Si tratta per lo più di società leader di dimensioni medio-grandi, spesso facenti parte di un medesimo gruppo sotto il controllo di una holding. Gli stessi gruppi sono costituiti secondo logiche differenti e non sempre in funzione di integrazioni produttive, orizzontali o verticali, con diversificazioni nella natura degli investimenti. Le società prese in esame stanno in una posizione mediana tra le residue grandi imprese rappresentanti dirette del declino industriale del paese e le piccole imprese con le quali condividono solo l'appartenenza al settore e, talora, legami di committenza. Il concetto di "chain value" introdotto da Porter (1985, 33) per disaggregare l'impresa in attività strategicamente rilevanti per il controllo dei costi e dei risultati è diventato uno schema sempre più adeguato a descrivere la frammentazione di processi produttivi esterni alle imprese che li comandano, come parte o meno di un gruppo, ma riguardanti fasi tra loro legate dalla produzione e commercializzazione del prodotto finale. La globalizzazione ha contribuito, anche nei settori dell'industria "leggera", all'affermazione e alla

<sup>2</sup> Le "4F" sono individuate in «Fashion and cosmetics; Food and wine; Furniture and ceramic tiles; Fabricated metal products, machinery and transport equipment».



diffusione di imprese conglomerate secondo catene di valore di investimento e commercio internazionale<sup>3</sup>. Ciò come vedremo rende problematiche le stesse divisioni settoriali specialmente per le differenziazioni di performance tra le imprese la cui attribuzione al settore di riferimento resta in base alla tradizione storica e al carattere prevalente della produzione manifatturiera.

Le società prese in considerazione hanno produzioni nei settori tessile e abbigliamento distinte per la qualità dei prodotti e per le quote di mercato, ma sono anche rappresentative di un maggior grado di mobilità nella dislocazione delle unità di produzione e nella capacità di diversificazione<sup>4</sup>. I due settori, per quanto, contigui e integrati, vanno distinti sotto più di un aspetto. Il settore tessile appartiene a una tecnologia “matura” con una filiera quasi lineare, scomponibile in fasi intermedie. Negli ultimi trentanni imprese del settore dell'abbigliamento, o tessili integrate nelle confezioni, per diverse esigenze di domanda e di cambiamento tecnologico hanno orientato le produzioni su linee diverse dalla produzione di massa<sup>5</sup>. L'integrazione verticale riscontrata nel settore tessile è dipesa in genere da esigenze di controllo sulla qualità del prodotto e di riduzione dei costi attraverso la razionalizzazione dei flussi di semilavorati (Morrone, 1992, 108-9).

### **3. Settori del tessile e abbigliamento in Italia: struttura e andamento**

Le industrie del tessile e dell'abbigliamento seppur considerate un settore maturo nelle economie avanzate occidentali, continuano a rivestire un ruolo nel sistema economico italiano sia in termini di imprese, che di addetti, di contributo al PIL e all'export. Nel 2013 i settori del tessile e abbigliamento rappresentavano oltre l'11% delle imprese, circa il 9% degli addetti e il 6,2% del

<sup>3</sup> Porter (1998, 250-6) parla anche di «segmentation matrices» per designare questo genere di strutture di imprese che per finalità strategiche si posizionano in determinati segmenti di attività. Su questi aspetti cfr. anche Marsh (2012, 17 e 243) per la dispersione su scala globale di attività manifatturiere da parte di imprese dei paesi avanzati, anche da parte di “micro-multinazionali”.

<sup>4</sup> Cfr. il caso delle imprese venete di calzature e abbigliamento esaminato da Gianelle e Tattara (2009).

<sup>5</sup> Per i cambiamenti tecnologici tipo “demand-driven” che hanno interessato i settori in epoca di globalizzazione v. Walter - Kartsounis - Carosio (eds. 2009). Un utile quadro d'insieme in Shishoo (ed. 2012).

valore aggiunto del totale dell'industria manifatturiera italiana, configurando l'Italia come il principale produttore ed esportatore europeo (Istat, 2016). L'evoluzione delle industrie del tessile e dell'abbigliamento nel corso del Secondo Dopoguerra ripercorre il sentiero di sviluppo del sistema economico italiano. Quali settori appartenenti alla categoria di attività produttive cosiddette leggere, ad elevata intensità di lavoro e a bassa tecnologia, hanno contrassegnato l'avvio del processo di industrializzazione di un paese, per poi ridimensionare il loro peso, in parallelo allo sviluppo dell'industria soprattutto verso i settori a più elevata intensità di capitale e di ricerca e sviluppo (Pent Fornengo, 1979; Giannetti Vasta, 2003 e 2005; Nordàs, 2004). I dati censuari confermano questa dinamica soprattutto per il tessile, che raggiunta la sua fase matura negli anni Cinquanta ha iniziato a declinare progressivamente dagli anni Sessanta, come evidenziato nella tabella 1 (Merlo, 2008; Pinchera, 2009). Un arretramento che inizialmente fu in parte compensato dall'ascesa del settore dell'abbigliamento, che si affermò come comparto industriale tra gli anni Sessanta e Settanta, attraverso la diffusione della produzione della confezione in serie e l'avvio del processo di formazione e articolazione della filiera della moda (Merlo, 2008, 91-115; Paris, 2006; Pinchera, 2009, 139-155).

	1951	1961	1971	1981	1991
Tessile	653.107	612.886	539.553	522.915	403.929
Abbigliamento	223.414	289.860	393.280	403.360	418.862
Totale	876.521	902.746	932.833	926.275	822.791
%	25,2	20,1	17,5	15,7	15,8

**Fonte:** nostre elaborazioni da Federico, 2003, tab. 2.1, 46.

La dinamica occupazionale dei settori leggeri fu, dunque, contraddistinta, da un lato dal processo di maturazione del sistema moda come settore industriale, attraverso lo spostamento del suo epicentro produttivo dal comparto tessile a quello dell'abbigliamento e dall'altro dal progressivo avanzare delle industrie pesanti, come la meccanica, la gomma e la siderurgia, che contribuirono in modo sostanziale alla crescita complessiva degli addetti del manifatturiero italiano e di conseguenza alla riduzione del peso dei comparti leggeri. Gli anni Settanta segnarono, quindi, un significativo cambiamento strutturale dei settori del tessile e dell'abbigliamento e più in generale del sistema industriale italiano, che favorì la diffusione delle piccole e medie imprese (Saviolo - Testa, 2000, 36-68; Testa, 2003, 699-734). Si trattò di un processo di radicale ristrutturazione.

turazione e decentramento produttivo determinato da un insieme composito di ragioni: recessione economica, evoluzione del mercato, alti costi del lavoro, nonchè caratteri strutturali del settore, come la scarsa rilevanza delle economie di scala. Tale processo non fu però generalizzato, ma interessò principalmente le imprese di grandi dimensioni, come il Gruppo Finanziario Tessile, Monti e Marzotto, ed in particolare quelle a partecipazione statale come la Lanerossi di Schio (Pent Fornengo, 1979, 72; Pinchera, 2009, 151-152). Tra il 1971 e il 1991 la quota di addetti degli esercizi dei settori del tessile e dell'abbigliamento con oltre 500 addetti si ridusse considerevolmente, passando rispettivamente da oltre il 16% a meno del 3% e dal 6,5% a meno del 2%, mentre parallelamente crebbe il peso degli addetti dei piccoli esercizi (cfr. tabella 2).

<b>Tessile</b>	Meno di 10 addetti	10-50 addetti	51-100 addetti	101-500 addetti	Oltre 500 addetti
1971	16,7	21,8	12,6	32,4	16,5
1981	24,8	29,2	12,4	26,6	7,0
1991	25,8	36,0	12,9	22,2	3,2
<b>Abbigliamento</b>					
1971	32,0	21,3	10,4	25,2	11,1
1981	27,4	36,1	11,2	18,6	6,7
1991	29,7	46,2	8,6	12,8	2,7

**Fonte:** nostre elaborazioni da Federico, 2003, 74-77.

Il progressivo processo di trasformazione del sistema industriale del tessile e abbigliamento, e più in generale del Made in Italy, in un insieme di piccole e medie imprese concentrate in poli locali ad alta specializzazione produttiva divenne l'emblema dell'affermazione a livello internazionale della moda italiana negli anni Ottanta, e più specificatamente del nuovo fenomeno del prêt-à-porter italiano, una produzione innovativa e qualificata di fascia più alta, frutto del nuovo "collegamento funzionale" tra industria e stilismo, tra produzione e creazione (Colli - Merlo, 2007; Merlo, 2011, 348-358). Tale successo comportò rilevanti cambiamenti nel sistema industriale della moda, stimolando il progressivo trasferimento dell'offerta da una produzione standardizzata e di massa alla fabbricazione di articoli più specializzati e ricercati, attraverso i quali valorizzare i caratteri qualitativi materiali, e soprattutto immateriali dei prodotti, come il design, lo stile e il valore simbolico della marca (Saviolo - Testa, 2000, 151-164; Gereffi, 1999 e Truett - Truett, 2014). Un processo di trasformazione che implicò importanti modificazioni nell'organizzazione e struttura delle imprese italiane della filiera

della moda, in relazione al crescente rilievo assunto dalle fasi finali del processo di produzione – dalle strategie di marca, alla comunicazione e, quindi, alla distribuzione – destinato ad accentuarsi nei decenni successivi.

Negli anni Novanta i mutamenti nell'economia internazionale sia lato dell'offerta, con il sopravanzare della produzione più a basso costo dei paesi di più recente industrializzazione, che della domanda, in relazione alla graduale contrazione dei consumi di abbigliamento nelle economie occidentali, e ad una crescente segmentazione del mercato, hanno portato a significativi cambiamenti nella geografia e struttura produttiva sia a livello nazionale che internazionale (cfr. tabella 3).

<b>Tessile</b>	<b>1996</b>	<b>2001</b>	<b>2006</b>	<b>2009</b>	<b>var. % 1996-2009</b>
<b>Addetti</b>	347.282	311.010	196.597	162.527	-53
<b>Imprese</b>	33.423	29.564	19.613	17.243	-48
<b>Abbigliamento</b>					
<b>Addetti</b>	348.079	298.619	273.947	244.354	-30
<b>Imprese</b>	51.923	43.780	39.306	34.695	-33
<b>Totale Fatturato T/A (mil €)*</b>	56.777	56.152	56.456	46.312	-18

\* Rivalutato al 2009.

**Fonti:** per addetti e imprese Istat, Censimento 1996 e 2001 e Istat, Rapporto sulla competitività 2013; per fatturato dati SMI e Eurostat.

In Francia i settori del tessile e abbigliamento tra il 1979 e il 2007 hanno registrato una caduta degli occupati di circa l'80%, che nel comparto dell'abbigliamento ha superato l'84%, più del doppio di quella italiana, riducendo il numero di addetti complessivi del T/A francese da circa 609.000 a poco più di 123.000<sup>6</sup>. Tale tendenza al forte ridimensionamento dell'occupazione dei settori leggeri non ha riguardato esclusivamente la Francia e, seppur in minor misura l'Italia, ma ha assunto una dimensione europea. Le filiere del tessile-abbigliamento dei principali paesi europei, dalla Francia, alla Germania, ormai giunte ad una fase di maturità, hanno avviato un progressivo processo di deindustrializzazione, che ha ulteriormente avvantaggiato l'avanzare dell'area asiatica. In primis in Francia, e in altre realtà europee si è realizzato un nuovo assetto strategico organizzativo e produttivo del sistema moda, che ha ridisegnato i caratteri delle macrofiliera, attraverso un processo di concentrazione che ha contribuito alla formazione di due principali holding della moda e del lusso: Kering e LVMH,

<sup>6</sup> Calcoli su dati OECD. Stat, sull'evoluzione del settore della moda in Francia si veda: Barrère – Santagata, 2005, 18-19.

la cui affermazione su scala mondiale è corsa parallelamente alla campagna di acquisizioni di aziende e marchi della moda soprattutto italiani, come Gucci, Emilio Pucci, Bottega Veneta e Loro Piana (Pinchera, 2009, 285-301).

Il processo di globalizzazione e di progressiva ridefinizione dei tradizionali mercati di sbocco della produzione italiana del tessile e abbigliamento ha sollecitato le principali imprese della moda a rivedere i loro assetti organizzativi e strategie produttive e distributive (Corbellini-Saviolo, 2009; Saviolo-Testa, 2000; Taplin, 2006; Taplin-Whiterton, 2004). Da un lato è avvenuto un generale processo di *upgrade* della produzione verso prodotti a maggior valore aggiunto, associato ad una selezione dei mercati di sbocco e ad una riduzione delle quantità prodotte, dall'altro si è proceduto ad incrementare gli investimenti su alcune fasi a valle del processo produttivo, delocalizzando ed esternalizzando le fasi manifatturiere in senso stretto (Ricchetti - Cietta, 2006, XV-XX). Un'evoluzione che ha ridimensionato i vantaggi competitivi del tradizionale paradigma organizzativo della moda italiana delle piccole e medie imprese e che soprattutto ha fortemente ridotto le quote di mercato a livello mondiale dei settori del tessile e dell'abbigliamento italiani, rispettivamente del 35% e 38% tra l'inizio degli anni Novanta e il 2009 (Pinchera, 2009, 251-255 e WTO).

Nel nuovo scenario competitivo la dimensione di impresa ha assunto un rilievo cruciale sia per preservare le aree di mercato che per tutelarsi da possibili acquisizioni. Ciononostante il settore tessile-abbigliamento italiano a differenza delle altre realtà europee risulta ancora sostanzialmente dominato dal sistema delle piccole e medie imprese sia in termini di addetti che di numero di imprese, seppure nel comparto dell'abbigliamento sia evidente una crescita del peso delle imprese medio-grandi (cfr. tabella 4).

<b>Tabella 4. Distribuzione di addetti per classe dimensionale di impresa, 2001-2011</b>					
<b>Tessile</b>	1-9 addetti	10-49 addetti	50-99 addetti	100-249 addetti	+ di 250 addetti
2001	20,7	34,1	13,5	13,9	17,8
2011	24,2	34,8	14,0	12,6	14,4
<b>Abbigliamento</b>					
2001	30,2	43,4	9,2	7,7	9,5
2011	32,8	37,4	7,0	6,8	16,0

**Fonte:** elaborazioni su dati Istat, Censimento industria 2001 e 2011.

La differente struttura del tessile-abbigliamento italiano è evidenziata dallo scarso peso delle imprese di grandi dimensioni, pur presenti in Italia, e dal basso livello di concentrazione del sistema produttivo rispetto ad altri paesi europei,

come denota il limitato ruolo rivestito dalle principali imprese e gruppi italiani sul fatturato totale dei comparti a livello nazionale ed europeo (cfr. tabella 5 e 6).

Paese	Tessile	Abbigliamento
Francia	28	35
Germania	20	46
Gran Bretagna	52	33
Italia	12	25

Fonte: Dati da Stengg, 2001, 10.

Gruppo	Settore	Fatturato in mil. €	Quota % fatturato nazionale	Posizione EU 15
Gruppo Marzotto SpA	Tessile	1,722	5	1
Gruppo Tessile Miroglio	Tessile	842	2	9
Ermenegildo Zegna Holdit	Tessile	686	2	11
Benetton gruppo	Abbigl.	2,098	5	4
Giorgio Armani SpA	Abbigl.	1,272	3	6
Max Mara Fashions	Abbigl.	1,087	3	9
Diesel SpA	Abbigl.	565	1	20

Fonte: nostre elaborazioni da Dunford, 2006, 45.

## 4. Le società leaders e followers nei dati Mediobanca

### 4.1. Settori labour-intensive e investimenti e disinvestimenti societari

L'analisi dei dati Mediobanca sulle principali imprese dal 1979 al 2009 conferma l'evoluzione strutturale dell'industria italiana del tessile-abbigliamento, rimarcando da un lato il progressivo declino del comparto tessile e l'ascesa dell'abbigliamento e dall'altro il complessivo ridimensionamento dei settori leggeri rispetto al sistema manifatturiero ed economico del paese (cfr. tabella 7).

	1979	1989	1999	2009
Tessile	5	1	2	1
Abbigliamento	-	3	6	8

Fonte: nostre elaborazioni dati Mediobanca.

La rappresentatività delle imprese del tessile e abbigliamento presenti nel database Mediobanca ha assunto un rilievo crescente tra il 1979 e il 2009 sia in termini di addetti che di fatturato (cfr. tabella 8).

**Tabella 8. Peso dei dati Mediobanca relativi alle principali imprese del tessile-abbigliamento su dati complessivi nazionali; addetti e fatturato in % (1979-2009)**

	1979	1989	1999	2009
<b>Addetti</b>	9,0	7,8	14,5	22,5
<b>Fatturato</b>	NA	24,0*	40,7	38,0

\* Dato 1990. **Fonte:** Nostre elaborazioni dati Mediobanca; OECD per addetti e SMI per fatturato.

Le società di dimensioni medio-grandi che entrano nel campione di Mediobanca hanno un volume di fatturato complessivo quasi pari a quello del settore di riferimento. Per le industrie “leggere” la copertura è minore e l’impresa-tipo è perciò anche meno rappresentativa<sup>7</sup>. La conduzione imprenditoriale-familiare è ancora presente nella fascia alta delle società tessili e di abbigliamento, anche se in parte la ragione sociale non è che un simbolo, un marchio di qualità e un residuo delle origini. Il fatturato complessivo delle società di grandi dimensioni rappresenta, comunque, una parte sempre più consistente del totale anche in questi settori (e specialmente per il tessile). Le piccole e medie industrie occupano spesso un ruolo di subfornitori, di lavorazioni esterne e, se indipendenti, svolgono lavorazioni di imitazione dei prodotti dei marchi più affermati. Nella fascia delle grandi imprese si trovano inoltre quelle più innovative e dinamiche e quelle che svolgono un ruolo di guida nell’orientare le scelte operative delle imprese minori.

Le società con funzioni di leader in settori maturi non si sono limitate, negli ultimi decenni, a reagire alla concorrenza dei paesi emergenti, della quale erano in parte promotrici dirette o indirette, ma hanno avviato – come vedremo – riorganizzazioni non sempre di successo nel comparto tessile e con risultati complessivamente migliori in quello dell’abbigliamento, verso il quale si sono mosse società che per tradizione provenivano dal primo.

Oltre a differenze strutturali già ricordate, nei due comparti ne sono emerse altre in conseguenza di scelte avvenute nell’arco dei tre decenni esaminati. I cam-

<sup>7</sup> Abbiamo mantenuto i riferimenti settoriali così come sono dati in ciascun annuario. In quelli più recenti le classificazioni sono state fatte secondo i criteri internazionali dei settori (cfr. la Nota metodologica in questo volume). La ricostruzione della serie è tratta dagli annuari Dc 1982, 1991, 2000, 2007, 2015, in modo da avere coorti di un numero stabile di società per ogni decennio circa, salvo per i dati riferiti agli anni '70 nei quali per ogni anno varia il numero di società che compongono l’aggregato.

biamenti di performance delle singole società sono intervenuti su una struttura patrimoniale rimasta abbastanza stabile per le imprese a produzioni prevalentemente tessili. Quelle dell'abbigliamento hanno invece aumentato dagli anni '90 al 2008 la quota delle immobilizzazioni materiali lorde con una dinamica di investimenti più attiva e una propensione all'espansione. Nel complesso la dotazione di immobilizzazioni tecniche rispetto al totale delle attività è poco inferiore al 40% nelle industrie tessili (con una fase di disinvestimenti attorno ai primi anni 2000). Per quelle dell'abbigliamento lo stesso rapporto è inferiore al 30% all'inizio degli anni '90 e raggiunge il picco del 50% nel 2009, seguito da una rapida flessione negli anni successivi alla crisi. Per le principali società non finanziarie il capitale fisso supera di almeno 10 punti percentuali tali livelli, compreso quello più alto delle società del settore dell'abbigliamento.

Il costo del lavoro rispetto al fatturato conferma la maggiore intensità di lavoro delle industrie tessili, superiore a quella di società che producono capi di abbigliamento. In queste ultime la flessione dell'indice negli anni '90 e nella prima parte del decennio successivo riflette gli investimenti realizzati con innovazioni tecnologiche riguardanti la diffusione di modelli a design digitale, ristrutturazione delle catene di vendita e dei flussi di magazzino e degli ordinativi, con una migliore rotazione del ciclo di produzione e di vendite.

#### **4.2. Le società: dimensioni relative e tendenze**

Le società esaminate sono elencate nel prospetto seguente dove, per dare un ordine di grandezza, sono presi a riferimento fatturato, dipendenti, attivo e un indice di redditività del capitale<sup>8</sup>. La scelta è dipesa principalmente dalla continuità di ragione sociale e di bilanci relativi, da una ripartizione tra società dei due gruppi con alcune tra le maggiori e altre di dimensioni relativamente più ridotte, alcune che sono capogruppo e altre semplici società operative. Le divergenze riscontrate – come vedremo – possono essere una conseguenza anche di questo, ma non solo.

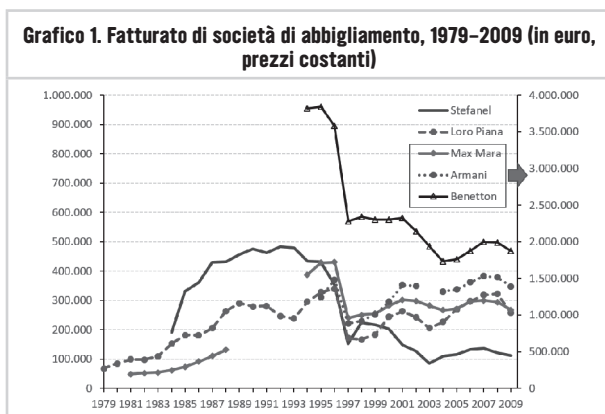
<sup>8</sup> In particolare, si tratta di: Benetton Group Edizione Holding, Giorgio Armani, Max Mara e dal 1996 Max Mara Fashion Group, Miroglio Tessile (poi Gruppo Miroglio), Marzotto Manifattura Lane Gaetano Marzotto & Figli, Stefanel, Loro Piana, Zegna Baruffa Lane Borgosesia, Cotonificio Olcese Veneziano, Lanificio Luigi Botto, Tessitura Monti poi Gruppo Tessile Monti, Vincenzo Zucchi. Le denominazioni della tabella sono usate nel testo per indicare le precedenti ragioni sociali.

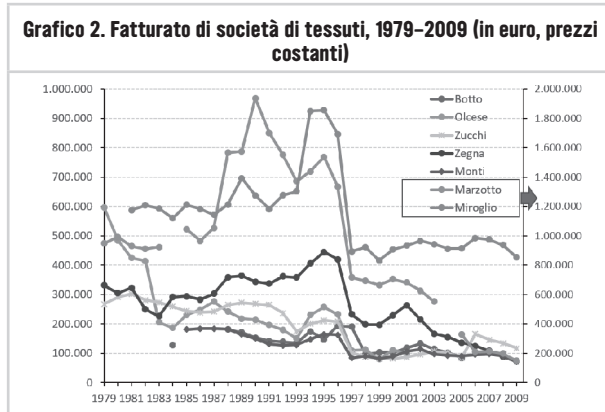


<b>Tabella 9. Società in ordine decrescente di fatturato al 1998 (in 000 di euro)</b>					
<b>Società</b>	<b>Fatturato</b>	<b>Dipendenti</b>	<b>Attivo</b>	<b>Roe</b>	
Benetton	1.980.053	7.235	2.666.989	13,2	
Armani	776.146	2.630	903.376	22,8	
Max Mara	845.182	2.846	1.037.352	15,4	
Miroglio	778.120	5.746	968.299	11,2	
Marzotto	585.307	5.098	631.225	11,4	
Stefanel	189.912	524	260.084	8,8	
Loro Piana	140.594	817	180.943	0,7	
Zegna	167.182	1.216	270.716	6,7	
Olcese	94.920	975	106.064	-26,3	
Botto	81.458	750	81.736	4,4	
Monti	76.500	991	88.596	24,9	
Zucchi	74.201	804	99.115	10,2	
Abbigliamento	6.989.466	28.457	6.892.651	15,9	
Tessile	4.196.392	29.079	4.703.560	12,1	
<b>Società non finanziarie</b>	<b>342.971.726</b>	<b>1.459.104</b>	<b>496.271.239</b>	<b>12,3</b>	

**Nota:** per le società il fatturato è lordo, per i settori è netto. **Fonte:** Mediobanca, *Le principali società italiane e Dati cumulativi di 2015 società* (2007).

I grafici 1 e 2 forniscono una prima indicazione delle dinamiche societarie nell'arco di tempo considerato.





Alcune prime osservazioni sono necessarie. La caduta del fatturato dal 1996 al '97 è dovuta principalmente al consolidamento dei conti a livello di gruppo. Nei gruppi del comparto abbigliamento la riduzione è molto più contenuta e il trend successivo non è per tutte le società discendente. A parte le sfasature di contabilità – per ora non corrette – quel che preme sottolineare sono le tendenze di fondo. Ma – come vedremo attraverso l'analisi di ratios – il passaggio risulta comunque importante e non riguarda solo aspetti contabili, ma investe l'organizzazione complessiva e le strategie aziendali.

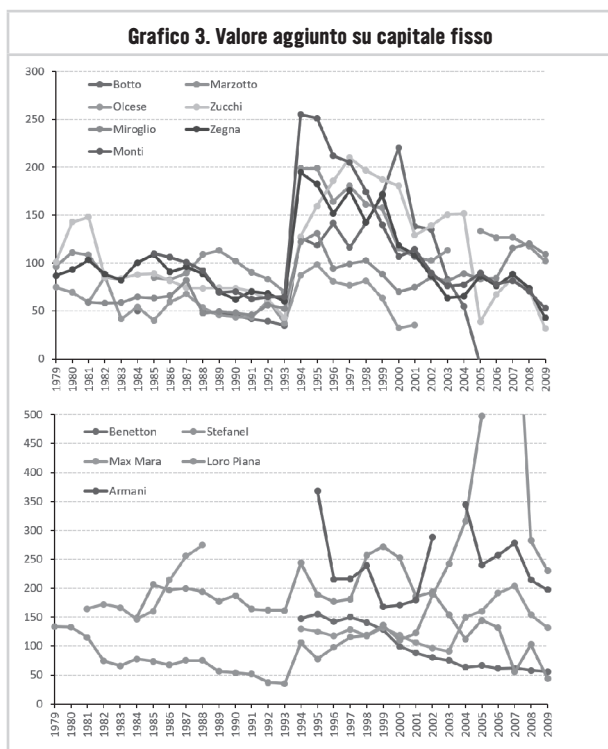
## 5. Produttività, reddito e debiti e dispersione delle dinamiche societarie

### 5.1. La produttività

Gli effetti di queste tendenze di fondo sugli indicatori di produttività sono molto differenziati. La produttività del capitale investito, misurata dal valore aggiunto per unità di capitale fisso (impianti e macchinari), subisce un declino costante sia per le società non finanziarie nel complesso che per le industrie dei due comparti "leggeri". La posizione relativa di queste ultime è più bassa di oltre 30 punti percentuali e con una tendenza alla convergenza prima del 2008 per l'intensificazione del capitale da parte delle società dell'abbigliamento.

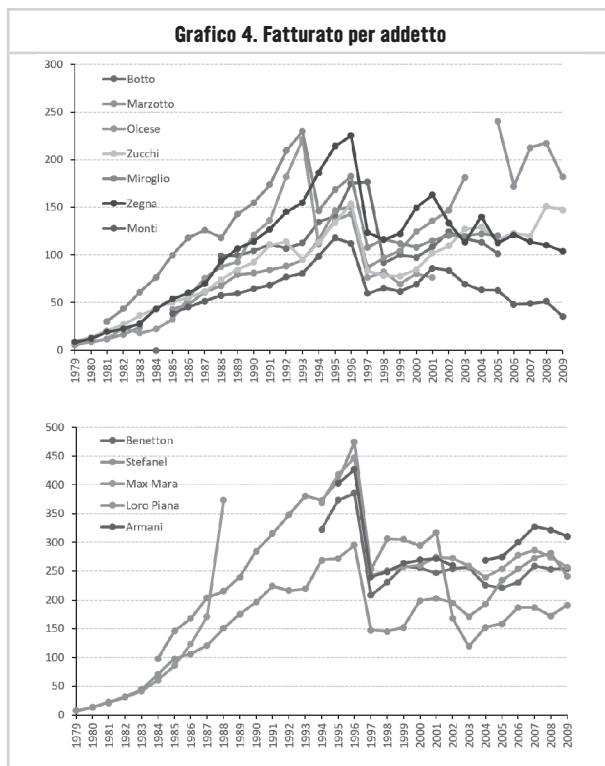
Negli ultimi 30 anni l'indice di produttività del capitale, ossia valore aggiunto per unità di capitale fisso (impianti e macchinari), subisce un declino molto forte per i due comparti esaminati rispetto alla flessione osservata per l'insieme delle società non finanziarie (grafico 3). Gli investimenti in immobi-

lizzi tecnici crescono costantemente. Il comparto dell'abbigliamento è quello con una minore intensità, con uno scarto di 30 punti percentuali circa rispetto al tessile e al totale (i cui valori tendono ad allinearsi dopo il 2007).



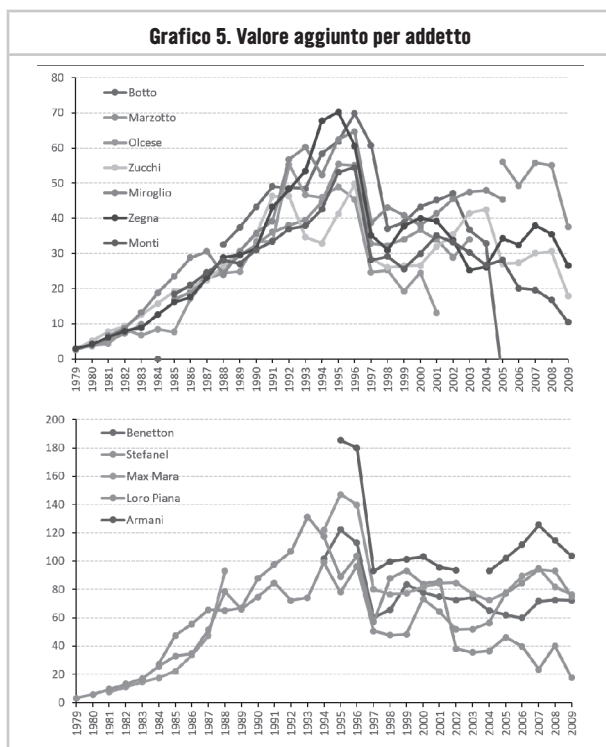
Anche per le singole grandi società tessili prese in esame il declino è vistoso. Il livello dell'indice è comunque superiore a quello dell'intero comparto (anche per le differenze contabili tra le fonti v. la Nota metodologica). Per le società considerate del comparto dell'abbigliamento l'andamento è più diversificato. La holding Benetton segue la media settoriale con un graduale declino dopo il 1997, giustificato da una crescita degli investimenti in capitale fisso. Più marcata è la riduzione per Stefanel dal 1998 in poi. Il capitale si riduce fino al 2003, ma nonostante ciò anche il valore aggiunto diminuisce. Per le altre gli andamenti sono a fasi alterne segno, in buona parte, di variazioni nelle consistenze dovute a investimenti e disinvestimenti. L'indice di Loro Piana cresce dal 1995 e s'impenna dal 2001 per la forte riduzione del capitale investito in immobilizzazioni tecniche. La crisi del 2007-8 smorza le tendenze degli anni precedenti. Un caso singolare è quello di Botto. La società, dopo

un picco in controtendenza raggiunto nel 2000, subisce una rapida caduta culminata nel 2005 (mentre per gli anni successivi mancano i bilanci). Ciò si spiega con la forte riduzione di valore aggiunto fin dai due anni precedenti a fronte di continui aumenti di capitale, il cui valore di bilancio raddoppia dal 2003 al 2005.



Se si considera la produttività delle vendite, misurata dal fatturato per addetto (grafico 4), si può osservare per ogni comparto una crescita sostenuta e costante, minore per l'abbigliamento rispetto a quella generale, e più alta rispetto al tessile. Per le singole società si riscontra una discontinuità nel 1996 attribuibile al consolidamento dei bilanci di gruppo. Nell'insieme il trend è crescente fino a quella data con due eccezioni nel comparto tessile. Infatti, per Miroglio e Olcese l'inversione di tendenza avviene nel 1993, sempre per effetto del consolidamento. Per Olcese la questione è più complessa, perché la riorganizzazione amministrativa e produttiva coincide con una forte crisi aziendale. Perdite consistenti emergono nell'esercizio 1993 per problemi già emersi negli anni precedenti.

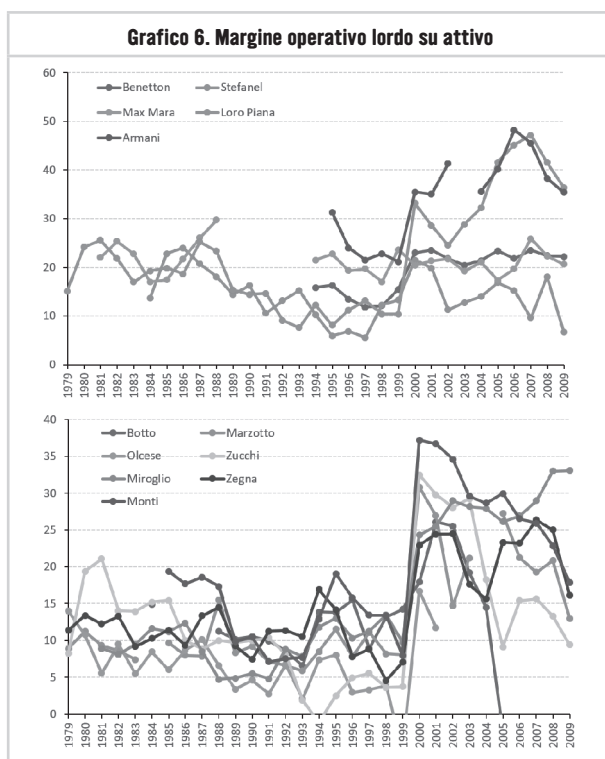
A parte tutto ciò, gli andamenti societari per gli anni seguenti sono contrassegnati da una generale stagnazione dell'indice del prodotto venduto per addetto, salvo – anche qui – alcune eccezioni relative, per il tessile, Marzotto che registra una crescita fino al 2005 e Zucchi fino al 2003. Più movimentata la situazione nel comparto dell'abbigliamento dove Loro Piana, società di tradizione tessile la cui produzione prevalente finisce per spostarsi verso l'abbigliamento, ha un indice in crescita dal 2003 al 2008. Meno pronunciata è la crescita per Armani all'incirca negli stessi anni. Stefanel è alle prese con un tracollo dal 2001 al 2003 e una risalita parziale negli anni seguenti.



Landamento del valore aggiunto per occupato (grafico 5), come indice della produttività del lavoro, non si discosta da quanto appena visto per le tendenze dell'indice del fatturato per addetto. Per le singole società considerate nel tessile la riduzione di produttività del lavoro dopo il 2000 è ancora più forte (rispetto al fatturato per addetto), in particolare per Monti dal 2001, Botto dal 2002, Zucchi con due cadute nel 2004 e 2008. Resistono le società dell'abbigliamento con l'eccezione di Stefanel che subisce un forte calo tra il 2001 e 2002 e un trend discendente dal 2005.

## 5.2. La redditività

L'andamento del margine operativo lordo rispetto alle attività (grafico 6) segue un andamento a campana se analizziamo i dati cumulati (*De*) generali e settoriali. Fino alla prima metà degli anni '90 i profitti nelle industrie tessili e abbigliamento crescono anche se con forti alternanze cicliche (rilevante quella nell'abbigliamento che ha profitti in caduta nella seconda metà degli anni '80). L'inversione di tendenza si colloca attorno alla metà degli anni '90, livello massimo dell'indice esaminato. Dopo di allora la discesa è abbastanza regolare con le seguenti caratteristiche. I livelli di profitto delle società di abbigliamento sono quasi sempre più elevati rispetto a quelli delle società non finanziarie. Sotto si posizionano le industrie tessili con un trend di fondo interrotto da repentine cadute e riprese.

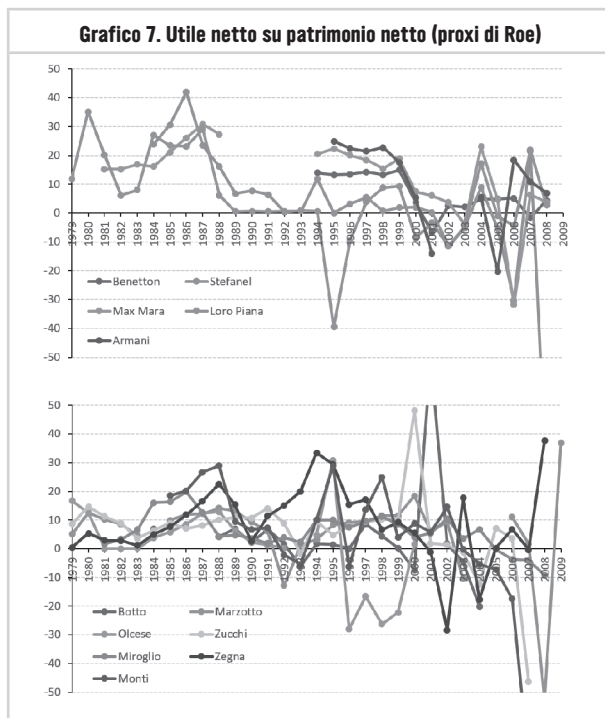


Le singole imprese considerate seguono percorsi molto diversi rispetto a quanto appena visto. Per alcune società di abbigliamento, come Stefanel e Loro Piana (misto abbigliamento e tessuti), i profitti si riducono tra la metà

degli anni '80 e la fine del decennio successivo. Loro Piana recupera redditività in alcuni esercizi con ricadute in anni di congiuntura internazionale più sfavorevole (2002, 2007 e 2009). A un livello più elevato e stabile si posiziona l'indice di redditività di Benetton, dopo il 2000, e di Max Mara, dal 1993 in poi. Il profilo e il livello dei profitti ottenuti da Armani e Loro Piana sembrano del tutto eccezionali dopo il 2000, con un andamento crescente fino al 2006 e 2007.

La situazione delle società tessili segue andamenti molto più variabili, con scarti significativi da una società all'altra. Fino al 2000 i profitti delle singole società si muovono attorno a un trend piatto. La Zucchi subisce una crisi nel 1993-94, l'Olcese nel 1999. Nel 2000 l'indice compie un salto per tutte le società per motivi di cambiamento dei criteri di contabilizzazione e aggregazione seguiti da Mediobanca. Gli anni seguenti sono anni difficili per quasi tutte le società. Solo Miroglio aumenta il proprio tasso di profitto, specialmente tra il 2006 e 2008, mentre in caduta continua è quello di Monti, Zucchi (dal 2004), Marzotto (dal 2005). Il declino è forte per Botto dal 2002 e in perdita nel 2005.

Se si considera un indice di redditività più ristretta, come l'utile netto rispetto al patrimonio netto (grafico 7), si hanno parziali conferme di quanto appena visto. Alcune società dell'abbigliamento ottengono risultati migliori di altre specialmente dopo il 2000. Benetton realizza utili relativamente più stabili e le perdite del 2001 e 2007 sono più contenute. Armani subisce invece una perdita maggiore nel 2001 e ancora nel 2005, parzialmente compensate dagli utili realizzati negli anni successivi. Si tratta di quelle società che avevano investito in capitale fisso e che avevano avviato una riorganizzazione interna secondo modelli gestionali volte a economie di costo e miglior utilizzo delle risorse disponibili. Loro Piana colma le perdite del 2002-3 e 2005-6 con utili più elevati nel 2004 e 2007. La situazione di Stefanel è più difficile per aver subito una grave crisi nel 1995, perdite dal 2000 al 2003 e una nuova grave crisi nel 2006 e 2008. Max Mara ha un anno molto duro nel 2006.



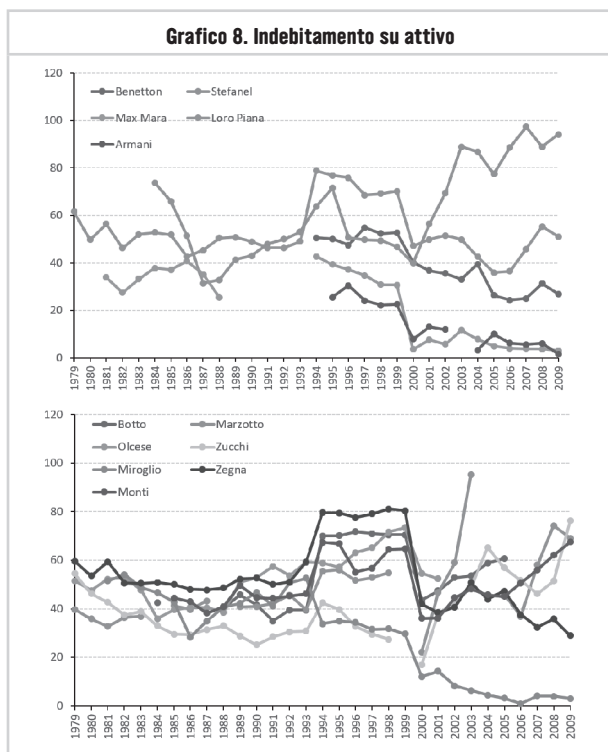
Per le società tessili le condizioni aziendali risultano ancora più precarie. Olcese subisce una crisi grave e lunga dal 1996 al '99 e esce dalla graduatoria di Mediobanca come società operativa dal 2002. Zucchi e Monti subiscono perite pesanti dal 2004 e specialmente nel 2006-7. Ugualmente gravi le perdite subite da Marzotto nel 2008 e peggiori nell'anno seguente. Zegna registra perdite consistenti nel 2002 e '04, ma recupera negli esercizi successivi. Tra tutte le società tessili è la più solida, almeno sotto questo profilo.

### 5.3. Un peso dei debiti molto variabile tra le società e nel tempo

Landamento dell'esposizione debitoria spiega come siano stati finanziati gli investimenti (non solo quelli in capitale fisso, ma anche quelli in attività finanziarie) e quanto il peso del debito abbia contribuito a schiacciare i profitti realizzati. Riguardo all'indebitamento, a livello generale si può osservare che le società non finanziarie mantengono un indebitamento relativamente stabile attorno al 63%, mentre nei settori "leggeri" si ha una riduzione dei debiti rispetto alle attività dall'inizio degli anni '90 in poi. Per essi livelli e tendenze si



sovrappongono tra la metà degli anni '90 e il 2004. Negli anni seguenti si hanno divaricazioni ampie. L'indebitamento continua a diminuire regolarmente, ma dal 2007 al 2015 passa dal 60 circa al 53%.



A livello di singole società (grafico 8) il quadro è molto più complesso e le divergenze ampie da una società all'altra e soprattutto rispetto agli stessi andamenti medi settoriali. Per le produzioni di abbigliamento alcune società, come Benetton e Loro Piana, seguono più da vicino l'andamento generale. Per entrambe il tasso d'indebitamento risulta comunque più basso. Attorno al 50% per Loro Piana negli anni successivi alla crisi del 2008 e tra il 27-30% per Benetton. Benetton è uno dei gruppi principali che - come accennato - ha ottenuto profitti e per questa via ha potuto ridurre l'esposizione debitoria. Invece Stefanel segue un percorso di forte crescita del debito in due ondate. La prima dalla metà degli anni '80 fino al culmine del '94. Il rientro su valori più bassi tra il '96 e il 2000 (attorno al 55% circa) è però seguito da una crescita sostenuta

con un picco dell'80% nel 2003 e stabilizzazione relativa negli anni seguenti. Le altre due società, Max Mara e Armani, avviano una riduzione drastica delle rispettive posizioni debitorie fin dalla metà degli anni '90, da livelli del 40% e del 30% rispettivamente, per schiacciare i propri tassi d'indebitamento nei primi anni 2000 sotto il 10% e ancora più in basso, a livelli irrisori, dopo il 2004. Nel loro caso, più che in quello di Benetton, la redditività ha contribuito a un alleggerimento così radicale che segnala – come mostreremo tra poco – anche cambiamenti strategici e organizzativi.

Il comparto tessile, come visto, ha livelli generalmente più elevati di indebitamento. Le società qui considerate mostrano una relativa stabilità nel lungo periodo con una parentesi di più alti livelli durante gli anni '90 per quattro delle sei società (Zegna, Botto, Olcese e Monti), le stesse che proseguono lungo una traiettoria più bassa negli anni più recenti, ma sempre in linea con quella di prima del '90. Ma è durante l'ultimo decennio che si verificano notevoli cambiamenti. Società più "virtuose" seguono alcune di quelle del comparto dell'abbigliamento. Tra queste Miroglio dopo il 2000 abbassa ulteriormente l'esposizione debitoria già in flessione dagli anni '90. Zegna ha un tasso d'indebitamento relativamente elevato, ma lo riduce dal 2005 sotto il 40%. Per le altre si osserva un debito in salita esponenziale per Marzotto, in crescita per Monti fin sopra il 60% negli ultimi anni, dopo la crisi. Lo stesso vale per Botto.

I dati relativi al peso degli oneri finanziari sui profitti non coprono tutto il periodo per le singole società, ma consentono di trarre alcune conclusioni sulle tendenze appena esaminate relative al debito. A livello complessivo (con dati fino al 2015) si osserva una forte riduzione degli oneri finanziari durante i primi anni '80. Il trend si interrompe verso la fine del decennio e i primi anni '90, ma prosegue poi fino a fine secolo. Da allora in poi gli interessi sui debiti rappresentano tra il 20 e il 30% dei profitti lordi. Negli anni della crisi l'indice cresce, ma solo per i due settori qui considerati per i quali si riduce dal 2010 in poi sotto il livello stesso delle società non finanziarie. Tali andamenti non si riscontrano a livello di singole società, i cui dati si interrompono nel 1999 per ragioni di restringimento della gamma di informazioni contabili negli annuari delle "Principali società". Quelle dell'abbigliamento mostrano un riduzione degli oneri finanziari che culmina al livello più basso del 7-8% circa per Loro Piana nel 1987 e per Stefanel e Max Mara nell'anno successivo. Dopo di allora la tendenza è fortemente crescente fino a picchi anche vicini al 100% per Stefanel e Loro Piana nel 1995. Allo stesso livello si presenta il carico di Benetton la cui serie compare dal 1994. Max Mara non ha seguito l'andamento delle prime due e ha contenuto gli oneri finanziari sotto il 20%. Dal '95 il carico di interessi riprecipita verso il basso: 40% per Benetton e Loro Piana, e attorno al 10 per Stefanel.

Se incrociamo queste tendenze con gli andamenti della redditività si può giustificare l'affannoso tentativo di uscire dall'indebitamento. Per alcune l'operazione riesce con successo, per le altre il debito pur in periodi di tassi d'interesse bassi, ma di prezzi instabili e in tendenziale contrazione, continua a esercitare una pressione sui ricavi aziendali e a rendere più difficili le operazioni di riconversione e di risanamento.

## 6. Conclusioni

Questo saggio presenta un primo studio descrittivo dei dati Mediobanca relativi alle imprese leggere italiane in un'ottica di lungo-medio periodo. L'analisi dei dati nel periodo compreso tra il 1979 e il 2009 evidenzia un generale trend di declino, seguito da una stagnazione del macro settore tessile-abbigliamento italiano, seppur con alcune differenze tra i due comparti. Nel caso del tessile l'arretramento sia in termini di fatturato che di dipendenti si interrompe negli anni Ottanta per poi riprendere e accentuarsi dagli anni Novanta e quindi assestarsi dal 2006 in poi. Per l'abbigliamento l'andamento del fatturato e dei dipendenti sono divergenti, e la discesa è meno marcata. Il fatturato aumenta dagli anni Ottanta e più spiccatamente agli inizi degli anni Novanta per poi assestarsi, mentre il numero degli addetti rimane sostanzialmente stabile dagli anni Ottanta in poi.

Il declino appare più accentuato nel tessile sia a livello aggregato che per le principali società del comparto, in termini di struttura, andamento e performance. Le crisi e il ridimensionamento delle storiche aziende Cotonificio Olcese, lanificio Botto, Tessitura Monti e Vincenzo Zucchi evidenziano le difficoltà del settore soprattutto per le imprese che non hanno inteso avviare processi di divesificazione e integrazione produttiva, come nel caso di Marzotto. Per l'abbigliamento il declino e la stagnazione segue una fase di sostenuta crescita negli anni Novanta che ha portato diverse aziende come Benetton, Armani e Max Mara ad acquisire una posizione di rilievo a livello internazionale agli inizi del 2000. Successivamente nel settore dell'abbigliamento le performance delle principali società si sono diversificate a seconda dei segmenti di mercato di riferimento. Nel caso delle imprese operanti nelle fasce più elevate, come Armani e Max Mara l'andamento è stato migliore, grazie a una maggiore redditività, mentre le società collocate nelle fasce più basse del *mass market* come Benetton, Miroglio e Stefanel hanno registrato cali consistenti di fatturato e utili, dovuti alla concorrenza dei gruppi del *fast fashion* Zara ed H&M.

La crescente competizione e l'allargamento dei confini di mercato ha, quindi, indotto importanti cambiamenti strutturali e strategici nelle imprese contrassegnati da processi di ampliamento e diversificazione produttiva, congiunti nella maggior parte dei casi alla costituzione di gruppi conglomerati, elementi che se da un lato hanno consentito il consolidamento e il posizionamento di alcune storiche imprese a livello internazionale, dall'altro non hanno necessariamente garantito il successo e/o preservato la proprietà dell'azienda. I casi di Vincenzo Zucchi nel tessile e di Loro Piana e Stefanel nell'abbigliamento risultano in tal senso esemplificativi. La Vincenzo Zucchi fondata nel 1920 e specializzata nella produzione di biancheria di lino per la casa, già negli anni Sessanta avviò una politica di sviluppo produttivo attraverso l'acquisizione di nuove società e l'ammodernamento degli impianti, suggellata nel 1986 dall'incorporazione di Bassetti, azienda leader del mercato italiano, e dalla costituzione del gruppo Zucchi-Bassetti, che nel 1989 ha debuttato alla Borsa di Milano. Negli anni Novanta il gruppo conosce una forte espansione internazionale soprattutto in Francia con l'acquisizione di nuove aziende come Jalla, Descamps e Norma, che segnano l'ingresso nella produzione di asciugamani e accappatoi in spugna. Agli inizi del Duemila nonostante nuovi accordi di licenza con importanti marchi come Armani e Ferrari, si manifestano i primi segnali di crisi, legati alla crescente concorrenza della produzione provenienti dai paesi asiatici e dall'ingresso sul mercato nazionale di nuovi competitor a basso costo, come Ikea, che già nel 2005 portano alla chiusura di numerosi stabilimenti sia in Italia che all'estero e alla conseguente riduzione degli organici e considerevoli perdite di esercizio. Negli anni più recenti, malgrado passaggi societari, tentativi di rilancio e riposizionamento attraverso nuove immissioni di liquidità e ristrutturazioni del debito, il gruppo non è riuscito a riassetare i bilanci, giungendo, quindi, nel 2015 all'avvio del concordato preventivo. La Loro Piana dopo aver assunto una leadership mondiale nella produzione di tessuti di cachemire e tessuti di fibre nobili ha sviluppato la produzione di linee di abbigliamento e accessori uomo e donna nel corso degli anni Ottanta e Novanta attraverso una struttura produttiva verticalizzata, articolata in due principali divisioni (tessile e luxury goods), e concentrata in 9 stabilimenti produttivi con circa 2.230 addetti totali, raggiungendo nel 2009 un fatturato pari a 394 milioni euro e un utile netto pari 9 milioni (ROE 5,67%). Risultati e numeri ragguardevoli, ma non sufficienti a misurarsi con i principali competitor a livello globale, e che hanno portato l'azienda familiare dopo sei generazioni e circa 90 anni di storia, a cedere nel 2013 l'80% della proprietà alla holding francese LVMH di Arnault. Nel caso di Stefanel l'allargamento e la diversificazione del gruppo aziendale ha condotto ad esiti decisamente più

infausti. L'impresa, nata nel 1959 con il marchio Maglificio Piave divenuto poi Stefanel nel 1979, ha conosciuto un intenso sviluppo negli anni Ottanta ampliando la produzione dalla maglieria allo sportswear e all'abbigliamento, incrementando in sei anni il fatturato di 4 volte e gli esercizi di vendita da uno a 550, per giungere, quindi, nel 1987 alla quotazione presso la Borsa di Milano. L'acquisizione nel 2002 del 50% di Nuance, una catena di negozi aeroportuali, che avrebbe dovuto segnare una svolta per il gruppo, ha comportato un forte indebitamento, aggravato dalla diminuzione del fatturato in relazione all'ascesa delle catene fast fashion. Neppure due aumenti di capitale, quattro ristrutturazioni del debito, un tentativo di riposizionamento del marchio su una fascia di mercato più alta e, quindi, nel 2011 la cessione di Nuance sono stati sufficienti a risanare i bilanci dell'azienda che ha recentemente avviato il concordato preventivo.

Il quadro generale che emerge dall'analisi ha evidenziato, quindi, più ombre che luci nell'andamento dei settori leggeri e in particolare nel tessile, soprattutto in confronto al "positivo" trend generale delle principali società italiane nel periodo. Un declino, però, che data la crescente complessità ed eterogeneità dell'assetto organizzativo delle principali imprese del tessile e dell'abbigliamento, trasformatesi nei decenni in gruppi polifunzionali e finanziari, soprattutto per l'abbigliamento si apre a nuovi ipotesi e prospettive di ricerca.

## Riferimenti bibliografici

- Barrère C., Santagata W. (2005), *La mode. Une économie de la créativité et du patrimoine à l'heure du marché*. Paris, La documentation Française.
- Berta G. (2004), *Metamorfosi. L'industria italiana fra declino e protezione*. Milano, Università Bocconi Editore.
- Bottazzi G., Dosi G. and Fagiolo G. (2002), *On the Ubiquitous Nature of Agglomeration Economies and Their Diverse Determinants: Some Notes*. in Quadrio Curzio - Fortis (eds. 2002).
- Colli A., Merlo E. (2007), "Family business and luxury business in Italy (1950-2000)", *Enterprise et Histoire*, 46, 112-124.
- Corbellini E., Saviolo S., (2009), *Managing Fashion and Luxury Companies*. Milan, Egea.
- Dunford M. (2006), Industrial Districts, Magic Circles, and the Restructuring of the Italian Textile and Clothing Chain". *Economic Geography*, 82(1), 27-59.
- Federico G. (2003), *La struttura industriale (1911-1996)*. in R. Giannetti, M. Vasta (eds.), *L'impresa italiana nel Novecento*. Bologna, Il Mulino, 41-88.
- Fortis M. (2016), *Production Districts and Their Relevance in the Italian Economy: A Few Analytical Profiles*. in Id. (ed.). *The Pillars of the Italian Economy. Manufacturing, Food & Wine, Tourism*. Cham, Switzerland, Springer, pp. 83-170.
- Gereffi G. (1999), "International trade and industrial upgrading in the apparel commodity chain". *Journal of International Economics*, 48, 37-70.

- Gianelle C., Tattara G. (2009). *Manufacturing abroad while making profits at home: the Veneto footwear and clothing industry*, in M. Morroni (ed.), *Corporate Governance, Organization and the Firm. Co-operation and Outsourcing in the Global Economy*. Cheltenham, UK - Northampton, MA, USA, Edward Elgar, pp. 206-34.
- Giannetti R., Vasta M. (2003). *Le prime 200 imprese manifatturiere (1913-1991)*. in R. Giannetti, Vasta M. (eds), *L'impresa italiana nel Novecento*. Bologna, Il Mulino, 149-186.
- Giannetti R., Vasta M. (2005). *Storia dell'impresa industriale italiana*, Bologna. Il Mulino.
- Kar M. (2015), *The Indian Textile and Clothing Industry. An Economic Analysis*. New Delhi - Heidelberg - New York - Dordrecht - London, Springer.
- Mariotti S., Cainarca G. (1986), "The evolution of transaction governance in the textile-clothing industry". *Journal of Economic Behavior and Organization*. 7, 351-74.
- Marsh P. (2012), *The new industrial revolution: consumers, globalization and the end of mass production*. Yale University Press. New haven and London.
- Mason E.S. (1939), "Price and production policies of large scale enterprise". *American Economic Review*, 29, 61-74.
- Mason E.S. (1949), "The current state of the monopoly problem in the United States". *Harvard Law Review*, 62, 1265-85.
- Menzel M.P., Fornahl D. (2010), "Cluster life cycles-dimensions and rationals of cluster evolution". *Industrial and Corporate Change*, 19(1), 205-238.
- Merlo E. (2008), *Moda italiana. Storia di un'industria dall'Ottocento a oggi*. Venezia. Marsilio.
- Merlo E. (2011), "Italian fashion business: Achievements and challenges (1970s-2000s)". *Business History*, 53, 344-362.
- Morroni M. (1992), *Production Process and Technical Change*. Cambridge, Cambridge University Press.
- Muthu S.S. (ed. 2017), *Sustainability in the Textile Industry*. Singapore, Springer.
- Nardozzi G. (2004), *Miracolo e declino. L'Italia tra concorrenza e protezione*, Roma-Bari, Laterza.
- Nordås H.K. (2004), *The Global Textile and Clothing Industry post the Agreement on Textiles and Clothing*. WTO Discussion Paper n. 5.
- Paris I. (2006), *Oggetti cuciti. L'abbigliamento pronto in Italia dal primo dopoguerra agli anni Settanta*, Milano, Franco Angeli.
- Parker, D., Scottish S.M. (1995), "The Impact of UK Privatisation on Labour and Total Factor Productivity", *Journal of Political Economy*, 42, n. 2, 201-20.
- Parker D., Wu H.L. (1998), "Privatisation and performance: a study of the British steel industry under public and private ownership", *Economic Issues*, 3, 31-50.
- Pent Fornengo G. (1979), *L'industria italiana dell'abbigliamento*. Bologna, Il Mulino.
- Pinchera V. (2009), *La moda in Italia e in Toscana dalle origini alla globalizzazione*. Venezia, Marsilio.
- Porter M.E. (1998), *Competitive advantage: creating and sustaining superior performance*. New York, The Free Press (1a ed. 1985).
- Ricchetti, M., Cietta E. (eds. 2006), *Il valore della moda. Industria e servizi in un settore guidato dall'innovazione*. Milano, Bruno Mondadori.
- Rullani E. (2002), *The Industrial Cluster as a Complex Adaptive System*. in A. Quadrio Curzio - M. Fortis (eds. 2002), *Complexity and Industrial Clusters Dynamics and Models in Theory and Practice*. Heidelberg - New York, Physica-Verlag, 35-61.
- Saviolo S., Testa S. (2000), *Le imprese del sistema moda. Il management al servizio della creatività*, Milano, Etas.

- Schwinge I. (2015), *The Paradox of Knowledge-Intensive Entrepreneurship in Low-Tech Industries. Evidence from Case Studies of the German Textile Industry*. Wiesbaden, Springer.
- Shishoo R. (ed. 2012), *The global textile and clothing industry. Technological advances and future challenges*. Oxford - Cambridge - Philadelphia - New Delhi, Woodhead Publishing.
- Stengg W. (2001), *The textiles and clothing industry in the EU. A survey*. Enterprise Paper, 2.
- Taplin I.M. (2006), "Restructuring and reconfiguration. The Eu textiles and clothing industry adapts to change", in *European Business Review*, 18(3), 172-186.
- Taplin I.M., Witerton J. (2004), "The European clothing industry: Meeting the competitive challenge", *Journal of Fashion Marketing and Management*, 8(3), 256-261.
- Truett L.J., Truett D.B. (2014), "A ray of hope? Another look at the Italian textile industry", *Empirical Economics*, 46(2), 525-542.
- Underhill G.R.D. (1998), *Industrial Crisis and the Open Economy. Politics, Global Trade and the Textile Industry in the Advanced Economies*. Houndmills, Basingstoke, Hampshire - London - New York, Macmillan Press - St. Martin's Press.
- Walter L., Kartsounis G.A., Carosio S., (eds. 2009). *Transforming Clothing Production into a Demand-driven, Knowledge-based, High-tech Industry. The Leapfrog Paradigm*, London, Springer.





# Struttura, comportamento e performance dell'industria farmaceutica italiana dal 1979 al 2008\*

ALBERTO BIANCHI

## 1. Introduzione

L'industria farmaceutica mondiale è stata interessata negli ultimi decenni da importanti mutamenti ambientali che hanno imposto nuove sfide imprenditoriali. Gli anni Novanta, in particolare, hanno segnato, l'inizio di una fase di discontinuità per tutto il settore derivata (Frezza 1994), rispetto al passato, da obiettivi di *policy* diversi in campo sanitario nonché da rilevanti cambiamenti tecnologici in campo chimico-farmaceutico. Riguardo alla regolazione pubblica del comparto, la necessità di contenimento della spesa sanitaria (inclusiva della spesa farmaceutica), cresciuta a causa del progressivo invecchiamento demografico delle popolazioni occidentali, ha condotto i governi ad adottare, sia pure con modalità differenti<sup>1</sup> (Lucioni 1982), con-

\* Un sincero ringraziamento va al Dr. Massimo Di Martino presidente e amministratore delegato di Abiogen Pharma S.p.A. per le osservazioni su alcuni temi affrontati nel presente lavoro. Naturalmente la responsabilità del saggio e degli errori in esso contenuti rimane soltanto mia.

<sup>1</sup> Nel Nord Europa e negli Stati Uniti si è cercato di stimolare una maggiore competizione nel settore farmaceutico responsabilizzando la classe medica nella prescrizione di medicinali con prezzi più bassi. Nei paesi del sud Europa, invece, come in Italia, si è agito sul controllo pubblico dei prezzi delle specialità medicinali.

trolli sempre più ampi ed estesi su tutta la filiera produttiva e sul mercato finale delle specialità medicinali condizionando in modo significativo la *performance* delle imprese. I nuovi indirizzi di *policy* hanno, inoltre, spinto verso legislazioni volte a stimolare una maggiore concorrenza settoriale favorendo la crescita del mercato secondario dei farmaci “generici”<sup>2</sup>, con un’ulteriore alterazione delle preesistenti condizioni di profittabilità aziendale. Per quanto concerne l’innovazione tecnologica, e in particolare nel campo delle biotecnologie, va ricordato il ruolo sempre più centrale svolto dalla ricerca scientifica nella competizione tra le imprese, determinando un legame più stretto, rispetto ai decenni passati, tra condotte innovative e *performance* delle società farmaceutiche.

Di fronte al mutato contesto ambientale sono state rimesse dunque in discussione le vecchie ortodossie (basate quasi esclusivamente sulla ricerca e sviluppo dei prodotti e sull’informazione tecnico-scientifica destinata ai medici e alle strutture sanitarie) che hanno garantito, dall’inizio del secondo dopoguerra<sup>3</sup>, i successi e gli ampi margini di profitto del settore farmaceutico, imponendo così, dal punto di vista manageriale, un profondo ripensamento sull’organizzazione e sulla *governance* dell’intero comparto.

Gli effetti di questo cambiamento sono stati avvertiti, sia pure con un certo ritardo, anche in Italia, dove l’industria farmaceutica presentava una maggiore vulnerabilità rispetto ad altri paesi, determinato da una strategia prevalentemente “imitativa” dei risultati della ricerca farmacologica estera stimolata dall’assenza di un regime nazionale di protezione brevettuale<sup>4</sup>.

Lo scopo di questo breve saggio è descrivere, tra la fine degli anni Settanta e l’inizio della crisi economica del 2008, come si è modificato il rapporto tra le imprese farmaceutiche, italiane e straniere, operanti nel paese e l’ambiente socio-economico influenzato dai cambiamenti internazionali appena descritti a cui si sono aggiunti, come si vede più avanti, mutamenti

<sup>2</sup> Si tratta, come è noto, di prodotti terapeutici che sono, a loro volta, copia di prodotti in commercio il cui brevetto è scaduto.

<sup>3</sup> Anche se l’origine dell’industria farmaceutica può essere fatta risalire alla seconda rivoluzione industriale, è necessario chiarire che il suo successo è solo di recente datazione. Anche se molte delle maggiori imprese farmaceutiche sono nate, sia in Italia che all’estero, nel XIX secolo, la loro importanza e visibilità si è concretizzata solo dopo la fine della seconda guerra mondiale (Chandler 2005).

<sup>4</sup> Come si evidenzia più avanti nel testo, l’adozione di una legislazione di tutela della proprietà brevettuale è avvenuta nel nostro paese soltanto nel 1978 ovvero un secolo dopo ai paesi sviluppati.

più specificamente nazionali<sup>5</sup>. L'analisi mira a descrivere le principali linee evolutive del settore farmaceutico nel suo complesso, focalizzando l'attenzione su struttura, condotta e *performance*. Le fonti privilegiate della ricerca, oltre all'ampia letteratura disponibile sul settore, sono rappresentate dai dati settoriali forniti dall'organizzazione industriale di categoria (Farmaindustria)<sup>6</sup> e dai dati di bilancio delle più grandi società del settore, pubblicati annualmente da Mediobanca<sup>7</sup>.

Il presente lavoro è così strutturato. Nel primo paragrafo, una volta evidenziati il peso e il contributo economico dell'industria farmaceutica nell'economia nazionale, l'indagine si sposta sui mutamenti strutturali del comparto soffermando, in particolare, l'attenzione sulla concentrazione societaria e sulla penetrazione del capitale estero. Nel secondo paragrafo, si delineano i principali cambiamenti di condotta (relativi all'offerta dei prodotti, all'organizzazione aziendale, alla scelta dei mercati di sbocco, ecc.) ascrivibili alle imprese farmaceutiche di proprietà italiana e alle filiali controllate dalle multinazionali straniere nel mutato contesto ambientale. Nel terzo si affrontano le questioni riguardanti la *performance* del settore farmaceutico attraverso l'utilizzo di alcuni indicatori di redditività ricavabili dai dati cumulativi di bilancio delle società farmaceutiche pubblicati da Mediobanca. Seguono, infine, alcune considerazioni conclusive. Va da sé che molti degli argomenti qui affrontati sono, per ragioni di spazio, solo abbozzati e meriterebbero un ulteriore approfondimento in successive ricerche.

<sup>5</sup> In pratica si cerca di descrivere la recente evoluzione del settore farmaceutico italiano "alla Chandler". L'autore americano, infatti, è stato uno tra i primi a testimoniare come il rapporto tra imprese e ambiente socio-economico, influenzato dai mutamenti dei mercati e della tecnologia, condizionasse le linee di sviluppo strategico e, in parallelo, la costituzione dei modelli organizzativi orientati al conseguimento di tali obbiettivi (Chandler 1993).

<sup>6</sup> Si tratta del rapporto annuale Indicatori farmaceutici pubblicato da Farmaindustria a partire dal 1978.

<sup>7</sup> Le pubblicazioni di Mediobanca, che sono state utilizzate per questa ricerca, sono due: 1) *Le principali società italiane* in cui sono riportate le più importanti voci di bilancio delle società, che superano un determinato volume di affari, per tutti i settori industriali e dei servizi; 2) *Dati cumulativi delle società italiane* in cui sono annualmente pubblicati, per ciascun comparto industriale e di servizio, relativamente ad un campione chiuso di grandi imprese, i dati aggregati di bilancio riferiti al decennio precedente. Per una descrizione accurata delle due fonti di Mediobanca si rimanda al saggio di Bianchi A., C. Brambilla, G. Conti, F. Lavista, V. Pinchera, *Un database storico su imprese e settori dell'economia italiana, 1979-2015: nota metodologica*, contenuto nel presente volume.

## 2. Le caratteristiche strutturali dell'offerta farmaceutica italiana

L'importanza dell'industria farmaceutica nel sistema paese, durante il secondo dopoguerra, è stata sicuramente superiore al rilievo che la stessa ha rivestito in termini occupazionali e di valore aggiunto creato. Come si può desumere dai dati censuari (grafico 1), gli addetti del settore non hanno mai superato l'1,5% dell'occupazione industriale e lo 0,35% della popolazione attiva nazionale. Sempre in base ai dati ISTAT (grafico 2), l'incidenza percentuale del valore aggiunto rispetto agli stessi aggregati, grazie all'elevata produttività settoriale del lavoro, è stata più alta del peso relativo degli occupati ma, a sua volta, non è andata mai oltre il 2,5% del valore aggiunto industriale e lo 0,5% del prodotto interno lordo.

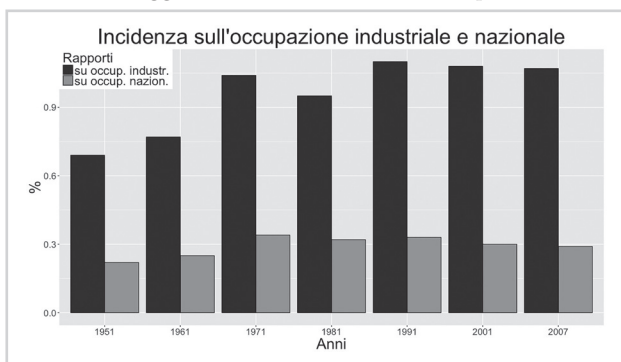


Grafico 1. Fonte: ISTAT

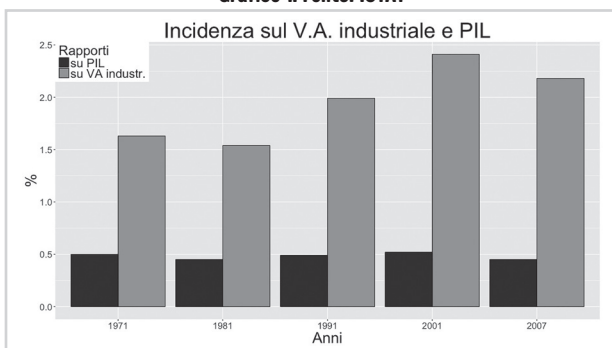


Grafico 2. Fonte: ISTAT

Pur dando un contributo relativamente contenuto al reddito nazionale e all'occupazione, il settore riveste, tuttavia, un ruolo strategico, nel nostro come negli altri paesi industrializzati, sia per le ripercussioni di carattere so-

ciali perché la salute della popolazione dipende dalla disponibilità di medicinali efficaci, sia per le forti interdipendenze esistenti con la ricerca scientifica di base perché una moderna industria farmaceutica apporta grandi vantaggi all'avanzamento delle conoscenze della comunità scientifica nazionale.

Lo studio della struttura produttiva dell'industria farmaceutica italiana consente di delineare un quadro di riferimento generale indispensabile per procedere in seguito all'analisi del comportamento delle imprese a fronte dei cambiamenti regolatori e tecnologici accennati in apertura. L'analisi dell'evoluzione dell'offerta farmaceutica può essere riassunta considerando alcune variabili fondamentali i cui valori biennali sono riportati, per il periodo preso in esame, nella tabella 1: il numero delle imprese e degli occupati del settore nonché il grado di concentrazione industriale e di penetrazione del capitale estero nel comparto farmaceutico italiano. Come si può vedere, nel 1979, l'anno da cui inizia la nostra indagine, le imprese farmaceutiche in attività sul territorio nazionale risultavano pari a 410 unità. Nel trentennio successivo si è verificata una continua e costante riduzione del numero delle aziende che sono passate complessivamente a 260 nel 2007. Sempre nello stesso arco temporale il numero totale dei dipendenti è aumentato di 6.604 unità, determinando di conseguenza un forte aumento della dimensione media aziendale da 188 dipendenti nel 1981 a 277 nel 2007.

**Tabella 1 - Imprese, concentrazione, capitale estero e addetti 1979-2007**

Anni	Numero totale imprese	Quota percentuale fatturato delle prime 25 imprese	Quota percentuale capitale estero	Addetti
1979	410	46,94%	51,99%	NA
1981	348	49,78%	52,10%	65.396
1983	345	49,11%	52,32%	NA
1985	335	48,71%	57,13%	62.252
1987	320	48,97%	58,33%	64.681
1989	305	50,08%	58,68%	67.672
1991	299	50,21%	61,36%	70.020
1993	297	49,02%	64,08%	68.600
1995	293	53,93%	69,54%	63.194
1997	286	55,46%	71,21%	64.936
1999	287	58,51%	70,84%	69.970
2001	294	64,15%	71,46%	79.029
2003	262	63,59%	71,43%	84.088
2005	270	77,29%	76,30%	74.000
2007	260	75,33%	77,40%	72.000

**Fonti:** ISTAT e Farmaindustria

In pratica, nel corso degli anni Ottanta e dei due decenni successivi, in Italia, come nel resto dell'occidente industrializzato, è proseguito il processo di ristrutturazione del comparto farmaceutico, avviatosi già dai primi anni Sessanta e caratterizzato dalla progressiva uscita dal mercato delle imprese più piccole e meno competitive.

A tale proposito va ricordato che alla fine del miracolo economico italiano il numero delle imprese farmaceutiche era piuttosto elevato (equivalente a 1.083 unità) e, a parte alcune eccezioni come Farmitalia, Lepetit e Carlo Erba, si trattava di realtà imprenditoriali medio-piccole concentrate su pochi prodotti o segmenti di nicchia<sup>8</sup> in un contesto internazionale già dominato dai colossi farmaceutici americani, giapponesi ed europei (inglesi, tedeschi, svizzeri) per i quali la ricerca e l'innovazione costituivano fattori fondamentali di competitività. Il processo di graduale ridimensionamento del numero delle imprese attive nella produzione di medicinali, comprese anche i piccoli «retrobottega delle farmacie»<sup>9</sup>, è iniziato a partire dal 1963, quando la tragedia del talidomide, uno dei maggiori disastri nella storia mondiale della farmacologia, dette una forte scossa al settore sollecitando, da parte dei governi dei maggiori paesi industriali, controlli più stringenti sui processi di ricerca e produttivi delle imprese farmaceutiche<sup>10</sup> per evitare il ripetersi di catastrofi simili. L'irrigidimento regolatorio determinò un innalzamento dei costi connessi all'intero processo produttivo risolvendosi, a sua volta, in un meccanismo di selezione

<sup>8</sup> Anche se il livello di competitività delle imprese farmaceutiche italiane è imparagonabile con quello delle grandi multinazionali straniere occorre, tuttavia, sottolineare l'importanza di singole figure imprenditoriali (come, ad esempio, Recordati, Angelini, Zambon, Bracco, Aleotti, Rovati, Dompé, ecc.) che pur partendo da realtà aziendali molto piccole hanno saputo dare, grazie alle proprie specializzazioni produttive, importanti impulsi alla crescita dell'industria farmaceutica italiana tra gli anni Sessanta e Settanta (Bianchi 2009).

<sup>9</sup> Con questo termine s'intendeva di solito quelle piccole imprese che si limitavano al confezionamento dei farmaci per conto terzi e che in passato operavano con le farmacie per la preparazione dei prodotti galenici (Frezza 1994).

<sup>10</sup> Il talidomide è stato un prodotto farmaceutico messo in commercio a cavallo tra gli anni Cinquanta e Sessanta dalla tedesca Grünenthal Group e somministrato alle donne in gravidanza per combattere le nausee. In seguito, fu accertato che il medicinale aveva causato la deformazione di ben oltre 10.000 bambini (focomelici). Ancora oggi le vittime del farmaco sono in attesa di un risarcimento economico da parte dell'impresa farmaceutica tedesca. Sebbene la tragedia fosse avvenuta nel vecchio continente, gli Stati Uniti si attivarono per primi nell'attuare un maggior controllo sull'attività di produzione dei medicinali, a cui seguirono poi analoghe iniziative anche da parte delle autorità governative dei paesi europei (Pammolli 1996).

per le imprese di limitate dimensioni. Era evidente, infatti, che le piccole realtà aziendali non avevano sufficiente massa critica per affrontare da sole la ricerca e lo sviluppo dei farmaci, visto che i maggiori *standard* di sicurezza richiesti dalle autorità di controllo allungavano di fatto l'*iter* d'approvazione di nuovi prodotti. Inoltre, sempre in quegli anni, le barriere tecnologiche divennero sempre più stringenti nel campo della sperimentazione e della ricerca scientifica, per cui si stava ormai già prospettando un mercato farmaceutico in cui l'unica competizione possibile era tra poche imprese in possesso di risorse e conoscenze superiori.

La graduale riduzione del numero delle imprese farmaceutiche trova conferma anche nell'aumento del grado di concentrazione industriale rappresentato, sempre in tabella 1, dall'incremento della quota di fatturato riconducibile alle prime 25 grandi imprese farmaceutiche presenti sul territorio italiano, passata tra il 1979 e il 2007 dal 46,94 al 75,33%. Tale fenomeno è stato accompagnato anche da un costante aumento del grado di penetrazione del capitale estero nell'industria farmaceutica italiana innalzandosi, sempre nello stesso arco temporale, dal 52 al 77,4%. Anche l'ingresso delle grandi multinazionali estere è iniziato in Italia dai primi anni Sessanta (Querini 1979) attraverso, da un lato, l'apertura di proprie filiali commerciali e stabilimenti produttivi, dall'altro, con l'acquisizione e la concentrazione finanziaria di imprese italiane medio-grandi. Nel primo caso si trattava di nuove realtà imprenditoriali aventi principalmente il compito di occuparsi delle fasi finali di trasformazione, confezionamento e distribuzione delle materie prime fornite dalla casa madre<sup>11</sup>. Nel secondo, invece, di acquisizioni strategiche volte a controllare la rete distributiva delle aziende assorbite vista l'appetibilità del mercato farmaceutico italiano posizionato, già dai tempi del "boom economico", tra le prime posizioni a livello mondiale (Querini 1983)<sup>12</sup>.

I primi ad intervenire nell'industria farmaceutica italiana sono stati i grandi gruppi americani, dopo la costituzione del mercato comune europeo come,

<sup>11</sup> Sfruttando gli incentivi fiscali o i finanziamenti agevolati previsti nell'ambito della Cassa per il Mezzogiorno, gli investimenti delle multinazionali farmaceutiche straniere si sono concentrate soprattutto nell'area del centro-sud d'Italia, in particolare nel Lazio. In altre regioni centrali come la Toscana, si nota invece una prevalenza di industrie nazionali mentre al nord, ad esempio in Lombardia, la presenza di imprese è piuttosto bilanciata tra quelle estere e quelle italiane visto che storicamente si trattava di un'area in cui l'investimento nel settore era avvenuto ad opera di imprenditori italiani e stranieri già dagli inizi del XX secolo (Frezza 2005).

<sup>12</sup> A partire dagli anni Ottanta, il mercato farmaceutico italiano si è quasi sempre posizionato tra la sesta e la settima posizione a livello mondiale.

ad esempio, Squibb, Abbott, Pfizer, Parke-Davis e Eli Lilly a cui è seguito, a cavallo tra gli anni Sessanta e Settanta, l'ingresso delle imprese europee come l'inglese SmithKline, la francese Delalande e la scandinava Byk Gulden (Frezza 2005). Nel trentennio successivo è proseguito a ritmo incalzante il processo di concentrazione finanziaria a seguito delle continue operazioni di fusione ed acquisizione caratterizzanti il settore farmaceutico mondiale. Gli effetti di questi sconvolgimenti societari sono riscontrabili nella classifica delle prime dieci grandi imprese farmaceutiche operanti in Italia, riportate in ordine al fatturato, su base quinquennale nella tabella 2. Tra queste spiccano le imprese derivanti da operazioni di fusione avvenute tra colossi mondiali come, ad esempio, Bristol-Myers Squibb nata nel 1989 tra due grandi aziende farmaceutiche americane, Bristol-Myers Company e Squibb Corporation oppure Novartis nata nel 1996 tra le svizzere Ciba-Geigy e Sandoz Laboratories o GlaxoSmithKline nata alla fine del 2000 tra la neozelandese Glaxo Wellcome e la britannica SmithKline Beecham. Meno rilevanti invece sono state, in termini dimensionali, le fusioni societarie italiane, tra cui la più emblematica è stata Farmitalia-Carlo Erba quando, nel 1979, la stessa Farmitalia è passata al gruppo Carlo Erba, con il quale si è fusa sotto il controllo della Montedison<sup>13</sup>. Oltre alle fusioni, va segnalato, per il trentennio 1979-2008, un cospicuo numero di acquisizioni di imprese italiane da parte di multinazionali estere rafforzando così la presenza del capitale straniero nell'industria farmaceutica italiana.

**Tabella 2 – Top ten delle imprese farmaceutiche in base al fatturato**

POSIZIONE	1979	1984	1989	1994
1°	BAYER	FARMITALIA	BAYER	BAYER
2°	FARMITALIA	BAYER	CIBA-GEIGY	CIBA-GEIGY
3°	CIBA-GEIGY	CIBA-GEIGY	FARMITALIA	FATER
4°	LEPETIT	LEPETIT	SIGMATAU	PHARMACIA
5°	SANDOZ	GLAXO	GLAXO	GLAXO
6°	ARTSANA	ARTSANA	BRACCO	FATER
7°	PIERREL	CYANAMID (PFIZER)	ARTSANA	BRISTOL MYERS SQUIBB
8°	ROCHE	SIGMATAU	ROCHE	SORIN
9°	CYANAMID (PFIZER)	ROCHE	MERCK-SHARP-DOHME	BRACCO
10°	SQUIBB	BOEHRINGER	FIDIA	ARTSANA

<sup>13</sup> In seguito, nel 1993, Kabi-Pharmacia ha acquisito Farmitalia-Carlo Erba creando il gruppo Pharmacia che, come si vede dalla tabella 2, ha continuato a restare tra le prime 10 posizioni delle più grandi imprese farmaceutiche operanti in Italia.



## Struttura, comportamento e performance dell'industria farmaceutica italiana

POSIZIONE	1999	2004	2009
1°	ARTSANA	MENARINI	COMIFAR
2°	MENARINI	ARTSANA	MENARINI
3°	COMIFAR	COMIFAR	ARTSANA
4°	GLAXO	PFIZER	SANOFI-AVENTIS
5°	PHARMACIA	GLAXOSMITHKLINE	PFIZER
6°	NOVARTIS	SERONO (MERCK)	SERONO (MERCK)
7°	BAYER	NOVARTIS	ABBOTT
8°	COMIFAR	BRACCO	GLAXOSMITHKLINE
9°	ROCHE	PHARMACIA	NOVARTIS
10°	FATER	ANGELINI	ANGELINI

**Fonte:** Mediobanca

A tale proposito, sono da ricordare, ad esempio, i trasferimenti di Italseber, Zambelletti e Manetti & Roberts alla multinazionale inglese Beecham, di Bonomelli e Italchemi a Glaxo, di Pagni a Parke Davis Warner Lambert, di Camillo Corvi a Rousell, di Ravizza a Basf, di Italchimici a Fisons e di Serono al gruppo Merck-Sharp-Dhome. Anche le maggiori imprese italiane si sono mosse nella stessa direzione, attraverso la costituzione di nuovi gruppi, tra cui il più rilevante è stato quello di Menarini<sup>14</sup> che ha acquisito Guidotti, Malesci, Lusofarmaco e Firma, a cui, poi, se ne sono aggiunti altri di dimensione minore come, ad esempio, quello di Marcucci che ha rilevato la maggioranza dell'Istituto Sieroterapico Italiano, di Farma Biagini e Sclavo e quello di Zambon che ha assunto il controllo di Simes e Tosi. Come si vede nel prossimo paragrafo, in cui si esamina la condotta del settore, le fusioni e i trasferimenti di proprietà delle imprese rispecchiano una strategia imprenditoriale ricollegabile non solo ai processi di concentrazione finanziaria attuatisi in quasi tutti i comparti farmaceutici occidentali ma anche ad una combinazione di cambiamenti manifestatisi, verso la fine degli anni Settanta, nella regolazione pubblica dell'industria farmaceutica nazionale.

<sup>14</sup> La Menarini, che è nata nel 1886 in un piccolo laboratorio della Farmacia internazionale di Napoli, fu trasferita nel 1915 a Firenze dove si costituì in forma di società in accomandita semplice sotto il controllo di Archimede Menarini e Edoardo Bruno. Nel 1975 la maggioranza fu assunta da Alberto Aleotti, già direttore generale, alla cui famiglia il gruppo fa attualmente capo. Al momento si tratta della maggiore società italiana con una produzione annua di circa 570 milioni di confezioni (di cui 450 milioni realizzati negli stabilimenti esteri, principalmente in Germania e Turchia). Oggi, nella diagnostica "in vitro" è il nono gruppo europeo, con una quota di mercato attorno al 3%.

### **3. La condotta delle imprese farmaceutiche**

Per comprendere appieno l'intensificazione dei processi di concentrazione finanziaria ed industriale realizzatasi nel settore farmaceutico italiano a partire dai primi anni Ottanta, occorre fare riferimento a due eventi particolarmente significativi riguardanti il controllo pubblico del comparto che anticiparono l'apertura del decennio: l'approvazione della legge sui brevetti avvenuta, a seguito di una sentenza della Corte Costituzionale, nel 1978 e l'istituzione del Sistema Sanitario Nazionale (SSN)<sup>15</sup> compiutasi sempre nello stesso anno. Le due novità regolative hanno avuto sicuramente un ruolo di primo piano nel determinare la condotta complessiva del settore nel trentennio preso in esame. In primo luogo, la disciplina sulla proprietà intellettuale dell'innovazione ha segnato una svolta nel comportamento delle imprese farmaceutiche italiane perché ha stimolato, nonostante alcune iniziali contrarietà emerse da un'associazione di settore alla nuova regolamentazione<sup>16</sup>, processi d'internazionalizzazione e diversificazione produttiva decisamente più marcati rispetto ai decenni passati. Anche se in termini comparativi la crescita all'estero delle imprese farmaceutiche italiane è stata quantitativamente non paragonabile all'espansione mondiale delle grandi multinazionali straniere, resta il fatto che nel trentennio si è avviato un forte consolidamento all'estero dell'industria nazionale sia attraverso l'apertura di nuove filiali produttive e commerciali in altri mercati europei e di altri continenti (come, ad esempio, la penisola Iberica, la Turchia, l'Asia e l'America Latina nelle esperienze di Menarini, Angelini e Chiesi) sia mediante la specializzazione della farmaceutica

<sup>15</sup> Prima dell'istituzione del Sistema Sanitario Nazionale (SSN) il diritto alla tutela della salute era riconosciuto al singolo individuo, ed esteso a suoi familiari, come lavoratore e non come singolo cittadino. Il sistema delle "casce mutue", tuttavia, creava sperequazioni tra gli assistiti, vista la disomogeneità delle prestazioni sanitarie da loro assicurate. Con la riforma del 1978, il servizio diventava dunque universalistico e la rimborsabilità del prezzo delle specialità medicinali veniva posta direttamente a carico del SSN.

<sup>16</sup> L'entrata in vigore della legge sulla proprietà brevettuale delle innovazioni ha avuto come effetto la costituzione di Farmaindustria, divenuta l'unica associazione industriale del settore farmaceutico. La nuova organizzazione di categoria prendeva il posto delle due precedenti aventi posizioni del tutto antitetiche di fronte all'approvazione della nuova legge. Da una parte, vi era Assofarma che riuniva le grandi multinazionali straniere e alcune grandi imprese italiane (come Recordati, Farmaitalia-Carlo Erba) che riconoscevano la necessità di estendere la brevettabilità dell'innovazione anche sul territorio italiano, dall'altra, Farmunione, che rappresentava per lo più le aziende medio-piccole, decisamente sfavorevole alla sua approvazione.

italiana, sui mercati internazionali, in segmenti di nicchia (come Sorin nel biomedicale, Bracco nei diagnostici e Serono nel Biotec). Inoltre, la nuova disciplina brevettuale è stata anche la molla che ha spinto le imprese italiane ad avviare collaborazioni con le grandi multinazionali allo scopo di rafforzare le sinergie in campo scientifico e tecnologico, come mostrato dagli accordi di ricerca interscambiati, ad esempio, tra Fidia e GeorgeTown Institute for the Neurosciences oppure tra il gruppo Merck e l'italiana Sigma Tau. In altri casi si è proceduto alla concentrazione di competenze tecnico-scientifiche in specifiche realtà aziendali come è stato, ad esempio, il caso della Pharmatrix Corporation, specializzata in tecnologie di Delivery, a cui ha partecipato dal 1987 anche l'italiana Recordati (Frezza 2005).

Infine, la tutela brevettuale ha avuto un ruolo di primo piano nello stimolo dell'attività di ricerca e sviluppo domestica. Storicamente, le aziende a capitale italiano, integrate dalla ricerca alla commercializzazione, come (per citare le più importanti) Menarini, Dompé, Bracco, Angelini, Sigma Tau, Italfarmaco hanno da sempre investito nelle proprie strutture di ricerca e hanno saputo gradualmente conquistare gli spazi lasciati liberi dalle grandi concentrazioni farmaceutiche specializzandosi in quelle «nicchie terapeutiche» (Sironi 2004) che richiedevano un'agilità produttiva che le grandi multinazionali non potevano avere. Durante l'arco temporale oggetto di questo studio, gli importanti risultati ottenuti nel campo cardiovascolare, oncologico e delle malattie respiratorie sono stati il frutto dei consistenti investimenti effettuati dalla ricerca farmaceutica condotta dalle imprese di proprietà italiana<sup>17</sup>. Dai dati forniti da Farmaindustria, si osserva che il settore farmaceutico nel 2007 rappresentava il 6,6% del totale della ricerca svolta in Italia ovvero un valore nettamente superiore al suo peso in termini di addetti e valore aggiunto (cfr. par. 2). In termini assoluti, le spese di ricerca e sviluppo (R&S) hanno collocato il comparto al terzo posto dopo i settori «aeronautica e mezzi di trasporto» e «apparecchi per le comunicazioni radio e TV», ma in termini relativi, misurando le spese di R&S rispetto al fatturato, il farmaceutico è stato il primo settore e gli investimenti in R&S per addetto sono stati più elevati della media industriale italiana (fatto 100 il settore farmaceutico, l'industria manifatturiera valeva 11) (Farmaindustria 2008). Tuttavia, da una comparazione con gli altri paesi industrializzati, si evince quanto sia stata poco consistente la ricerca farmaceutica italiana rispetto ad altri contesti nazionali.

<sup>17</sup> Le strutture di R&S delle imprese italiane hanno da sempre avuto il vantaggio di poter seguire indirizzi strategici propri. Le filiali delle multinazionali, invece, godono di minore libertà dovendo subire le decisioni strategiche stabilite dalla rispettiva casa madre.

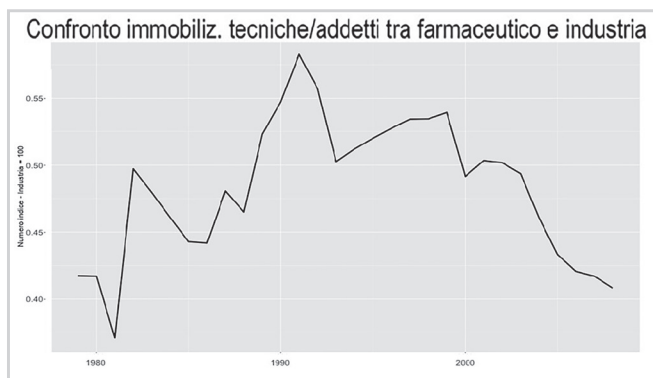
Nel 2005, infatti, la ricerca farmaceutica italiana equivaleva a 839 milioni di euro contro i 30.458 degli USA, i 6.747 del Giappone, 6.484 della Gran Bretagna, di 3.696 della Francia, i 3.820 della Germania e i 2.297 della Svizzera. Le ragioni di questa persistente discrepanza sono dipese, come si vede tra poco, anche dagli stessi processi di concentrazione industriale, che hanno coinvolto tutta la filiera (dalla ricerca alla produzione dei farmaci), adottati dalle grandi multinazionali a partire dagli anni Novanta in poi.

Con l'introduzione dell'altro cambiamento istituzionale, rappresentato dal Sistema Sanitario Nazionale che dettava il passaggio da un meccanismo frammentario di rimborsabilità posticipata fondato sulle mutue ad uno schema di rimborsabilità diretta da parte dello stato basato sul potere di scegliere i farmaci mutuabili, è comparsa la necessità, soprattutto da parte delle multinazionali straniere, di acquisire e sfruttare nel mercato italiano conoscenze e strutture che avessero un collegamento diretto con il nuovo interlocutore pubblico. È proprio nella prima metà degli anni Ottanta che nel settore farmaceutico italiano sono comparse e hanno preso sempre più piede intese di *cooperative marketing* tra imprese straniere e nazionali allo scopo di esercitare un'adeguata capacità di *lobbying* verso i nuovi centri pubblici decisionali<sup>18</sup>. L'affermazione del modello di *co-marketing* è risultata del resto vantaggiosa anche per talune imprese nazionali che hanno potuto sfruttare la propria capacità di conoscere, condizionare e influire sul mercato italiano, quale punto di forza per attrarre accordi commerciali con le imprese straniere ed ottenere i loro prodotti farmaceutici in licenza. Questi intrecci hanno rappresentato, in diverse occasioni, il punto di partenza per l'avvio di acquisizioni di imprese italiane da parte di multinazionali straniere, al fine di assorbire la loro intera struttura distributiva e tutto il loro *know-how* commerciale.

In linea generale, il comparto farmaceutico italiano ha assunto dagli anni Ottanta in poi una connotazione sempre più commerciale e meno manifatturiera. Infatti, l'obbiettivo dei colossi farmaceutici mondiali di conseguire una maggiore penetrazione del mercato farmaceutico italiano attraverso lo sviluppo e il consolidamento dell'area commerciale è coinciso con la nuova strategia di realizzare una più marcata razionalizzazione

<sup>18</sup> A tale proposito Leonardo Frezza definisce gli accordi di *cooperative marketing* come «azione comune delle forze commerciali, o parti di esse, di due o più aziende (farmaceutiche) per raggiungere una concentrazione di risorse ed esercitare un impatto sulle strutture sanitarie tale da alterare rapidamente la situazione concorrenziale all'interno della classe terapeutica di riferimento» (Frezza 1994).

delle strutture produttive. I crescenti costi in termini di tempo e d'investimento, associati allo sviluppo di nuovi prodotti<sup>19</sup> (Paci 1990), hanno spinto le multinazionali a sfruttare quanto più possibile le economie di scala dei rispettivi processi concentrando le attività produttive in pochi e grandi stabilimenti. In particolare, questo mutamento strategico ha condotto alla progressiva smobilitazione degli impianti più piccoli e periferici che non godevano di economie di costo paragonabili a quelle dei grandi centri produttivi gestiti direttamente dalle grandi case madri estere dove invece è avvenuta, col tempo, la riassunzione della manifattura dei principi attivi attraverso una maggiore integrazione produttiva e un più efficace controllo tecnologico su tutta la filiera. Il progressivo disinvestimento degli impianti di produzione ha riguardato anche l'Italia, paese tradizionalmente senza proprie multinazionali (con la sola eccezione di Menarini) e a vocazione prevalentemente "imitativa". Tale fenomeno può essere colto dall'andamento del volume degli investimenti fissi lordi delle maggiori aziende operanti in Italia come riportato da Mediobanca (grafici 3 e 4).



**Grafico 3 - Fonte: Mediobanca**

Come si può vedere dal grafico 3, il settore farmaceutico ha presentato, per tutto il trentennio, un rapporto tra immobilizzi tecnici ed addetti decisamente inferiore a quello dell'industria manifatturiera italiana. Ciò dimostra quanto nell'industria farmaceutica il capitale fisso abbia svolto un ruolo meno crucia-

<sup>19</sup> Nel corso degli anni vi è stato un allungamento dei tempi necessari perché un dato principio attivo superasse tutte le fasi della ricerca e della sperimentazione fino alla commercializzazione sul mercato: dai 3-5 anni negli anni Sessanta, dai 9 ai 12 anni negli anni Ottanta, ai 12-15 anni dagli anni Novanta fino ai giorni nostri.

le all'interno del processo produttivo comparativamente a tutti gli altri settori industriali, mentre, come si è visto prima, hanno più rilevanza, rispetto alla media industriale del paese, gli investimenti in ricerca e sviluppo<sup>20</sup>. Raffrontato all'aggregato industriale (preso come base = 100), il suddetto rapporto riferito al settore farmaceutico tende a raggiungere il suo apice intorno ai primi anni Novanta (avvicinandosi a 60), per poi scendere bruscamente e aumentare gradualmente alla fine del decennio e, infine, declinare progressivamente nel nuovo millennio.

La crescente smobilitazione degli impianti produttivi nel settore farmaceutico italiano, a partire dagli anni Novanta, viene ulteriormente confermata dal rapporto tra investimenti fissi e fatturato, come mostrato dal grafico 4, seguendo più o meno lo stesso andamento fatto registrare dallo stesso indice calcolato per l'aggregato industriale.

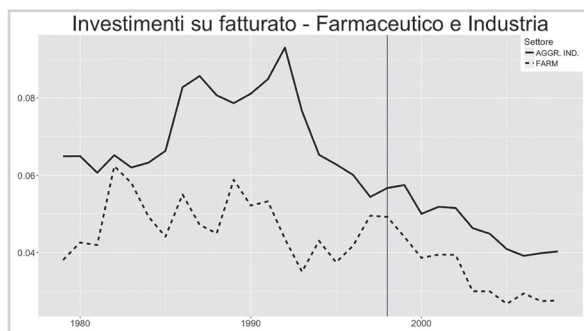


Grafico 4 - Fonte: Mediobanca

La razionalizzazione dei processi industriali, condotta principalmente mediante la centralizzazione in un'unica unità produttiva di stabilimenti sparsi in aree o paesi diversi, è stata resa possibile grazie ad un'ulteriore intensificazione del processo di concentrazione finanziaria condotto sempre attraverso

<sup>20</sup> In linea generale, gli impianti farmaceutici si dividono tra quelli che sono necessari per la produzione delle sostanze di base (sintesi, fermentazione ed estrazione) e quelli che sono utilizzati per la realizzazione dei prodotti farmaceutici finali (perfezionamento, dosaggio ed inscatolamento). Mentre i primi richiedono un maggior utilizzo di capitale fisso e possono presentare vantaggi di scala, i secondi comprendono macchine che svolgono prevalentemente operazioni semplici e non richiedono pertanto elevati immobilizzi tecnici (Paci 1990).

operazioni societarie di acquisizione e fusione. Ad esempio, emblematico è stato il processo d'integrazione produttiva seguito dalla multinazionale inglese SmithKline-Beecham che ha concentrato in un'unica unità industriale gli stabilimenti di imprese sotto il proprio controllo (Smith Kline & French, Zambelletti, Isf e Manetti & Roberts). Un'altra impresa britannica, Fisons, acquisita dal gruppo francese Rhône-Poulenc-Rorer, ha integrato produttivamente gli impianti italiani di Italcimici mentre Glaxo-Wellcome, ha proceduto alla ristrutturazione e concentrazione delle attività produttive degli stabilimenti di Verona e Parma. Sulla stessa falsariga si sono mosse le filiali di multinazionali tedesche e svizzere come, ad esempio, Ciba-Geigy, concentrando le proprie produzioni italiane presso lo stabilimento di Torre Annunziata oppure Basf che ha proceduto ad affidare ad un unico centro produttivo le attività delle controllate Bioresearch, Ravizza e Knoll. La stessa dinamica è stata registrata anche per le imprese italiane come, ad esempio, Serono che ha razionalizzato le proprie produzioni investendo nello stabilimento di Bari e Sandoz che ha concentrato sia le proprie attività produttive sia quelle delle imprese acquisite (come Samil).

#### **4. I risultati economici settoriali**

La *performance* economica delle imprese farmaceutiche operanti sul territorio nazionale dipende, in larga parte, dalla regolazione pubblica dei prezzi amministrati delle specialità medicinali. La titolarità del potere di fissare i prezzi dei farmaci, fino al 1993 riconosciuta al Comitato Interministeriale Prezzi (CIP) e in seguito al Comitato Interministeriale per la Programmazione Economica (CIPE), è passata alla competenza dell'Agenzia Italiana del Farmaco (AIFA) a partire dal 2004. Fino ai primi anni Novanta, i risultati economici delle aziende farmaceutiche sono stati sostanzialmente buoni potendo godere di una relativa stabilità assicurata da un atteggiamento equilibrato e non restrittivo sia da parte dell'autorità di regolazione in termini di controllo dei prezzi, sia da parte dei governi in termini di spesa pubblica sanitaria. Dopo gli scandali di "farmacopoli", è iniziata una fase di netta cesura rispetto al passato in cui l'attenzione delle istituzioni si è sempre più focalizzata sul costo sociale del trattamento terapeutico associato alle specialità medicinali e sulla disponibilità delle risorse pubbliche da destinare alla copertura della spesa sanitaria. Dal mutato atteggiamento regolatorio sono derivati dunque provvedimenti di prezzo più rigidi che hanno condizionato la *performance* reddituale delle imprese operanti nel settore farmaceutico italiano.

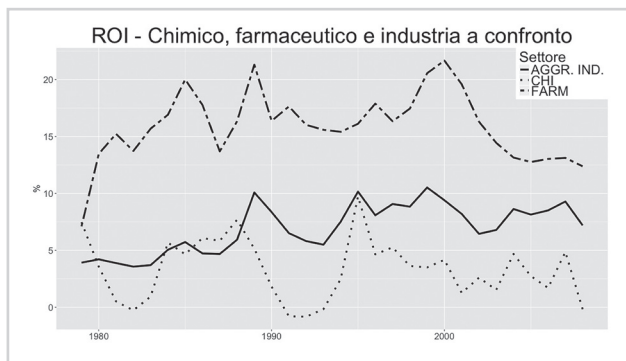


Grafico 5 - Fonte: Mediobanca

Come si può vedere dal grafico 5, in cui viene messo confronto l'andamento del ROI (Return on Investment)<sup>21</sup> (Mediobanca, vari anni) delle principali società farmaceutiche con la dinamica dello stesso indice calcolato, sempre per grandi imprese, per il comparto chimico e l'intero aggregato industriale italiano, la redditività della gestione caratteristica farmaceutica ha subito una flessione nella prima metà degli anni Novanta per poi risalire, verso la fine del decennio, allo stesso livello degli anni Ottanta e subire di nuovo una brusca caduta a partire dal nuovo millennio. Tale tendenza mostra una certa aderenza con le decisioni prese dalle autorità di regolazione in merito alle classi di rimborsabilità e ai livelli di prezzo attribuiti ai farmaci in commercio. Un primo rilevante cambiamento regolatorio è emerso nei primi anni Novanta, quando la revisione del prontuario terapeutico nazionale, modificando a più riprese le classi di rimborsabilità, ha provocato andamenti altalenanti del fatturato e conseguentemente della redditività delle imprese farmaceutiche. Gli interventi di controllo sui prezzi sono stati ancora più stringenti nel nuovo millennio. La fase più critica è stata quella compresa tra il 2001 e il 2006 in cui i prezzi sono scesi complessivamente del 14,2% a fronte di un aumento dell'inflazione pari al 12% (Farmaindustria 2007)<sup>22</sup>.

Anche negli altri paesi si è registrato nella prima metà del nuovo millennio un calo dei prezzi dei medicinali, ma in Italia la riduzione è stata molto

<sup>21</sup> Come è noto, il ROI è un indice calcolato come rapporto tra il risultato operativo (o della gestione caratteristica) e il capitale investito netto operativo (Galdabini 1983).

<sup>22</sup> Nel trentennio oggetto d'analisi, i prezzi dei medicinali non hanno tenuto lo stesso passo dei costi di produzione. Complessivamente i prezzi hanno subito un aumento medio del 112,76% contro un incremento dei salari pari al 296,42% e del costo delle materie prime farmaceutiche pari al 263,74%.



più repentina. Da uno studio che ha messo a confronto i prezzi al pubblico delle prime 500 specialità medicinali vendute in Italia con quelli esistenti, per confezioni uguali, in altri mercati europei, risulta che, durante il 2006, nel nostro paese i prezzi sono stati inferiori a quelli delle altre nazioni ben nel 70,4% dei casi (Farmaindustria 2007). È evidente che in un simile contesto le dimensioni del mercato dei “generici” sono in Italia piuttosto contenute soprattutto a fronte della esiguità dei margini di manovra per i “genericisti” causata dal livello contenuto dei prezzi per i medicinali di marca al momento della scadenza del brevetto (Pammolli 1996). Da questo punto di vista, le imprese farmaceutiche *unbranded* hanno lamentato in anni recenti l'assoluta mancanza di condizioni concorrenziali che invece sono risultate maggiormente presenti in altri mercati europei (tedesco e svizzero) ed extraeuropei (statunitense, canadese e giapponese)<sup>23</sup>.

La riduzione della redditività aziendale nel comparto farmaceutico italiano è stata, inoltre, accompagnata anche da una progressiva erosione dei margini di autofinanziamento. Come si può vedere dal grafico 6, in cui viene riportato l'andamento del rapporto tra autofinanziamento e fatturato per le più grandi società farmaceutiche presenti in Italia (Mediobanca, vari anni), si è assistito, soprattutto nel nuovo millennio, ad un consistente peggioramento della capacità delle imprese di finanziarsi attraverso la creazione di risorse proprie comparativamente ai livelli raggiunti negli anni Ottanta. Questa tendenza sembra testimoniare un progressivo indebolimento di uno dei maggiori punti di forza su cui hanno fatto affidamento, nei passati decenni, le imprese del settore farmaceutico italiano per far fronte alla scarsa sensibilità che il settore bancario italiano ha quasi sempre mostrato nell'uso degli strumenti di *project financing* in ambito industriale e, in particolare, farmaceutico<sup>24</sup>.

<sup>23</sup> La pressione competitiva dei “generici” negli altri mercati si è sentita molto di più rispetto al nostro paese soprattutto per i medicinali destinati a curare grandi malattie sociali (che richiedono il consumo di ampi volumi) come il diabete, l'ipercolesterolemia, l'ipertensione, l'osteoporosi, ecc.. Si stima che l'uscita di un prodotto generico sul mercato può decurtare i prezzi dei prodotti *branded* entro un range compreso tra il 30 e il 40%.

<sup>24</sup> L'esperienza di altri paesi ha mostrato invece quanto la finanza di progetto sia stata molto importante nel sostenere iniziative nel comparto farmaceutico. Ad esempio, il National Institutes of Health (NIH) è un'agenzia del Dipartimento della salute e dei servizi umani degli Stati Uniti che effettua, ogni anno, la valutazione di molti progetti di ricerca e alla fine ne finanzia pochi destinando loro, tuttavia, considerevoli risorse. Per assolvere questi compiti sono necessarie competenze tecniche in grado di interpretare le potenzialità scientifiche ed industriali dei progetti presentati che ancora oggi latitano nel sistema creditizio italiano.

In ultima analisi, secondo alcuni autori, la riduzione degli investimenti nel settore, oltre a dipendere dalla razionalizzazione della struttura produttiva delle imprese (cfr. par. 3), è derivata anche dalla minore capacità di autofinanziamento delle stesse provocata dall'insufficiente dinamica dei prezzi amministrati (Paci 1990).

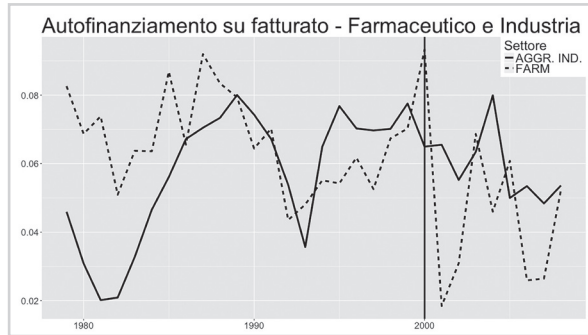


Grafico 6 - Fonte: Mediobanca

## 5. Conclusioni

La storia dell'industria farmaceutica qui narrata per il trentennio 1979-2008 ricomincia, per alcuni aspetti, quella dei decenni precedenti segnati da una graduale ristrutturazione settoriale, caratterizzata da un processo di concentrazione finanziaria operata prevalentemente dalle grandi multinazionali straniere nel tentativo di penetrare il mercato italiano che, fin dal "boom economico", è stato uno dei più grandi a livello mondiale. Inoltre, non va sottovalutato l'esempio di alcuni imprenditori italiani che hanno dimostrato un attivo impegno nell'innovazione, nel tentativo di assicurare alle proprie aziende una prospettiva di crescita basata sulla competitività piuttosto che sul perseguimento di una mera strategia imitativa che, comunque, è stata per lungo tempo stimolata nel nostro paese a causa dell'assenza di una legislazione brevettuale fino alla fine degli anni Settanta. Questo è stato il caso, ad esempio, di Menarini che, in risposta alla concorrenza delle multinazionali straniere, ha avviato un processo di concentrazione societaria costituendosi come gruppo e di diverse imprese, come Recordati, Sigma Tau, Bracco, Sclavo, Dompé, Chiesi, Zambon, Italfarmaco e Fidia, che hanno attuato investimenti significativi in ricerca e sviluppo.

Tuttavia, a partire dagli anni Novanta, sono da segnalare alcuni elementi di discontinuità rispetto al passato, quando sono cambiate le strategie di *management* dei grandi gruppi farmaceutici mondiali basati sulla centralizzazione della produzione in pochi e grandi complessi per sfruttare al meglio le economie di scala di tutta la filiera farmaceutica (dalla ricerca fino al prodotto finito). Tale fenomeno ha interessato anche l'Italia ed ha comportato una riduzione degli stabilimenti sul territorio nazionale. Inoltre, a questo fenomeno si è aggiunto anche un mutamento delle condizioni di redditività delle imprese. Infatti, dagli elevati livelli di marginalità economica registrati negli anni Ottanta, il settore ha iniziato a sperimentare, soprattutto dagli inizi del nuovo millennio, un graduale peggioramento della propria *performance* economica a causa di una regolazione pubblica, in termini di controllo dei prezzi delle specialità medicinali e di contenimento della spesa farmaceutico-sanitaria, sempre più stringente. L'insufficiente dinamica dei prezzi amministrati ha fatto così venire meno nel nostro paese la pressione competitiva proveniente dal mercato dei prodotti a brevetto scaduto che, invece, ha conosciuto una forte espansione in altri paesi industrializzati.

## Riferimenti bibliografici

- Bianchi A., Fanfani, T. (2009), *Il settore farmaceutico pisano. Un incontro tra ricerca e impresa in cento anni di storia*, Pacini Editore, Pisa.
- Cesali S. (1979), *Le imprese farmaceutiche*, Giuffrè Editore, Milano.
- Confindustria (2012), *Struttura e performance della filiera della salute (2007-2012)*, Roma.
- Chandler A.D. (1993), *Strategia e struttura: storia della grande impresa americana*, Franco Angeli, Milano.
- Chandler A.D. (2005), *The remarkable story of the evolution of the modern chemical and pharmaceutical industries*, Cambridge (Mass.), Harvard University Press.
- Farmaindustria (vari anni), *Indicatori farmaceutici*, Farmaindustria, Roma.
- Ferraguto G., Lucioni, C., Onida, F. (1983), *L'industria farmaceutica italiana. L'innovazione tecnologica*, Il Mulino, Bologna.
- Frezza L. (1994), *Farmaci e imprese*, Il Sole 24 ORE, Milano.
- Frezza L. (2005), *Industria farmaceutica e management*, Il sole 24 ORE, Milano.
- Galdabini G., Gros-Pietro G.M., Rolfo, S. (1983), *L'industria farmaceutica italiana. Analisi economico-finanziaria del settore*, Il Mulino, Bologna.
- Lucioni C. (1981), «Metodi di controllo dei prezzi dei medicinali in Italia», in *Economia pubblica*, 11, 2.
- Maggioni V. (1984), *L'industria farmaceutica*, Editoriale Scientifica, Napoli.
- Mediobanca (vari anni), *Le principali società italiane*, Mediobanca, Milano.
- Mediobanca (vari anni), *Dati cumulativi delle società italiane*, Mediobanca, Milano.
- Paci R. (1986), «Alcuni aspetti dell'attività innovativa delle imprese farmaceutiche in

- Italia. Una indagine econometrica”, in *Rivista internazionale di scienze sociali*, 94, 4.
- Paci R., Ratti, U. (1987), “La spesa per la ricerca nell’industria farmaceutica privata”, in *Rivista di politica economica*, 77, 10.
- Pac, R. (1990), *Innovazione tecnologica e intervento pubblico nell’industria farmaceutica*, Franco Angeli, Milano.
- Paganelli O. (2002), *Analisi di bilancio. Indici e flussi*, Utet, Torino.
- Pammolli F. (1996), *Innovazione, concorrenza e strategie di sviluppo nell’industria farmaceutica*, Edizioni Angelo Guerini e associati S.p.A., Milano.
- Querini G., Sanna, F. (1983), *L’industria farmaceutica italiana. Il mercato internazionale*, Il Mulino, Bologna.
- Querini G. (a cura di) (1979), *La concentrazione industriale. Problemi teorici e considerazioni empiriche con particolare riferimento all’industria farmaceutica*, Franco Angeli, Milano.
- Scherer F.M. (1980), *Industrial Market Structure and Economic Performance*, Houghton Mifflin Company.
- Sironi V.A. (2004), “Dal passato le strategie per il futuro”, in *Farmanews Italia*, settembre.
- Wardell W.M. (1975), L. Lasagna, *Regulation and Drug Development*, in American Enterprise Institute, Washington D.C..

# Analisi di struttura-comportamento-performance: l'industria pesante

CARLO BRAMBILLA E FABIO LAVISTA\*

## 1. Introduzione

Il saggio presenta i primi risultati di una più ampia ricerca a più mani che ha come obiettivo quello di analizzare i cambiamenti strutturali avvenuti in seno all'industria italiana nel corso degli ultimi quarant'anni. In particolare la ricerca mira a dare conto della trasformazione settoriale che ha caratterizzato il comparto manifatturiero nei decenni che vanno dal generale rallentamento dell'economia nazionale registratosi alla metà degli anni settanta alla più recente crisi finanziaria internazionale avviatasi nel corso del 2008. Le scelte di politica economica indotte dalle difficoltà degli anni settanta sembrano infatti aver avviato una fase di ridimensionamento dell'industria italiana, continuata nei venti anni successivi e culminata nella crisi odierna, dopo un decennio di progressiva perdita di competitività (Traù, 1999; De Nardis, Traù, 2005; Perulli, Pichierri, 2010).

Per dare conto di queste trasformazioni la ricerca si è posta l'obiettivo di analizzare l'evoluzione dei singoli settori industriali sotto il profilo qualitativo e quantitativo, facendo uso principalmente dei dati riportati nella pubblicazione *Le principali società italiane*, curata e pubblicata da Mediobanca dal 1966 a oggi (in merito alla banca dati che su tale fonte è stata costruita, la sua struttura e i problemi che il suo utilizzo comporta, si veda più sopra la "Nota metodologica" a pp. 235-239). I dati di bilancio delle principali imprese italiane

\* Il saggio, pensato unitariamente dai due autori, è da attribuirsi a Carlo Brambilla per i paragrafi 3, 4 e 5, a Fabio Lavista per i paragrafi 1 e 2.

riportati in questo annuario, costituendo un campione rilevante dell'industria manifatturiera nazionale, permettono di ricostruirne sinteticamente, anno per anno, l'evoluzione della struttura settoriale. Quello che si è cercato di comprendere è come l'evoluzione della struttura abbia influenzato il comportamento delle imprese e la loro *performance* sui mercati e, implicitamente quindi, la stessa evoluzione settoriale. Il riferimento teorico implicito va chiaramente al modello struttura-comportamento-*performance* (Scherer, Ross, 1990) citato anche nel titolo di questo saggio, nella consapevolezza però che tale modello non è sempre in grado di spiegare il comportamento delle imprese, e di conseguenza i loro risultati sul mercato, dal momento che spesso fattori istituzionali esogeni ne guidano o obbligano l'azione.

Il saggio illustra alcuni risultati preliminari che concernono il settore metallurgico, uno dei comparti industriali che, per via della sua rilevanza strategica sotto il profilo economico e militare, ha da sempre rappresentato nel corso del Novecento un buon esempio di quanto detto sopra: un settore in cui il modello struttura-comportamento-*performance* trova ampie difficoltà di applicazione. Si tratta infatti di un settore in cui le politiche pubbliche di matrice nazionale e, almeno per quanto riguarda il caso europeo, internazionale hanno da sempre rivestito un ruolo fondamentale nel determinare le opportunità di investimento, le scelte degli attori coinvolti e, in fin dei conti, la struttura stessa del mercato.

Di seguito vengono presentate le prime elaborazioni che è stato possibile svolgere sulla banca dati costruita nei mesi passati. Dette elaborazioni, pur scontando alcuni limiti che il *dataset* ancora presenta e di cui si dirà più diffusamente in seguito, permettono di tracciare un quadro abbastanza chiaro dell'evoluzione del settore metallurgico nel nostro paese e di confermare alcune delle conclusioni cui è arrivata la letteratura che si è occupata di questo comparto e in particolare del processo di ridefinizione della sua struttura a seguito delle privatizzazioni degli anni novanta.

Il saggio è suddiviso in quattro paragrafi: il primo si sofferma sinteticamente sulle dinamiche nazionali e internazionali del settore metallurgico, e di quello siderurgico in particolare (dal momento che è quest'ultimo quello maggiormente rappresentato entro il campione Mediobanca); il secondo si concentra sul principale *shock* istituzionale che ha caratterizzato la storia del comparto nel periodo considerato, ossia il processo di privatizzazione avviato in concomitanza con la crisi monetaria del 1992; il terzo presenta i risultati dell'analisi quantitativa; il quarto infine trae alcune conclusioni dall'analisi, ne mostra i principali limiti e traccia un'agenda per gli approfondimenti futuri.

## 2. Una visione d'insieme

### 2.1. L'evoluzione del mercato internazionale

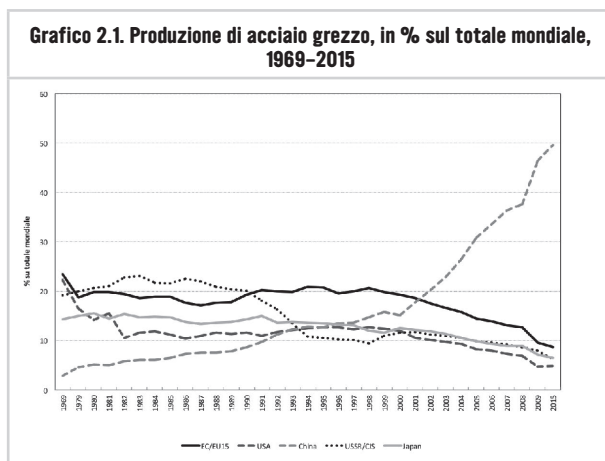
Il settore metallurgico – e in particolare all'interno di questo il comparto siderurgico – tende per sua stessa natura verso la concentrazione. Se si considerano le caratteristiche di un settore industriale che secondo la letteratura dovrebbero favorire la strutturazione del mercato intorno a un ristretto nucleo di imprese di grandi dimensioni (Scherer, Ross, 1990, pp. 97-140) queste sono tutte presenti nel caso considerato: le economie di scala e di scopo che, specie a livello di impianto, è possibile realizzare, così come i vantaggi, in termini di compressione dei costi e di reperimento delle necessarie risorse finanziarie, offerte dalla larga scala delle operazioni. Ciò è particolarmente vero per quanto concerne la cosiddetta siderurgia a ciclo integrale, quella cioè che porta alla produzione di prodotti in acciaio facendo uso di altoforni, che è il modello verso il quale si è orientato gran parte del settore a livello europeo a partire dalla fine della seconda guerra mondiale (Balconi, 1991, pp. 43-78). A fare della concentrazione una caratteristica del settore hanno contribuito tra gli anni cinquanta e gli anni settanta non solo ragioni di ordine tecnologico, che hanno progressivamente fatto crescere la dimensione media degli impianti (Cockerill, 1974), ma anche scelte di tipo politico.

Lo sviluppo di un'industria metallurgica e siderurgica ha infatti evidenti implicazioni strategiche sotto il profilo militare. Proprio in riferimento a queste implicazioni è avvenuta a livello europeo la riorganizzazione del settore negli anni della guerra fredda: la prima e più importante istituzione creata all'inizio degli anni cinquanta per coordinare il consumo di materie prime e la distribuzione dei prodotti siderurgici, la Comunità europea per il carbone e l'acciaio (Ceca), trovava le sue ragioni d'essere nella volontà statunitense di integrare i mercati europei in funzione antisovietica, incentivare la costruzione di un meccanismo di difesa europeo, stemperando le tensioni esistenti tra i vari paesi del continente, e in particolare tra Francia e Germania. Proprio la possibilità di accedere alla ricca dotazione di risorse tedesche da parte degli altri paesi europei sarà il nucleo centrale dell'accordo che avvierà nel 1951 il processo di integrazione economica continentale (Gillingham, 2004; Berend, 2016).

In questo contesto un ruolo molto importante, specialmente in paesi come Francia, Regno Unito (almeno fino alla fine degli anni sessanta) e Italia lo hanno avuto le fonti di finanziamento pubbliche, attraverso incentivi settoriali o l'azione diretta delle industrie di stato. Ciò è particolar-

mente evidente nel caso italiano dove le imprese a partecipazione statale hanno ricoperto una funzione centrale nella costruzione della siderurgia a ciclo integrale fino dai suoi esordi (Bonelli, 1982; Balconi, 1991; Ranieri, 2015).

A partire dall'inizio degli anni settanta, il settore siderurgico ha attraversato a livello mondiale una profonda ristrutturazione che ha portato alla diffusione di acciaierie di dimensioni relativamente più contenute, a seguito della progressiva affermazione dei processi di conversione ad ossigeno puro, che hanno permesso di ridurre la scala della dimensione ottima minima degli impianti, ma soprattutto a causa della ridefinizione del peso relativo dei vari paesi produttori. Come si può vedere nella tabella 2.1, tra il 1969 e il 1979 Europa e Stati Uniti hanno cominciato a perdere posizioni nei confronti della concorrenza giapponese e delle accresciute capacità produttive sovietiche, ma è soprattutto nei decenni successivi che la fotografia è andata modificandosi in maniera radicale.



Fonte: Worldsteel Association, *Steel statistical yearbook*, ad annos.

Ciò che emerge piuttosto chiaramente è uno spostamento della leadership nella produzione dai paesi di “vecchia” industrializzazione verso i paesi emergenti, le cui quote di mercato divengono via via più consistenti a fronte dei fenomeni di deindustrializzazione nei paesi sviluppati. In particolare Europa e Stati Uniti sperimentano fin da subito una rapida diminuzione della loro quota di mercato, che va accentuandosi, dopo una



leggera ripresa nel corso degli anni novanta, dalla fine del secolo in poi. Il Giappone, che nel corso dei decenni precedenti era stato il grande concorrente dei paesi di prima industrializzazione, attraversa anch'esso un declino, che si accentua dall'inizio del nuovo secolo. Anche il terzo grande produttore, la Russia, mostra un trend del tutto simile, con una caduta che ne dimezza la quota sulla produzione mondiale all'inizio degli anni novanta, in corrispondenza della crisi politico-istituzionale che pose fine all'Urss e si accompagnò a gravissime difficoltà economiche, alla caduta della produzione e del reddito. I nuovi produttori, in particolare Cina, ma anche India, Brasile e Corea del Sud, erosero via via sempre maggiori quote del mercato mondiale a partire dagli anni novanta, con una forte accentuazione nella crescita della propria quota di mercato dall'inizio del nuovo secolo. Come si può vedere dal grafico, attualmente la Cina è il maggior produttore mondiale di acciaio con una quota di mercato che si attesta intorno al 50%.

Questi sviluppi hanno inevitabilmente cambiato la siderurgia mondiale e le strategie dei paesi produttori, così come delle singole imprese. Tuttavia, come è stato messo in evidenza da alcuni studi, nonostante oggi la competizione sul prezzo sia una caratteristica importante del settore è altrettanto vero che pochi governi hanno lasciato che il destino di questo comparto fosse determinato dalle sole forze del mercato. La struttura oligopolistica della metallurgia e della siderurgia, le economie di scala in queste realizzabili e le barriere all'entrata di origine politica hanno in un certo senso fatto sì che la competizione sul prezzo fosse una conseguenza degli investimenti pregressi in tecnologia. Gli stati hanno perseguito delle politiche di investimento nel settore metallurgico e siderurgico e, una volta raggiunti sufficienti livelli di efficienza, le imprese che si sono avvantaggiate di queste politiche hanno iniziato a competere a livello internazionale. Soprattutto, la dimensione strategica del settore ha stimolato una sempre più elevata concentrazione degli investimenti effettuati dai paesi a industrializzazione tardiva (D'Costa, 1999).

Se i cambiamenti introdotti nell'ambito della tecnologica e dell'organizzazione delle filiere produttive spiegano in parte la nuova geografia della produzione di acciaio, le determinanti del cambiamento sembrano quindi da ricercarsi anche – se non soprattutto – nelle diverse scelte strategiche in tema di specializzazione settoriale prese dai vari paesi coinvolti. In altre parole, la ristrutturazione cui il settore è andato incontro a partire dalla fine degli anni settanta è stata il risultato di deliberate strategie di investimento e innovazione.

**Grafico 2.2. Produzione di acciaio grezzo, in % sul totale mondiale e percentuale italiana sul totale EC/EU15, 1969–2015**



**Fonte:** Worldsteel Association, *Steel statistical yearbook*, ad annos.

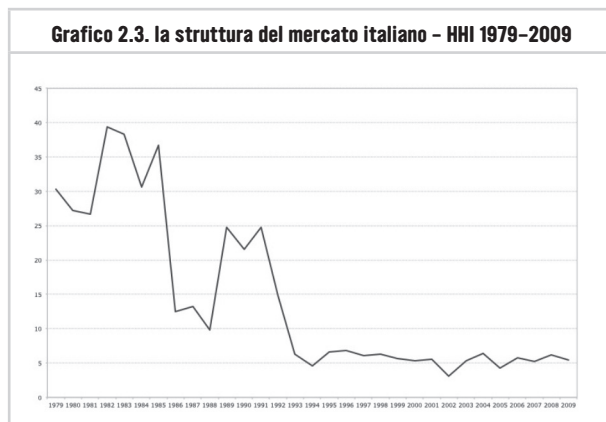
Si consideri in proposito il Grafico 2.2, che mostra le produzioni di acciaio grezzo di Europa, Italia e Corea del Sud in percentuale sul totale mondiale, nonché la produzione italiana in percentuale su quella europea tra il 1969 e il 2015. Come si può notare, vi è stato un progressivo calo delle produzioni europee e in questo contesto l'Italia si è collocata nel più ampio *trend* comunitario: la sua quota di produzione si è ridotta in percentuale sul totale mondiale, ma si è mantenuta su percentuali relativamente stabili in relazione a quella europea. Dopo un forte incremento della rilevanza della produzione italiana in relazione a quella continentale tra il 1969 e il 1979, a seguito dei grandi investimenti pubblici nel Sud Italia, in particolare con il raddoppio del centro siderurgico Iri di Taranto (Balconi, 1991, pp. 223-250), le produzioni italiane si attestano per tutti gli anni successivi su percentuali che oscillano tra il 16 e il 18% di quelle europee, mentre queste perdono progressivamente terreno sul totale mondiale. Una tendenza quest'ultima che si spiega con le scelte di politica industriale delle autorità comunitarie che, per fronteggiare la crisi da sovrapproduzione e la concorrenza internazionale che hanno investito il settore a partire dalla metà degli anni settanta, hanno avviato una progressiva e concordata riduzione delle capacità produttive. Alternando un insieme di sostegno ai prezzi, protezione dagli eccessi della concorrenza internazionale, incentivi alla riconversione produttiva, e subordinando la possibilità di concedere aiuti pubblici alle imprese in crisi al taglio delle produzioni, le autorità comuni-

tarie preposte al settore siderurgico ne hanno stimolato il ridimensionamento, con l'obiettivo di adeguare i livelli produttivi comunitari a quelli di consumo, nella convinzione che il ruolo strategico del settore fosse, se non venuto meno, almeno diminuito, data l'ormai raggiunta maturità dei sistemi industriali europei (Balconi, 1991, pp. 185-214 e 361-399).

Al contrario invece, in altri paesi che vivevano fasi differenti del processo di sviluppo, come è ad esempio il caso della Corea del Sud, già negli anni settanta la possibilità di entrare nel novero dei paesi produttori di acciaio attivi a livello internazionale fu vista come un'opportunità per sostenere un processo di sviluppo *export-led* alla vigilia della definitiva crisi dei modelli di crescita basati sulla sostituzione delle importazioni, che sarebbero stati di lì a poco resi impraticabili dal generale innalzamento dei tassi di interesse seguito al cosiddetto *Volker Shock* del 1979 (Frieden, 2006, pp. 413-434). Nel corso degli anni settanta il governo sud-coreano lanciò un ampio programma di investimento e sviluppo a favore dei settori ad alta intensità di capitale, l'*Heavy Chemical and Industry Drive*, tra i quali figurava tra gli altri anche la siderurgia, vista come un utile supporto allo sviluppo di altri comparti manifatturieri, attraverso la produzione di beni intermedi, ma anche come uno strumento che avrebbe permesso di incrementare la quota di esportazioni del paese (Lane, 2017).

## 2.2. L'evoluzione del mercato italiano

In Italia, dalla fine della seconda guerra mondiale fino all'inizio degli anni ottanta, il settore metallurgico – e nell'ambito di questo in particolare il comparto siderurgico – è stato caratterizzato da una netta preminenza della grande impresa pubblica facente capo all'Iri. Nonostante questo, i livelli di concentrazione registrati sono meno elevati di quelli che nel medesimo arco temporale è possibile ravvisare in altri paesi europei, per via della persistente importanza della siderurgia da rottame (Cockerill, 1977, pp. 47-48) che, a partire dagli anni sessanta, facendo uso di forni elettrici e di laminatoi per prodotti lunghi, ha dato vita al fenomeno delle cosiddette miniacciaierie: gruppi e imprese che, specializzati inizialmente su prodotti siderurgici a minor valore aggiunto, nei decenni successivi hanno fatto della flessibilità e dell'innovazione il loro punto di forza, diventando in alcuni casi importanti realtà produttive a livello internazionale (Balconi, 1991, pp. 155-170, 341-360 e 513-542).



Fonte: elaborazioni da *Le principali società italiane*

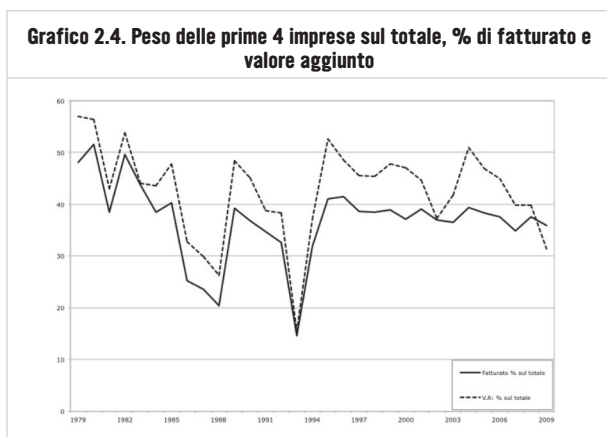
Il Grafico 2.3 riporta l'indice di *Herfindahl-Hirschman* per il periodo 1979-2009 e, in linea con quanto detto sopra, la concentrazione settoriale risulta piuttosto elevata sino ai primi anni novanta, pur in presenza di un trend discendente nel corso dei dieci anni precedenti. La significativa caduta nell'indice tra il 1985 e il 1987, seguita da un'altrettanto significativa ripresa, può essere fatta probabilmente risalire ai diversi interventi di riorganizzazione e ristrutturazione aziendale e di gruppo che interessarono la Finsider nel corso del decennio (Ranieri, 2015). La forte caduta dell'indice e il suo assestamento a livelli intorno al 5% a partire dal 1993 è invece da mettere in relazione con il processo di privatizzazione e ristrutturazione dell'intero settore avviato nel 1992 (Bemporad, Reviglio, 2001; Tonon, 2000; Artoni, 2014).

La "polverizzazione" del mercato, con la scomparsa dei grandi gruppi pubblici dalla metà degli anni novanta e il ridimensionamento del settore rispetto al mercato mondiale, sembra trovare riscontro anche nei dati relativi all'incidenza delle importazioni. Se nel periodo 1969-1990 il rapporto tra importazioni e consumi apparenti di prodotti metallurgici oscillò intorno al 35%, a partire dal 1993-94 l'indice mostra un *trend* crescente, che si attesta oltre il 40% dal 1998 in avanti per raggiungere valori intorno al 55% negli ultimi anni considerati<sup>1</sup>.

<sup>1</sup> Elaborazioni su OECD, STAN Database on Structural Analysis (dati estratti il 29/12/2016).

A fronte di una riduzione della concentrazione complessiva, comunque, il settore rimase dominato dalle maggiori imprese, come emerge chiaramente dal Grafico 2.4. Piuttosto interessante è il confronto tra questo e il grafico precedente: nella prima parte del periodo, sino al 1993, gli indici dei due grafici si muovono sostanzialmente in parallelo, a ulteriore conferma della estrema rilevanza dell'impresa pubblica nel settore. Nella seconda metà del periodo, conclusosi il processo di privatizzazione della siderurgia Iri, la quota di mercato delle prime quattro maggiori imprese si assesta, in termini di fatturato, intorno al 38-40%, su livelli più bassi, ma non molto discosti da quelli fatti registrare dalle maggiori imprese pubbliche alla fine degli anni settanta, testimoniando come, in un settore che ha ridotto le sue dimensioni, siano comunque le grandi imprese a rivestire ancora un ruolo preminente.

Il dato relativo al valore aggiunto, inoltre, non sembra mostrare una dinamica diversa nell'ultima parte del periodo rispetto alla fine degli anni settanta. Sembra infatti che il rapporto tra le quote di mercato espresse in termini di valore aggiunto e di fatturato delle prime quattro imprese – fatto salvo il periodo turbolento 1981-1994, caratterizzato da crisi, ristrutturazioni e privatizzazioni – sia rimasto lo stesso tra inizio e fine periodo, cosa che suggerirebbe una certa difficoltà delle imprese di più grandi dimensioni, che sono anche quelle che hanno maggiormente beneficiato dei processi di privatizzazione, nell'accrescere o mantenere la propria efficienza complessiva (vedi oltre).



Fonte: elaborazioni da *Le principali società italiane*

### 3. Le privatizzazioni

Nella storia recente della metallurgia italiana uno spartiacque fondamentale sono stati indubbiamente i primi anni novanta, quando fu avviato il complesso processo di privatizzazione delle imprese a partecipazione statale nell'ambito della crisi finanziaria del 1992 (Ciocca, 2007, pp. 307-3015; Rossi, 2007, pp. 74-94; Amato, Graziosi, 2013, pp. 141-226; Devillanova, 2014; Craveri, 2016). Una parte rilevante di questo processo ha infatti riguardato la siderurgia pubblica, che aveva fatto capo fino alla fine degli anni ottanta alla Finsider, la *sub-holding* di settore dell'Iri, messa in liquidazione volontaria nel febbraio del 1988, cedendo gran parte delle sue partecipazioni alla ricostituita Ilva, l'azienda che sarebbe stata messa sul mercato di lì a poco (Ranieri 2015).

Il processo di privatizzazione fu innescato dalla crisi finanziaria dei primi anni novanta (Goldstein, Nicoletti 1996; De Nardis 2000; Artoni 2014), ma la sua genesi risale agli anni precedenti ed è da ricercare nella lunga fase di difficoltà delle partecipazioni statali apertasi con il rallentamento dell'economia mondiale registratosi negli anni settanta e continuata poi con i tentativi di riorganizzazione dell'industria pubblica degli anni ottanta, specialmente di quella che faceva capo all'Iri (Silva 2013). Per quello che riguarda in specifico l'industria siderurgica, bisogna poi tenere conto del fatto che le privatizzazioni avvennero nel quadro della politica di ristrutturazione del comparto voluta dalla Comunità Europea che, a partire anche in questo caso dalla seconda metà degli anni settanta, cominciò a imporre prezzi minimi per alcune categorie di prodotto e a favorire le ristrutturazioni e le riconversioni per contenere la capacità produttiva (Balconi, 1991).

Alla fine del 1993 le spinte alla ristrutturazione di matrice europea si concretizzarono nell'accordo siglato dal ministro degli Esteri italiano, Beniamino Andreatta, e il commissario europeo alla Concorrenza, Karel Van Miert, con il quale veniva concessa all'Iri la facoltà di rispondere in pieno dei debiti pregressi contratti dalle sue controllate fino al dicembre del 1993; e dal successivo accordo tra il ministro dell'Industria, Paolo Savona, e Van Miert, attraverso il quale il progetto di ristrutturazione del settore siderurgico messo a punto dall'Iri nei mesi precedenti veniva accettato dall'Unione, a patto che entro il 1994 venisse realizzato un consistente numero di dismissioni (Curli, 2014).

A seguito degli accordi tra il governo italiano e l'Unione europea, dunque, la privatizzazione dell'industria siderurgica italiana fu avviata: nel 1994 fu liquidata la Nuova Ilva e di lì a poco le principali imprese siderurgiche a controllo pubblico sarebbero state cedute a gruppi industriali privati. Tra il

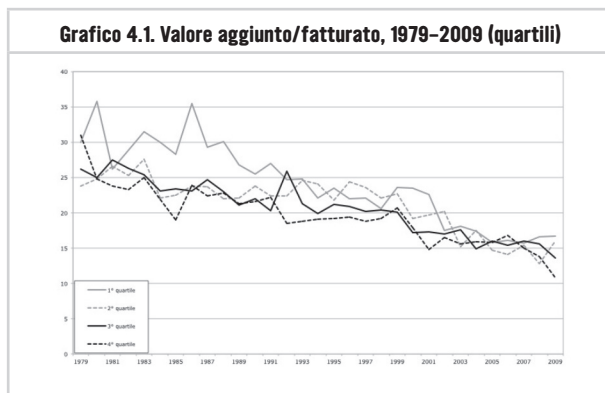
1994 e il 1995 la Acciai speciali Terni e l'Ilva laminati piani furono vendute rispettivamente al gruppo tedesco Thyssen-Krupp e al gruppo Riva. Nel 1996, poi, Techint acquistò Dalmine, mentre si procedeva alla cessione delle altre imprese di più piccole dimensioni rimaste ancora sotto il controllo pubblico (Ranieri, 2015).

## 4. Analisi 1979–2009

Come accennato più sopra, il lavoro utilizza un *dataset* nel quale sono riportati i dati di bilancio – stato patrimoniale e conto economico – delle maggiori imprese del settore metallurgico dal 1979 al 2009, cioè relativi al trentennio che partendo dalla crisi del settore attraversa gli anni della sua ristrutturazione e privatizzazione, fino a giungere all'inizio del nuovo secolo. Per l'analisi dei cambiamenti avvenuti nella struttura, nel comportamento e nella *performance* delle imprese metallurgiche in questo trentennio sono stati individuati quattro indicatori: due di efficienza o *performance* (valore aggiunto su fatturato; fatturato su costi di esercizio); uno di solvibilità o solidità patrimoniale/leva finanziaria (indebitamento su attivo), uno infine di redditività (margine operativo lordo su attivo). I quattro indici di bilancio sono filtrati sia per quartile di fatturato, al fine di valutare se si siano verificati comportamenti differenti in relazione alla dimensione d'impresa, sia per assetto proprietario (proprietà pubblica o privata), dal momento che è proprio sotto questo profilo che sono avvenuti i maggiori cambiamenti nel periodo considerato. A questo proposito si tenga presente che per l'intero periodo di esistenza dell'impresa pubblica entro il campione Mediobanca, e cioè fino al 2000, le imprese pubbliche sono per lo più imprese di medio-grandi dimensioni, mentre è possibile osservare una più equa distribuzione tra le classi di fatturato da parte di quelle private.

I Grafici 4.1. e 4.2. mostrano l'andamento dell'indice del valore aggiunto sul fatturato rispettivamente per quartili di dimensione e per assetto proprietario<sup>2</sup>.

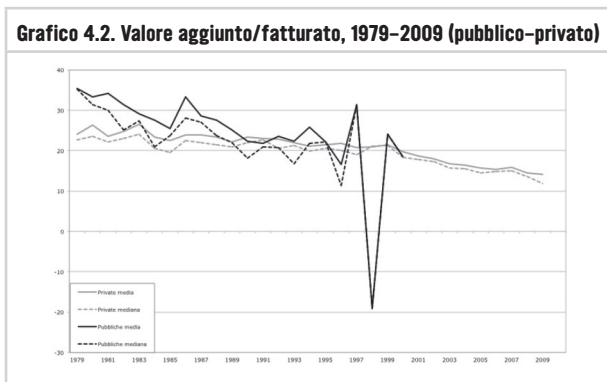
<sup>2</sup> Sebbene solitamente si utilizzi il rapporto tra valore aggiunto e fatturato come una misura di integrazione verticale, data la scarsa – se non nulla – incidenza dell'*outsourcing* nel settore considerato, si è ritenuto di poter utilizzare questo indice come indicatore di efficienza insieme a quello relativo al rapporto tra fatturato e costi di esercizio.



**Fonte:** elaborazioni da *Le principali società italiane*

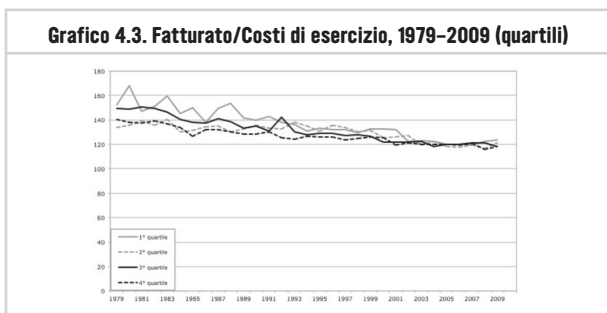
Ciò che emerge chiaramente dal primo grafico è il *trend* discendente dell'indice nel corso del periodo e relativamente all'intero settore, pur con oscillazioni anche marcate, soprattutto nella prima metà del periodo, e differenze tra i diversi gruppi dimensionali. A tale proposito è evidente la miglior *performance* – sino alla fine del secolo – delle imprese di minori dimensioni, rispetto a quelle maggiori. Ciò sembra evidenziare una incongruenza con quanto emerge dal Grafico 4.2. relativamente alle imprese di proprietà pubblica, che sino alla fine degli anni ottanta presentano un andamento dell'indice più positivo di quelle private, nonostante il loro gruppo sia composto da imprese di grandi dimensioni. La discordanza tra i due grafici trova probabilmente spiegazione nella maggiore varianza esistente tra i livelli di *performance* delle imprese pubbliche, del concentrarsi dei loro risultati negativi in alcune imprese specifiche. Quest'ipotesi, suggerita anche dalla maggiore distanza tra media e mediana nel caso delle imprese pubbliche potrà essere tuttavia confermata solo in seguito ad ulteriori approfondimenti e alla soluzione dei due principali problemi che ancora il *dataset* presenta e di cui si dirà più avanti.





Fonte: elaborazioni da *Le principali società italiane*

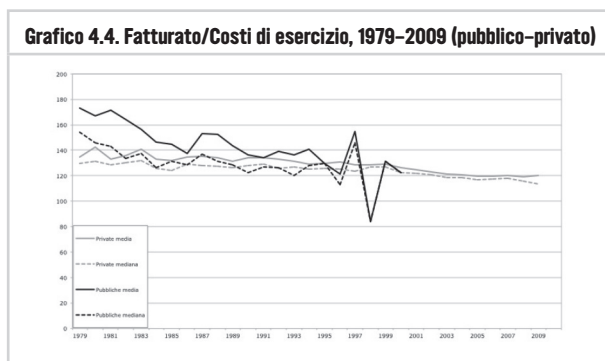
In termini generali, le imprese pubbliche, pur assottigliando il proprio vantaggio rispetto a quelle private, si mantennero su livelli più elevati di queste sino quasi al termine del processo di privatizzazione. Le oscillazioni che si osservano nell'indice tra 1996 e 1999, così come la forte caduta del 1998, dipendono essenzialmente dalle code del processo di ristrutturazione e liquidazione delle ultime due/tre imprese rimaste in mano pubblica e in sostanziale dissesto. Il picco negativo del 1999, infine, è dovuto alla Società Bagnoli, l'ultima impresa siderurgica ancora di proprietà pubblica, che quell'anno presentava un fatturato negativo. L'altro dato che sembra emergere abbastanza chiaramente è il fatto che il processo di ristrutturazione e di privatizzazione degli anni novanta non sembra aver portato a una maggiore efficienza del settore: appare infatti evidente una convergenza verso livelli molto bassi dell'indice da parte di tutti i quartili dimensionali.



Fonte: elaborazioni da *Le principali società italiane*

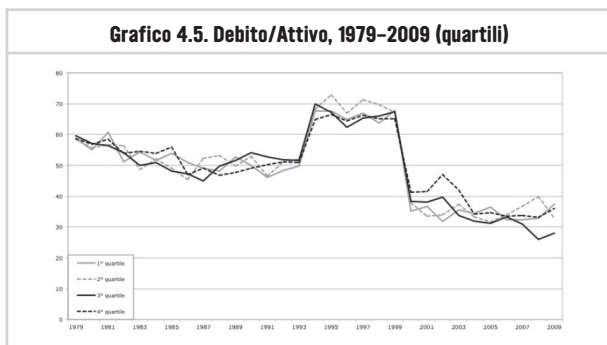
Non molto differente appare il quadro se osservato attraverso l'andamento del rapporto tra fatturato e costi di esercizio, ossia analizzando la capacità delle imprese di produrre dei margini una volta sostenuti i costi delle produzioni. Nel Grafico 4.3., sia le imprese di minori dimensioni sia quelle medio-grandi (3° quartile) mostrano – sino all'inizio degli anni novanta – una *performance* in termini di efficienza maggiore rispetto a quelle medio-piccole e alle maggiori. È questa forse la differenza più rilevante tra questo grafico e il Grafico 4.1. Tuttavia il *trend* generale, di tutte le classi dimensionali, è anche in questo caso discendente. La convergenza tra i quartili risulta inoltre in questo caso massima: dall'inizio del nuovo secolo tutti i quartili si assestano sullo stesso livello.

Differenze interessanti emergono anche dall'andamento del rapporto tra fatturato e costi di esercizio tra le imprese di proprietà pubblica e quelle private (Grafico 4.4). Sino all'avvio delle privatizzazioni, nel 1994, le imprese pubbliche – pur avendo sofferto una diminuzione nella propria capacità di coprire adeguatamente i costi operativi nel corso del decennio precedente – si posizionano ancora al di sopra di quelle private. Queste ultime, infatti, mostrano un andamento dell'indice sostanzialmente stabile, ma piuttosto “mediocre”, se comparato a quello ben superiore delle imprese pubbliche. Terminata la fase di ristrutturazione del settore, dalla fine degli anni novanta, l'indice anziché riprendersi, si assesta su un livello medio che – benché positivo – appare decisamente basso. Tale fenomeno è probabilmente il frutto di una concomitanza di fattori, sintetizzabili in quel cambiamento strutturale a livello di industria cui s'è fatto cenno più sopra. Sembrerebbe insomma che le imprese italiane siano state meno pronte rispetto ai produttori dei paesi emergenti nel fare fronte ai cambiamenti istituzionali – tra cui vanno annoverate le liberalizzazioni e le privatizzazioni – e ai processi di innovazione tecnologica, come ad esempio il rapido diffondersi e ammodernarsi dei cosiddetti *mini-mills*.



Fonte: elaborazioni da *Le principali società italiane*

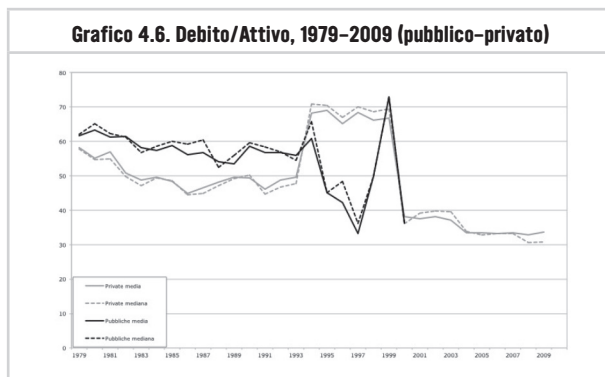
Venendo ora a considerare il tasso di indebitamento delle imprese metallurgiche, bisogna rilevare come questo fosse piuttosto elevato alla fine degli anni settanta (60%), presenti nel corso decennio successivo un *trend* discendente sino al 1988-89, per attestarsi quindi alla vigilia delle privatizzazioni intorno al 50%. Nello stesso sotto-periodo le differenze tra quartili evidenziano una migliore *performance* delle imprese minori e delle medio-grandi (Grafico 4.5).



Fonte: elaborazioni da *Le principali società italiane*

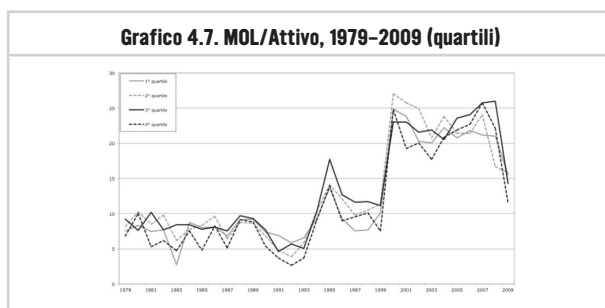
Con l'avvio del processo di privatizzazione, l'indice balza a livelli molto elevati – i più alti del periodo – restandovi sino a fine decennio. Il significativo aumento della leva finanziaria è la diretta conseguenza dell'acquisizione della siderurgia pubblica da parte di gruppi privati finanziata attraverso l'indebitamento. Come emerge chiaramente dal confronto tra i grafici 4.5 e 4.6, infatti, il tasso di indebitamento delle imprese pubbliche, in calo nel corso degli anni ottanta, crolla bruscamente durante i primi anni novanta, a fronte di una crescita considerevole di quello delle imprese private. Forse ancor più rilevante è il fatto che mentre sino ai primi anni novanta per le imprese private la mediana sia inferiore o al più coincida con la media, dal 1993 sino al 2004-2005 circa, metà delle imprese private presenti un tasso di indebitamento superiore a quello medio, una situazione che ricalca appieno quella delle imprese pubbliche per quasi l'intero periodo, e che rientra solo negli ultimi anni considerati.

Si può infine notare come, tra il 1999 e il 2000 si registri una brusca caduta del tasso di indebitamento, seguita da un *trend* calante per il resto del periodo, che porta l'indice ad assestarsi a livelli pari a circa la metà di quelli raggiunti a fine anni settanta.

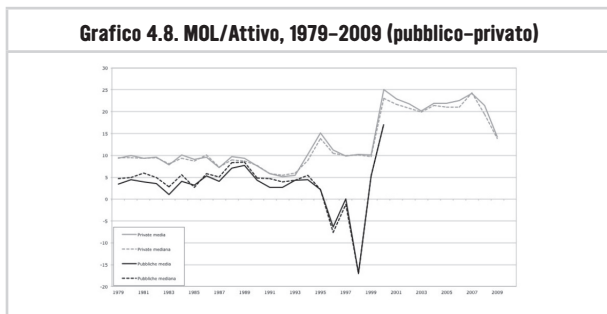


**Fonte:** elaborazioni da *Le principali società italiane*

Il calo è molto probabilmente legato al miglioramento del quadro economico nella seconda metà degli anni novanta e alla capacità delle imprese di raggiungere in questo contesto più elevati livelli di redditività, come suggeriscono i grafici 4.7 e 4.8, che mostrano il rapporto tra MOL e totale attivo. A partire dall'avvio dei processi di privatizzazione questo indicatore, che ha visto le imprese pubbliche registrare *performance* peggiori rispetto a quelle private per tutto il periodo considerato, è infatti andato decisamente migliorando, a testimonianza del fatto che, nell'ambito di un generale *trend* positivo di tutto il comparto, le ristrutturazioni, prima, e la messa sul mercato dell'industria pubblica, poi, sembrano aver generato un miglioramento della redditività delle imprese acquisite.



**Fonte:** elaborazioni da *Le principali società italiane*



Fonte: elaborazioni da *Le principali società italiane*

Vi potrebbe essere tuttavia un'altra spiegazione dell'andamento del rapporto tra MOL e totale attivo: la sua crescita potrebbe essere determinata non da un incremento del MOL, bensì da una diminuzione del totale attivo derivante da eventuali svalutazioni delle immobilizzazioni seguite al processo di privatizzazione. Un'interpretazione quest'ultima che porterebbe a riconsiderare i risultati dell'intero processo in termini di incremento della redditività delle imprese e che sembrerebbe maggiormente in linea con quanto suggeriscono gli indicatori di efficienza di cui si è parlato sopra e anche con l'analisi dei livelli di autofinanziamento, che per tutto il periodo considerato permangono prossimi allo zero. Un'ipotesi, tuttavia, che per essere verificata necessita di alcuni ulteriori affinamenti al *dataset* che sono ancora in corso e di cui si dirà brevemente nel paragrafo conclusivo

## 5. Conclusioni e agenda

L'analisi quantitativa dell'evoluzione del settore metallurgico italiano permette di trarre alcune conclusioni preliminari su come il settore abbia risposto ai cambiamenti tecnologici, economici e istituzionali degli ultimi decenni. Bisogna innanzitutto considerare il fatto che questa risposta, almeno per quanto concerne il sotto-settore siderurgico, si pone entro il più ampio quadro della crisi europea del comparto. Come si è visto, gli andamenti italiani, pur nella loro peculiarità, rispecchiano e risentono di alcune scelte operate a livello europeo. In secondo luogo l'analisi ha evidenziato il ruolo fondamentale giocato, nel tracciare le linee di sviluppo del settore, dal processo di privatizzazione innescato dalla crisi finanziaria del 1992. A tale proposito sembrerebbe possi-

bile concludere che esso abbia permesso un moderato recupero di redditività e abbia garantito la sopravvivenza delle componenti maggiormente dinamiche del comparto. Allo stesso tempo la scarsa solidità finanziaria degli acquirenti, che nella maggior parte dei casi hanno utilizzato la leva finanziaria per portare a termine l'acquisizione delle imprese pubbliche collocate sul mercato, ha prodotto una limitata capacità di autofinanziamento, che si è tradotta nell'impossibilità – o nella mancanza di volontà – di produrre nuovi investimenti, con effetti negativi sull'efficienza delle imprese. In questo senso il lavoro di analisi sembrerebbe confermare le ambiguità del processo di privatizzazione, in termini di miglioramento della redditività delle imprese e di vantaggi per il consumatore finale, che sono state messe in luce nell'analisi retrospettiva eseguita dalla Corte dei conti a distanza di circa un quindicennio dall'avvio delle prime dismissioni (Corte dei conti 2010).

Per poter trovare conferma questa lettura necessita tuttavia di ulteriori approfondimenti e in particolare del completamento qualitativo del *dataset*, che è al momento ancora in fase di perfezionamento. La pubblicazione di Mediobanca su cui il *dataset* si basa non riporta infatti un codice univoco per le imprese, allo stato attuale non è stato pertanto possibile seguire la storia delle singole aziende e dei singoli gruppi, né riporta una chiara classificazione dei settori. Tutte le imprese rientrano in un generico settore metallurgico: in larga parte, stanti le soglie di fatturato che hanno escluso dall'annuario solo le imprese di più piccole dimensioni, esse fanno parte del comparto siderurgico, ma questo non è sempre vero, con evidenti effetti distorsivi.

La creazione di un'identificazione univoca per le imprese e la migliore specificazione dei settori di appartenenza, attraverso l'utilizzo di classificazioni standard (Ateco o Isic) permetterà un'analisi più approfondita del comportamento delle imprese e la possibilità di meglio misurare la loro efficienza e le loro *performance* reddituali, dando maggiore robustezza alle ipotesi qui avanzate.

## Riferimenti Bibliografici

- Amato, G., Graziosi, A. (2013), *Grandi illusioni. Ragionando sull'Italia*, il Mulino.
- Artoni, R. (2014), *Storia dell'IRI. 4. Crisi e privatizzazione*, Laterza.
- Balconi, M. (1991), *La siderurgia italiana (1945-1990). Tra controllo pubblico e incentivi del mercato*. Il Mulino.
- Berend, I.T. (2016), *The History of European Integration: A New Perspective*, Taylor and Francis.

- Bonelli, F. (1982), *Acciaio per l'industrializzazione. Contributi allo studio del problema siderurgico italiano*, Einaudi.
- Cockerill, A. (1974), *The Steel Industry. International Comparisons of Industrial Structure and Performance*, Cambridge University Press.
- Ciocca, P. (2007), *Ricchi per sempre? Una storia economica d'Italia (1796-2005)*, Bollati Boringhieri.
- Corte dei conti (2010), *Obiettivi e risultati delle operazioni di privatizzazione di partecipazioni pubbliche*, documento reperibile al seguente indirizzo: [http://www.corteconti.it/export/sites/portalecdc/\\_documenti/controllo/sez\\_centrale\\_controllo\\_amm\\_stato/2010/delibera\\_3\\_2010\\_g\\_relazione.pdf](http://www.corteconti.it/export/sites/portalecdc/_documenti/controllo/sez_centrale_controllo_amm_stato/2010/delibera_3_2010_g_relazione.pdf)
- Curli, B. (2014), *Il «vincolo europeo»: le privatizzazioni dell'IRI tra Commissione europea e governo italiano*, in *Storia dell'IRI. 4. Crisi e privatizzazione* a cura di R. Artoni, Laterza, pp. 181-256.
- Craveri, P. (2016), *L'arte del non governo. L'inesorabile declino della Repubblica italiana*, Marsilio.
- De Nardis, S., Traù, F. (2005), *Il modello che non c'era. L'Italia e la divisione internazionale del lavoro industriale*, Rubettino.
- De Nardis, S. (2000), *Le privatizzazioni italiane*, il Mulino.
- Devillanova, C. (2014), *Il quadro macroeconomico negli anni Novanta*, in *Storia dell'IRI. 4. Crisi e privatizzazione* a cura di R. Artoni, Laterza, pp. 39-58.
- Frieden, J. A. (2006), *Global Capitalism. Its Fall and Rise in the Twentieth Century*, W. W. Norton.
- Gillingham, J. (2004), *Coal, Steel, and the Rebirth of Europe, 1945-1955. The Germans and French from Ruhr Conflict to Economic Community*, Cambridge University Press.
- Goldstein, A., Nicoletti, G. (1996), *Le privatizzazioni in Italia, 1992-1995: motivi, metodi e risultati*, in *La finanza pubblica italiana dopo la svolta del 1992* a cura di A. Monorchio, il Mulino.
- Lane, N., *Manufacturing Revolutions. Industrial Policy and Networks in South Korea*, job market paper reperibile al seguente indirizzo: <http://nathanlane.info>.
- Perulli, P., Pichierri, A. (2010), *La crisi italiana nel mondo globale: economia e società del Nord*, Einaudi.
- Ranieri, R., (2015), *La siderurgia Iri dal Piano Sinigaglia alla privatizzazione*, in *Storia dell'Iri*, vol. 5, *Un gruppo singolare. settori, bilanci, presenza nell'economia italiana* a cura di F. Russolillo, Laterza, 5-199.
- Bemporad, S., Reviglio, E. (2001), *Le privatizzazioni in Italia: 1992-2000*. Edindustria.
- Rossi, S. (2007), *La politica economica italiana 1968-2007*, Laterza.
- Scherer, F. M., Ross, D. (1990), *Industrial market structure and economic performance*, Houghton Mifflin Company.
- Silva, F. (2013), *Storia dell'IRI. 3. I difficili anni '70 e i tentativi di rilancio negli anni '80*, Laterza.
- Traù, F. (1999), *La "questione dimensionale" nell'industria italiana*, il Mulino.





# Longevità e modelli di business dei social network online: un'analisi empirica

ANDREA MANGANI

## 1. Introduzione

Gli effetti di rete (o esternalità di rete) sollevano serie questioni antitrust, poiché la regola del “chi vince prende tutto” può condurre a mercati molto concentrati. In realtà, gli effetti di rete creano problemi anche per la semplice definizione del mercato e per la realizzazione di studi empirici di settore. Infatti, con effetti di rete poche imprese raggiungono una massa (critica) di utenti/consumatori che possa innescare un'espansione automatica della rete stessa. Il ricercatore può quindi osservare soltanto le caratteristiche e il comportamento delle imprese “vincenti”. Quelle “perdenti” lasciano tracce evanescenti, spesso inadeguate per l'analisi empirica.

Tale situazione è tipica nel settore dei *social networking sites* (d'ora in avanti, SNS). Data l'assenza di rilevanti barriere all'entrata nel mercato e l'apparente semplicità di creare una comunità virtuale online, dalla metà degli anni '90 molti soggetti hanno inseguito il successo economico mediante il lancio di SNS generalisti o specializzati (Albarran, 2013). Oggi conosciamo in dettaglio le caratteristiche di chi ce l'ha fatta, ma abbiamo poche informazioni su chi ha fallito: le tracce digitali dei perdenti svaniscono rapidamente. L'informazione disponibile è quindi viziata da questa asimmetria, e complica sia l'analisi empirica del settore sia le applicazioni manageriali e di politica economica.

Per risolvere almeno in parte questo problema, il presente lavoro costruisce un database dei SNS apparsi tra gennaio 1995 e marzo 2015. Il campione include 224 SNS, 181 dei quali attivi e 43 cessati. L'analisi empirica si concentra

sui fattori tecnologici e strutturali (dell'impresa e del settore) che possono influenzare la probabilità di sopravvivenza dei SNS.

Gli studi economici dei SNS sono principalmente teorici, nonostante il numero delle analisi empiriche stia crescendo velocemente. Ad esempio, alcuni autori hanno analizzato il comportamento degli utenti dei SNS (Kossinets e Watts, 2006; Lin, 2008; Shriver et al., 2013; Wilcox e Stephen, 2013; Tucker, 2014), mentre altri si sono concentrati sulla forza degli effetti di rete (Henkel e Block, 2013), sulle strategie delle imprese (Miller et al., 2009) e sulla connettività *cross-border* delle comunità virtuali (Rohn, 2014). Tuttavia, la maggior parte delle analisi empiriche consiste in studi di singoli SNS, con un'attenzione particolare ai principali soggetti del settore, come Facebook e Twitter. Nessun autore ha analizzato la questione del successo/fallimento dei SNS in aggregato. Quindi la prossima sezione si rifarà alla letteratura "tradizionale" sulla sopravvivenza delle imprese, discutendone i maggiori risultati in relazione al settore dei SNS. La terza sezione descrive il processo di raccolta dei dati e illustra le statistiche descrittive, mentre la quarta mostra i risultati dell'analisi empirica, che possono essere così riassunti. Da un punto di vista statico e dinamico, soltanto tre elementi sono associati sistematicamente alla sopravvivenza dei SNS. Primo, se un SNS si rivolge a un pubblico specializzato, la sua probabilità di sopravvivenza è maggiore. Secondo, quando un SNS è acquisito da un'impresa esterna, la probabilità di cessare le attività aumenta considerevolmente. Infine, maggiore è l'esperienza di un SNS nel mercato, maggiore la probabilità di interrompere le attività. La quinta sezione descrive alcuni sviluppi dei modelli di business che non sono stati considerati nell'analisi empirica. L'ultima sezione conclude il lavoro, descrive i suoi limiti e suggerisce alcune implicazioni di politica economica.

## **2. Sopravvivenza delle imprese: teoria, lavori empirici e applicazione ai SNS**

La letteratura economica sulla sopravvivenza delle imprese affronta da tempo lo studio dei fattori interni ed esterni alle imprese che hanno un impatto sulla loro capacità di rimanere in un mercato. Una delle caratteristiche d'impresa più discusse riguarda le dimensioni, in termini di fatturato o numero di dipendenti. In generale, i lavori empirici mostrano che le imprese più grandi hanno una maggiore probabilità di sopravvivenza (Aldrich e Auster, 1986; Mata e Portugal, 1994; Wagner, 1994; Geroski et al., 2002; Esteve-Perez e Mañez-Castillejo, 2008; Strotmann, 2007; Varum e Rocha, 2012). Nell'industria dei SNS, la dimensione d'impresa è una variabile di difficile misu-

razione. In termini di ricavi, quota di mercato o utenti, ogni SNS entra nel mercato con una dimensione pari a zero. Inoltre, il numero di dipendenti di un SNS è dato raramente disponibile. La terza e la quarta sezione, quindi, si concentreranno su altri fattori che possono influenzare la probabilità di sopravvivenza di un SNS, e in particolare: *focus* del SNS (SNS generalista oppure specializzato), possibili restrizioni alla registrazione degli utenti, metodi di finanziamento (pubblicità, servizi *premium*, ecc.), intensità concorrenziale nel paese di appartenenza, operazioni particolari (fusioni, acquisizioni, *partnership*) ed esperienza del SNS nel mercato.

I lavori empirici precedenti hanno preso in considerazione alcune strategie di impresa per valutarne l'impatto sul tasso di sopravvivenza. Ad esempio, Audretsch (1995), Esteve-Perez e Mañez-Castillejo (2008), e Cefis e Marsili (2005) sostengono che le imprese che fanno R&D hanno una maggiore probabilità di rimanere sul mercato. Anche le strategie di differenziazione possono essere rilevanti per la sopravvivenza delle imprese. Hannan e Freeman (1977), in un contributo ampiamente citato, sostengono che le imprese "generaliste" hanno la capacità di affrontare eventi imprevisti, mentre quelle specializzate fronteggiano maggiori rischi di cessazione delle attività. Tuttavia, studi più recenti mostrano che la differenziazione intra-industriale può aumentare la probabilità di rimanere nel settore (Stern e Henderson, 2005; Zahavi e Lavie, 2013). Inoltre, nonostante la domanda dei prodotti specialistici sia soggetta a maggiore variabilità, una strategia di nicchia consente una crescita dimensionale più rapida (Cressy, 2006). Anche gli studi di management enfatizzano la tensione tra servizi generalisti e specializzati (si veda, ad esempio, Deephouse, 1999). In relazione ai SNS, rivolgersi a un pubblico generalista facilita il raggiungimento di una massa critica di utenti, che assicura ricavi pubblicitari ed eventualmente anche la monetizzazione del servizio mediante pagamento diretto degli utenti (Canzer, 2006). I "grandi numeri" sono importanti per i SNS per due motivi. Primo, le esternalità dirette di rete implicano che il valore di appartenere al social network aumenta all'aumentare dei suoi componenti. Secondo, un'audience estesa aumenta la disponibilità a pagare di molti inserzionisti. Tuttavia, un SNS generalista compete con molte altre piattaforme e deve aggiustare costantemente le sue caratteristiche per soddisfare le preferenze eterogenee degli utenti. Invece un SNS specializzato si può concentrare su un gruppo più ristretto e omogeneo di utenti e occupare una nicchia di mercato a bassa intensità concorrenziale (Laudon e Traver, 2007).

Alcune particolarità del settore dei SNS possono avere importanza nell'analisi di sopravvivenza. In primo luogo, i SNS possono escludere dalla registrazio-

ne alcune categorie di utenti. Ad esempio, gli individui con meno di 13 anni non possono registrare un account su Facebook. L'effetto delle restrizioni all'accesso non è scontato. Da un lato le restrizioni riducono artificialmente il numero potenziale di utenti, così come i possibili ricavi pubblicitari. Dall'altro, le restrizioni garantiscono agli utenti un minimo livello di omogeneità e intimità, aumentando così l'utilità di far parte del social network (Subrahmanyam et al., 2008).

In relazione alle possibili fonti di ricavo, alcuni SNS offrono servizi addizionali (premium) dietro il pagamento di un prezzo. Come per le restrizioni, questa pratica può avere effetti contrastanti. I servizi premium ampliano la varietà dei contenuti, e il SNS può offrire distinti pacchetti di servizi a prezzi diversi. In presenza di potere di mercato, i SNS possono discriminare sul prezzo tra gli utenti e quindi incrementare ricavi e profitti (Wang et al., 2005; Enders et al., 2008). Tuttavia, i servizi premium possono essere percepiti come un tentativo del SNS di sfruttare la disponibilità a pagare degli utenti. I SNS potrebbero quindi perdere gli utenti con una mentalità "free" circa la fornitura di servizi e contenuti via Internet (Dou, 2004).

Considerazioni simili valgono per la vendita di spazio pubblicitario e informazioni private agli inserzionisti. La pubblicità tradizionale (come quella dei *banners*) garantisce risorse necessarie per migliorare le funzionalità del SNS e aumentare la probabilità di sopravvivenza, ma molti utenti trovano la pubblicità display particolarmente fastidiosa (Kelly et al., 2010). Certamente, l'osservazione del comportamento degli utenti all'interno delle comunità virtuali permette al social media e agli inserzionisti di personalizzare gli annunci pubblicitari. I consumatori tendono quindi a percepire le inserzioni più allineate ai propri interessi e ciò può aumentare l'efficacia della comunicazione pubblicitaria nel suo complesso (Anand e Shachar, 2009; Lambrecht e Tucker, 2013). Tuttavia, il presupposto della combinazione ottimale tra preferenze degli utenti e caratteristiche degli annunci è la cessione di informazioni riservate agli inserzionisti, un aspetto per il quale molti utenti dei social media sono seriamente preoccupati (Taylor et al., 2011; Lee, 2013).

Alcuni SNS vendono beni e servizi online direttamente agli utenti registrati. Se le attività di e-commerce hanno successo, il SNS ha un reddito addizionale per migliorare i servizi offerti e rimanere sul mercato (Buhler et al., 2015)<sup>1</sup>. Inoltre, l'e-commerce aumenta le interazioni tra gli utenti del

<sup>1</sup> Si è soliti distinguere tra transazioni endogene, che si verificano quando una piattaforma offre beni materiali o immateriali che gli utenti possono acquistare direttamente dal SNS, e transazioni esogene, riferite alla situazione in cui il SNS vende contenuti di terze parti ai suoi utenti o facilita le transazioni tra gli utenti stessi (Enders et al., 2008, p. 208).

SNS, la loro fiducia nella piattaforma e la probabilità di transazioni future (Hajli, 2014). Tuttavia, affinché la vendita di servizi online risulti profittevole il numero di transazioni deve essere elevato (Armstrong e Hagel, 2000; Zeng e Reinartz, 2003). Inoltre, alcuni utenti potrebbero percepire le attività di vendita come uno svilimento della principale missione del SNS, cioè la creazione di una comunità virtuale basata su legami personali e svincolata da logiche commerciali (Huang e Benyoucef, 2013).

Ritornando all'ampia letteratura sulla sopravvivenza delle imprese, un risultato ricorrente è la *liability of newness*: controllando per le dimensioni d'impresa, i primi anni di attività presentano un maggiore rischio di fallimento (Stinchcombe e March, 1965; Mata e Portugal, 1994), nonostante alcuni lavori suggeriscano una non linearità della relazione tra "età" dell'impresa e probabilità di sopravvivenza (Wagner, 1994; Fackler et al., 2013). L'età dell'impresa è una proxy della sua esperienza, che serve a costruire salde relazioni con fornitori e consumatori e a convincere nuovi consumatori a provare il bene o il servizio (Bruderl e Schussler, 1990). Queste considerazioni si applicano anche ai SNS. Nonostante l'idea di base di una comunità virtuale sia spesso la causa principale del suo successo, l'esperienza accumulata nel tempo può rafforzare le relazioni con gli utenti, facilitare il miglioramento del servizio e attrarre un numero crescente di utenti e inserzionisti. In effetti, alcuni lavori mostrano che la fiducia degli utenti nel social network di riferimento aumenta col trascorrere del tempo (Dwyer et al., 2007; Fogel e Nehmad, 2009). Tuttavia, altri autori segnalano la possibilità di una *liability of aging* (Barnett, 1990; Barron et al., 1994; Ranger-Moore, 1997). In breve, le imprese diventerebbero progressivamente incapaci di generare innovazioni all'aumentare della loro permanenza nell'industria. Col trascorrere del tempo, le imprese sviluppano infatti rigidità organizzative che non facilitano l'acquisizione di nuova conoscenza e nuove routine (Leonard-Barton, 1992). Invece le imprese più giovani hanno vantaggi di apprendimento poiché possono esplorare routine innovative senza dover disimparare quelle vecchie. Gli elementi che favoriscono le imprese più giovani sono, nel settore dei SNS, particolarmente importanti, perché ogni anno generazioni successive di utenti entrano nel "mercato" ed esprimono preferenze nuove ed eterogenee.

I lavori empirici sulla sopravvivenza delle imprese considerano anche fattori esterni all'impresa, come l'intensità concorrenziale del mercato o dell'area geografica di appartenenza. Quando il numero di imprese è elevato rispetto alle dimensioni complessive del mercato, la probabilità di

cessazione dell'attività sono maggiori (Hannan e Carroll, 1992). Allo stesso tempo, la concentrazione geografica di imprese simili o appartenenti allo stesso settore può generare economie esterne di scala, che tendono a compensare gli svantaggi della maggiore concorrenza (Audretsch e Feldman, 1996; Porter, 1998)<sup>2</sup>. Alcuni autori ritengono che la maggior disponibilità di risorse umane e di lavoro qualificato sia legata a una maggiore propensione all'innovazione e che ciò aumenti l'efficienza delle imprese (Fritsch et al., 2006). Nell'industria dei SNS, l'intensità concorrenziale e le economie esterne sono difficili da valutare. Per semplicità, la prossima sezione assume che il numero di SNS attivi nello stesso paese costituisca una proxy dell'intensità concorrenziale. In relazione agli aspetti geografici, lo sviluppo di un SNS avviene sostanzialmente davanti a un personal computer, la cui localizzazione è, in fondo, irrilevante. Tuttavia, più della metà dei SNS sono stati lanciati negli Stati Uniti, dove la West Coast e la Silicon Valley presentano un'incredibile concentrazione di risorse umane specializzate nella programmazione (Cha, 2013). Quindi la prossima sezione terrà conto della localizzazione dei SNS nell'analisi dei fattori che possono influenzare la sopravvivenza nel mercato.

I lavori sulla sopravvivenza delle imprese normalmente non includono la partecipazione ad operazioni di fusione o acquisizione tra le variabili esplicative. Dal punto di vista dell'impresa target tale aspetto può apparire irrilevante, dal momento che essa è destinata a dissolversi nella nuova entità legale. Tuttavia, un'impresa diventa obiettivo di un'acquisizione nel caso in cui generi profitti e ricavi rilevanti, abbia un marchio riconosciuto oppure conti su un ampio bacino di clienti/utenti. Una volta conclusa l'operazione di acquisizione, l'interruzione delle sue attività o la scomparsa del suo nome/marchio dal mercato possono causare una perdita di valore per l'impresa acquirente (Basu, 2006; Clark et al., 2010), specialmente nell'industria dei SNS, dove il successo di mercato dipende da molteplici utenti che potrebbero sospettare che la piattaforma di riferimento cambi identità (Boyd e Ellison, 2007). Data l'interdipendenza tra le azioni degli utenti, un iniziale e piccolo calo del numero di utenti può trasformarsi in un rapido crollo degli accessi e condurre al collasso del SNS. Naturalmente, l'obiettivo di un'acquisizione può essere anche la semplice eliminazione di un pericoloso concorrente (Kwoka e Elhauge,

<sup>2</sup> In generale, gli *spillover di conoscenza* associati alla prossimità geografica possono aumentare le chance di sopravvivenza delle imprese (Krugman, 1991) anche in presenza di concorrenti numerosi e aggressivi.

2012). In questo caso, la chiusura della target company sarebbe l'esito naturale dell'intera operazione. Dall'altro lato, le fusioni verticali o conglomerali non comportano necessariamente la fine dell'impresa acquisita (come nel caso dell'acquisizione di MySpace da parte di News Corporation, completata nel 2005: l'operazione fu un fallimento, ma MySpace non scomparve). Un SNS può agire anche come acquirente di imprese esterne. Non è possibile in questo caso prevedere l'impatto di un'acquisizione esterna sulla probabilità di sopravvivenza del SNS, poiché la letteratura sulle fusioni ha da tempo sottolineato che, dal punto di vista empirico, le acquisizioni di successo eguagliano quelle fallimentari (si veda, per una rassegna, Ismail et al., 2011; Das e Kapil, 2012). Le prossime sezioni includeranno comunque, tra le variabili di controllo, la partecipazione sia passiva sia attiva ad operazioni di acquisizione. Per la stessa ragione, l'analisi empirica terrà conto della eventuale partecipazione dei SNS a partnership con entità giuridiche separate<sup>3</sup>.

In generale, la letteratura teorica ed empirica sulla sopravvivenza delle imprese ha discusso in dettaglio l'impatto che numerosi fattori possono avere sulla permanenza delle imprese nel mercato. I fattori esplicativi possono essere sia interni sia esterni alle imprese. Tuttavia, non esiste un elemento che risulta sistematicamente associato a una maggiore o minore probabilità di sopravvivenza, e l'impatto dei vari fattori su quest'ultima variabile sembra variare da settore a settore.

### 3. Statistiche descrittive

I siti web di social network (SNS) sono “servizi via web che permettono agli individui di costruire un proprio profilo pubblico o semi-pubblico all'interno di un sistema chiuso, articolare una lista di altri utenti con i quali condividere una connessione, e poter osservare la loro lista di connessioni e quelle di altri all'interno del sistema stesso” (Boyd e Ellison, 2008, p.211). Usando questa definizione, questo paper considera i SNS “citati” dalla letteratura accademica. In particolare, nel marzo 2015 è stata lanciata in EBSCOhost una ricerca usando la parola chiave “social networking websites”. La ricerca ha prodotto 407 SNS menzionati nei primi 1000 risultati (la lista

<sup>3</sup> Ci sono molti casi di partnership nell'industria dei SNS. Ad esempio, Twitter e Yandex (il più diffuso motore di ricerca russo) annunciarono una collaborazione il 21 febbraio 2012. L'accordo prevedeva che Yandex mostrasse i tweet di informazione più rilevanti immediatamente dopo i risultati di ricerca degli utenti (Prodhan, 2012).

completa dei SNS e delle corrispondenti pubblicazioni è disponibile su richiesta). Sono stati scartati i SNS senza sufficienti informazioni sulle loro caratteristiche, caratteristiche necessarie per l'analisi empirica proposta (si veda più avanti)<sup>4</sup>. La lista dei 273 SNS così rimasti è stata distribuita a 10 studenti remunerati ai quali è stato richiesto di confermare che ogni osservazione corrispondesse effettivamente a un SNS (gli studenti erano liberi di verificare l'informazione con qualunque mezzo desiderassero). Per ogni SNS, era necessario che almeno 8 studenti su 10 confermassero che al nome indicato corrispondesse un SNS. Il campione finale è risultato così composto da 224 SNS.

Ogni SNS del campione è definito dalle seguenti caratteristiche: data di lancio; data di chiusura se non più attivo; tipo di audience del SNS (generalista o specializzato<sup>5</sup>); possibili restrizioni all'accesso (basate, ad esempio, sull'età degli utenti); presenza o meno di servizi premium; vendita o meno di spazi pubblicitari; presenza o meno di attività di e-commerce; paese di origine (in particolare, se il SNS era statunitense oppure no); possibili acquisizioni in cui il SNS è stato il target oppure l'acquirente<sup>6</sup>; presenza o meno di attività di partnership con imprese non appartenenti al settore.

Per tenere conto almeno in parte delle dimensioni dei SNS, i dati comprendono il massimo numero di utenti registrati (*accounts*), disponibile per solo 153 SNS. Il numero di utenti di un social network è una *time-variant covariate* e non è disponibile per tutti gli anni e per tutti i SNS. Qui, l'idea è verificare se un numero minimo di utenti è sufficiente a influenzare positivamente la sopravvivenza nel mercato di un SNS.

La Tabella 1 mostra le statistiche descrittive delle variabili qualitative. Il 70% dei SNS utilizza la pubblicità come fonte di finanziamento, mentre pochi SNS vendono beni o servizi online. La maggior parte dei SNS ha il quartier generale negli Stati Uniti (59.8%), mentre il campione è diviso a metà tra SNS generalisti e specializzati. I SNS hanno generalmente com-

<sup>4</sup> I dati sulle caratteristiche strutturali dei SNS sono stati raccolti all'interno delle pubblicazioni presenti in EBSCOhost. I dati raccolti sono stati integrati da altre informazioni presenti sui siti web dei SNS ancora attivi.

<sup>5</sup> È stato chiesto ad altri dieci studenti di stabilire se gli SNS fossero specializzati o generalisti. Le classificazioni degli studenti sono risultate altamente congruenti le une con le altre.

<sup>6</sup> I dati riguardanti le operazioni di fusione e acquisizione sono stati controllati mediante il database Zephyr.



piuto poche acquisizioni mentre 61 di essi sono stati acquisiti da imprese esterne.

**Tabella 1. Statistiche descrittive**

	<b>attivi</b>	<b>cessati</b>	<b>totale</b>
specialized	98	14	112
restrictions	45	11	56
premium	34	7	41
advertising	120	30	150
e-commerce	14	2	16
us	105	29	134
partnership	55	7	62
M&A	19	2	21
target	42	19	61
Obs.	181	43	224

## 4. Analisi empirica

Prima di condurre l'analisi di durata (*duration analysis*), i dati raccolti possono far luce sulle relazioni di tipo statico tra le variabili esplicative. La Tabella 2 mostra i risultati di una *probit analysis* dove la variabile dipendente è uguale a 1 se il SNS era ancora attivo e a 0 altrimenti. Le variabili esplicative indicate in Tabella 2 sono: *specialized*, uguale a 1 se il SNS si rivolge esplicitamente a un pubblico particolare, a 0 altrimenti; *restrictions*, uguale a 1 se il SNS impone limitazioni all'accesso, a 0 altrimenti; *premium*, uguale a 1 se il SNS offre contenuti o servizi addizionali dietro il pagamento di una somma positiva di denaro, a 0 altrimenti; *advertising*, uguale a 1 se il SNS vende "spazio pubblicitario" agli inserzionisti, o altrimenti; *e-commerce*, uguale a 1 se il SNS vende direttamente beni e servizi sul sito web ai suoi utenti, a 0 altrimenti; *competition*, il numero dei SNS attivi nello stesso Stato in ogni periodo considerato (in questo caso, il mese); *us*, uguale a 1 se il SNS ha la sua base negli Stati Uniti, a 0 altrimenti; *partnership*, uguale a 1 se il SNS è coinvolto in un accordo di collaborazione, joint venture, partnership, etc. con imprese esterne, a 0 altrimenti; *M&A*, uguale a 1 se il SNS ha acquisito imprese esterne, a 0 altrimenti; *target*, uguale a 1 se il SNS è stato acquisito da imprese esterne, a 0 altrimenti.

<b>Tabella 2. Probit analysis: effetti marginali sulla probabilità di sopravvivenza</b>				
	<b>I</b>	<b>II</b>	<b>III</b>	<b>IV</b>
<i>specialized</i>	0.747** (0.299)	0.768** (0.300)		
<i>restrictions</i>	0.094 (0.238)		0.080 (0.233)	0.074 (0.236)
<i>Premium</i>	0.069 (0.265)			0.044 (0.265)
<i>advertising</i>	-0.086 (0.220)	-0.074 (0.221)	10.116 (0.217)	
<i>e-commerce</i>	0.269 (0.449)	0.229 (0.453)	0.173 (0.417)	0.242 (0.428)
<i>competition</i>	0.019 (0.019)	0.019 (0.019)	-0.012 (0.140)	-0.003 (0.210)
<i>us</i>	-0.310 (0.215)	-0.333 (0.217)	-0.258 (0.209)	-0.264 (0.211)
<i>partnership</i>	0.150 (0.249)	0.146 (0.248)	0.205 (0.245)	0.124 (0.245)
<i>M&amp;A</i>	0.419 (0.419)	0.459 (0.424)	0.308 (0.401)	0.313 (0.409)
<i>target</i>	-0.500** (0.214)	-0.505** (0.215)	-0.521** (0.212)	-0.552*** (0.213)
<i>premium*restrictions</i>		0.797 (548)		
<i>specialized*premium</i>			0.213 (0.372)	
<i>specialized*advertising</i>				0.512* (0.274)
<i>constant</i>	0.490 (0.496)	0.494 (0.498)	1.351*** (0.348)	0.980*** (0.358)
<i>LL</i>	-100.155	-98.968	-103.443	-101.827
<i>Pseudo R2</i>	0.086	0.096	0.056	0.071
<i>n</i>	224	224	224	224
<b>Note.</b> I coefficienti in tabella sono effetti marginali. Livello di significatività: ***0.01, **0.05, *0.10. Errori standard in parentesi.				

La Tabella 2 mostra che soltanto due variabili sono associate sistematicamente alla probabilità di sopravvivenza. In primo luogo, se il SNS si rivolge a un pubblico specifico, avrà una maggiore probabilità di rimanere sul mercato. In secondo luogo, se il SNS è stato oggetto di una acquisizione la probabilità di sopravvivenza è significativamente più bassa. L'impatto delle altre variabili

non è significativo. Le regressioni II, III e IV includono alcuni termini di interazione, che comunque non modificano il risultato principale, nonostante il ricorso alla pubblicità sembri più importante per la sopravvivenza dei SNS dedicati a un pubblico particolare<sup>7</sup>. Questi risultati comunque ignorano completamente il fatto che la variabile “durata” dei SNS è necessariamente “troncata”.

La variabile di interesse nella *duration analysis* è il tempo che intercorre tra l'inizio di un evento e la sua fine o la fine del periodo di osservazione. Quindi l'osservazione consiste in una *cross section* di durate  $t_1, t_2, \dots, t_n \in T$ , dove  $T$  è una variabile casuale discreta o continua. L'analisi stima la probabilità che l'evento “fallimento” avvenga nel periodo successivo. In questo lavoro, la variabile dipendente è il periodo di sopravvivenza del SNS, calcolato come differenza tra il tempo  $t$  e la data di lancio del SNS, mentre l'evento “fallimento” coincide con la cessazione del sito web del SNS. Quindi la variabile  $T$  è necessariamente troncata. La probabilità cumulata della variabile causale  $T$  è

$$F(t) = \int_0^t f(s)ds = Pr(T \leq t)$$

dove  $f(s)$  è la distribuzione continua di probabilità. La probabilità che il SNS sia attivo per un periodo almeno uguale a  $t$  è data dalla seguente funzione di sopravvivenza

$$S(t) = 1 - F(t) = Pr(T \leq t)$$

e la probabilità che il “fenomeno” giunga a conclusione nell'intervallo successivo di tempo ( $\Delta$ ) è

$$l(t, \Delta) = Pr(t \leq T \leq t + \Delta | T \geq t)$$

Il tasso di rischio è il tasso al quale gli intervalli sono completati dopo una durata pari a  $t$  (posto che durino almeno fino a  $t$ ):

$$\begin{aligned} \lambda(t) &= \lim_{\Delta \rightarrow 0} \frac{Pr(t \leq T \leq t + \Delta | T > t)}{\Delta} \\ &= \lim_{\Delta \rightarrow 0} \frac{F(t + \Delta) - F(t)}{\Delta S(t)} = \frac{f(t)}{S(t)} \end{aligned}$$

<sup>7</sup> Le considerazioni seguenti hanno suggerito l'inclusione in Tabella 2 dei termini di interazione: 1) la combinazione di restrizioni e *premium fee* può effettivamente delineare l'audience di un SNS (*premium\*restrictions*); 2) d'altra parte, il pagamento di una *fee* può garantire la fornitura di servizi e contenuti realmente specialistici (*specialized\*premium*); 3) la pubblicità può essere efficace e garantire entrate rilevanti se si rivolge ad utenti specifici che cercano contenuti particolari (*specialized\*advertising*). Le regressioni mostrate nella Tabella 2 sono state replicate utilizzando altri termini di interazione ma i risultati non cambiano.

Per valutare l'impatto delle variabili esplicative sulla probabilità di sopravvivenza, il metodo chiamato *Cox Proportional Hazard Regression* stima il parametro  $\lambda$ . In particolare, la funzione di rischio  $h_i(t)$  del SNS  $i$  è

$$h_i(t) = h(t, x_i) = h_0(t) \exp(x_i' \beta)$$

dove  $h_0(t)$  è una arbitraria funzione di rischio di riferimento,  $x_i$  è un vettore delle variabili esplicative osservate degli  $i$ -esimi SNS e  $\beta$  è il vettore dei parametri incogniti da stimare. Un segno negativo del coefficiente oppure un tasso di rischio inferiore a uno significano che il tasso di rischio si riduce, cioè che la corrispondente variabile esplicativa è associata a una maggiore probabilità di sopravvivenza.

La Tabella 3 mostra i risultati di una regressione Cox, in cui le variabili esplicative sono quelle definite per l'analisi probit. La tabella include anche la variabile *age* che misura i mesi trascorsi dal lancio del SNS, mentre la variabile *gdp* è la variazione del PIL nazionale su base mensile.

**Tabella 3. Cox Proportional Hazard Regression. Sopravvivenza dei SNS, 1995–2015**

	HR(1)	HR(2)	HR(3)	HR(4)	HR(5)	HR(6)
<i>specialized</i>	0.383** (0.181)	0.361** (0.171)			0.181*** (0.117)	0.158*** (0.108)
<i>restrictions</i>	0.980 (0.354)		0.933 (0.330)	0.966 (0.346)	0.739 (0.369)	0.714 (0.394)
<i>premium</i>	0.731 (0.479)			0.703 (0.306)	0.893 (0.517)	1.603 (0.896)
<i>advertising</i>	1.245 (0.531)	1.199 (0.420)	1.273 (0.445)		1.630 (0.802)	0.902 (0.407)
<i>e-commerce</i>	0.482 (0.333)	0.504 (0.380)	0.554 (0.408)	0.502 (0.373)	0.530 (0.448)	1.111 (0.929)
<i>competition</i>	0.986 (0.031)	0.984 (0.030)	1.025 (0.223)	1.013 (0.024)	0.995 (0.043)	0.938 (0.043)
<i>us</i>	1.661 (0.570)	1.660 (0.564)	1.575 (0.538)	1.557 (0.528)	1.366 (0.572)	1.382 (0.619)
<i>partnership</i>	0.811 (0.320)	0.816 (0.319)	0.704 (0.276)	0.801 (0.314)	0.903 (0.458)	0.902 (0.477)
<i>M&amp;A</i>	0.429 (0.320)	0.419 (0.311)	0.477 (0.353)	0.502 (0.372)	0.301 (0.314)	0.623 (0.498)
<i>target</i>	2.245** (0.723)	2.216** (0.705)	2.376*** (0.756)	2.416*** (0.779)	2.081** (0.854)	
<i>age</i>	1.284*** (0.083)	1.279*** (0.083)	1.287*** (0.084)	1.268*** (0.083)	1.349*** (0.118)	1.329*** (0.120)
<i>gdp</i>	0.783 (0.527)	0.581 (0.271)	0.479 (0.385)	0.985 (0.783)	0.628 (0.611)	0.672 (0.601)
<i>accounts</i>					0.998 (0.006)	
<i>premium*restrictions</i>		0.240 (0.245)				
<i>specialized*premium</i>			0.632 (0.398)			
<i>specialized*advertising</i>				0.507 (0.222)		
LL	-191.915	-190.648	-194.336	-193.059	-106.200	-95.399
Prob>chi	0.000	0.000	0.000	0.000	0.001	0.000
n	224	224	224	224	153	163

**Note.** HR=hazard ratio. Un hazard ratio minore di uno significa che la variabile corrispondente è associata a una maggiore probabilità di sopravvivenza, e viceversa. Tutte le variabili sono dummy, a parte *competition* (numero di concorrenti nella stessa nazione), *age* (in mesi), *gdp* (in miliardi di dollari) e *accounts* (in migliaia).

La Tabella 3 mostra che i SNS che si rivolgono a un pubblico specifico hanno una probabilità di sopravvivere molto maggiore di quelli generalisti. Quindi il vantaggio di disegnare un SNS per un pubblico particolare più che compensa lo svantaggio di rinunciare a un'audience molto estesa. In effetti metà dei SNS censiti è di tipo generalista, per cui in quel segmento la competizione è molto intensa. Gli inserzionisti, la principale fonte di reddito dei SNS, certamente sono attratti dai "grandi numeri", ma allo stesso tempo prestano attenzione all'omogeneità degli utenti registrati presso un social network, maggiore nel caso di SNS specializzati.

Un secondo risultato riguarda l'età dei SNS. I "vecchi" SNS hanno una maggiore probabilità di cessare l'attività. La letteratura sulla sopravvivenza delle imprese assume che l'età è una misura indiretta dell'esperienza delle imprese, cruciale per conservare una quota di mercato maggiore di zero. Nell'industria dei SNS, gli effetti di rete sono pervasivi e gli utenti devono sostenere alti costi di cambiamento per passare da una comunità virtuale all'altra. Tuttavia, i SNS devono adattare le loro caratteristiche alle ondate di nuovi utenti che ogni mese e anno entrano nel mercato. Le preferenze dei nuovi utenti sono in continuo cambiamento e sono per lo più ignote. I nativi digitali entrano nell'industria ed esprimono le loro preferenze semplicemente unendosi a uno o più social network senza sostenere alcun costo. In questo ambiente turbolento, i SNS più "giovani" hanno mostrato una maggiore abilità di sopravvivere, almeno nel periodo sotto esame.

I SNS che sono stati oggetti di un'operazione di acquisizione hanno una probabilità di chiudere i battenti doppia rispetto a quelli che sono rimasti indipendenti. La significatività delle variabili *age* e *target* può avere una spiegazione "dinamica". Alcune compagnie entrano nell'industria dei SNS mediante un'acquisizione e ottengono immediatamente una massa di utenti che aumenta la probabilità di rimanere sul mercato. In altre parole, un'impresa target potrebbe continuare a svolgere attività di social network con un altro nome e questo fenomeno andrebbe ad aumentare la probabilità di sopravvivenza delle imprese che apparentemente sono entrate da poco nel mercato. Osservazioni casuali all'interno del campione mostrano però un quadro eterogeneo: alcune imprese acquirenti non proseguono le attività dell'impresa acquisita, mentre altre incorporano l'impresa acquisita, così com'è, nelle proprie pagine web, mantenendone anche il nome

originale<sup>8</sup>. Per controllare questo possibile *bias*, la Colonna HR(6) della Tabella 3 non considera le acquisizioni. I risultati sono confermati: la specializzazione e l'età dei SNS continuano a influenzare la probabilità di cessare le attività.

Quando la motivazione dell'acquisizione è la soppressione di un concorrente pericoloso, la chiusura della target company potrebbe essere un esito scontato. Tuttavia, la fine del SNS acquisito non è l'ovvia conseguenza di fusioni verticali o conglomerali. Per controllare questo aspetto, le acquisizioni sono state classificate come orizzontali e non orizzontali. Un'acquisizione è definita orizzontale se la società acquirente controllava un altro SNS al momento dell'acquisizione (quindi la stessa impresa avrebbe potuto completare acquisizioni sia orizzontali sia non orizzontali; ad esempio, Yahoo acquistò Flickr nel 2005 e Tumblr nel 2013; la seconda acquisizione è orizzontale, la prima no). La Tabella 4 mostra che il 25% dei SNS ha chiuso i battenti dopo essere stata oggetto di un'acquisizione orizzontale, mentre le acquisizioni non orizzontali hanno condotto alla chiusura dell'impresa target in circa metà dei casi.

<b>Tabella 4. Imprese target: tipo di acquisizione e sopravvivenza</b>		
	<b>attivi</b>	<b>cessati</b>
Orizzontali	27	9
Non orizzontali	13	12

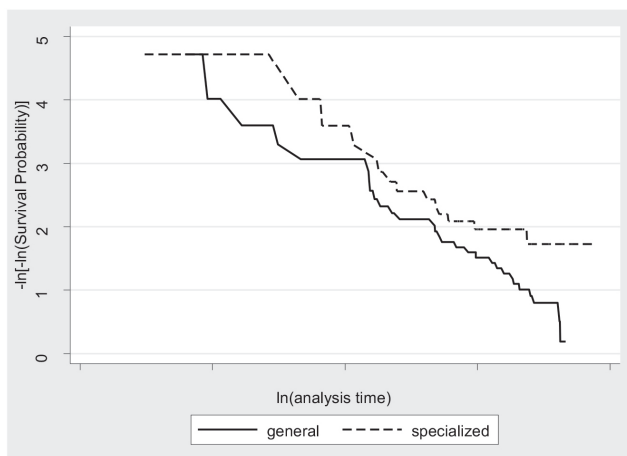
Un'analisi più rigorosa (analisi *probit*, in cui la variabile dipendente è una dummy uguale a 1 se il SNS è cessato, o altrimenti, e le variabili esplicative sono quelle indicate in Tabella 3, con un numero di osservazioni pari a 61), mostra che la probabilità di chiusura non è influenzata dal tipo di acquisizione. In conclusione, le acquisizioni orizzontali che "eliminano" un concorrente oppure ne cambiano il nome non costituiscono una pratica frequente. Rimane il fatto che le acquisizioni sono associate significativamente a un minor numero di SNS nel mercato.

<sup>8</sup> Talvolta il processo di acquisizione ha un esito diverso: l'impresa obiettivo viene chiusa e gli account degli utenti sono dirottati verso altri servizi, insieme ad alcune funzionalità dell'impresa target. Ad esempio, Amazon acquistò PlanetAll nel 1998 e lo chiuse due anni dopo, annunciando che gli utenti di PlanetAll avrebbero potuto proseguire le loro esperienze social all'interno dei servizi e-commerce di Amazon (<https://www.cnet.com/news/amazon-to-shut-planetall-absorb-features/>).

I diversi metodi di finanziamento delle comunità virtuali online (pubblicità, e-commerce, contenuti premium) e restrizioni all'accesso non influenzano la probabilità di sopravvivenza. Anche la ricerca di espansione diretta o indiretta (come le acquisizioni di imprese esterne e le partnership), la localizzazione geografica e l'andamento economico generale hanno un impatto scarso o nullo sulla probabilità di chiudere i battenti. Le regressioni II, III e IV includono i termini di interazione descritti nella nota 7. I coefficienti dei termini di interazione non sono significativi e non modificano i risultati della prima regressione. La regressione V comprende il numero massimo di account registrati dai SNS (*accounts*), disponibile solo per 153 imprese. Il tasso di rischio (*hazard ratio*) è vicino a uno e non significativo.

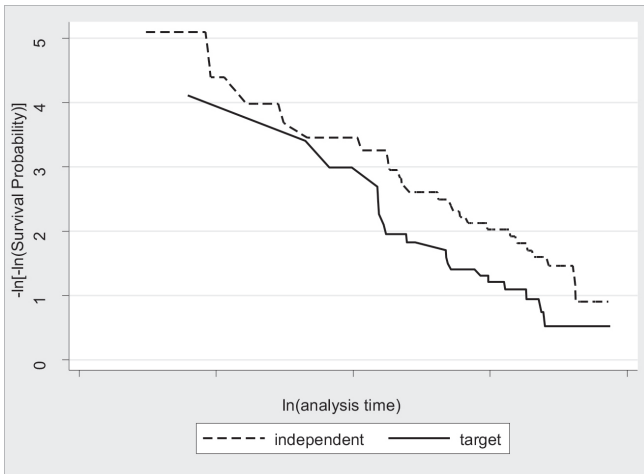
Le Figure 1 e 2 mostrano il grafico log-log relativo alle variabili *specialized* e *target*. Infatti il *Cox regression model* assume che il tasso di rischio (*hazard ratio*) sia proporzionale nel corso del tempo. L'assunzione di rischi proporzionali non è violata, dal momento che le linee sono sufficientemente parallele. Inoltre, un test di Schoenfeld mostra che non vi è prova che l'assunzione dei rischi proporzionali sia nel complesso violata.

**Figura 1. Funzioni log-log: SNS specializzati e generalisti**





**Figura 2. Funzioni log-log: SNS indipendenti e obiettivo di acquisizioni**



## 5. L'evoluzione dei modelli di business

La sezione precedente ha assunto che le caratteristiche tecnologiche e le strategie di mercato dei SNS fossero classificabili in ampie categorie, e questo per condurre un'analisi di sopravvivenza su un campione più ampio possibile. In quel modo, tuttavia, si perdono alcune particolarità dei diversi modelli di business adottati, a cui si accennerà brevemente in questa sezione. In relazione al concetto di modello di business, si assumerà qui che esso corrisponda al modello di ricavo: come i SNS “fanno soldi”. In realtà il modello di business di un'impresa, secondo la letteratura manageriale-aziendale, è definito da più elementi (Cha, 2013; Rohn, 2015): la creazione di valore (natura del prodotto/servizio; scelta tra attività mono-prodotto o multi-prodotto; caratteristiche distributive; posizionamento dell'impresa), il mercato obiettivo (tipo di consumatori; aree geografiche interessate; dispersione geografica; interazione con i clienti), la fonte delle competenze (aspetti tecnologici e di marketing; *supply chain management*; gestione della marca; *networking*) e il modello di ricavo (fonti di ricavo; politiche di prezzo). Assumere che il modello di business si riassume nelle modalità di generazione di ricavo è in qualche modo riduttivo, ma consente di entrare direttamente nel dettaglio di alcune pratiche commerciali tipiche dei SNS.

Le sezioni precedenti, ad esempio, hanno impiegato la variabile *advertising* per indicare se un SNS vende spazi pubblicitari e/o informazioni private agli inserzionisti. Il rapporto tra social media e inserzionisti può comunque svilupparsi in modo eterogeneo. La modalità più tradizionale è la vendita di “spazio”

all'interno delle pagine web. Anche se ritenuta fastidiosa e invadente da molti utenti dei SNS, la classica pubblicità *display* resiste nei media online. Una forma più interessante e innovativa del rapporto tra social media e inserzionisti è la *syndication*. Il termine nasce negli Stati Uniti per indicare la vendita di programmi televisivi o radiofonici, già trasmessi, a network indipendenti, per repliche nelle stesse o in differenti aree geografiche. Sul web, syndication sta ad indicare, in generale, una situazione in cui un sito di informazione fornisce contenuti ad un numero arbitrario di altri siti che li ridistribuiscono in qualche forma. Ad esempio, Facebook ha nel passato messo a disposizione di altri siti web, come Google e Bing (Microsoft), il profilo pubblico dei suoi utenti, affinché esso fosse collegato ai risultati delle ricerche online. La stessa cosa l'ha fatta Twitter, sempre con motori di ricerca come Google e Yahoo (Google abbandonato tali accordi quando ha lanciato, peraltro senza molto successo, la propria piattaforma social, Google+). La syndication è quindi un accordo tra produttore e distributore che permette al produttore di ammortizzare i costi di creazione di contenuti e ai distributori di utilizzare i contenuti pagando un prezzo basso o addirittura pari a zero. Secondo alcuni analisti media (Clarke, 2008; Cong e Du, 2008; Hmedeh et al., 2011), la syndication ha un grande potenziale di crescita come fonte di ricavo.

Altre differenze sono osservabili nelle modalità di vendita di spazio pubblicitario agli inserzionisti. Ad esempio, MySpace utilizzava quattro modalità di vendita: classica pubblicità *display*, sponsorizzazioni, annunci legati alle ricerche degli utenti e annunci prima o dopo la trasmissione in streaming di audio o video musicali. Twitter invece ha articolato la vendita di spazi pubblicitari tra *promoted tweet*, *promoted trend* e *promoted account*. I primi sono legati alle ricerche degli utenti, mentre i secondi permettono agli utenti di visualizzare le tendenze in termini di tempo, contesto ed eventi, promosse dagli inserzionisti. Infine, i *promoted account* suggeriscono account che al momento l'utente non segue ma che potrebbe trovare interessanti.

Anche nell'ambito dell'e-commerce esiste una certa eterogeneità di comportamento tra i SNS. Ad esempio, Facebook ha utilizzato (fino al 2013) *Facebook credits*, una moneta virtuale, per l'acquisto di giochi, applicazioni e altri oggetti virtuali disponibili solo su Facebook. Su MySpace gli utenti potevano scaricare musica a pagamento fornita da piattaforme come iTunes e Amazon.

Vi sono infine alcune fonti di ricavo strettamente legate alla specializzazione del social media. LinkedIn è un esempio particolarmente interessante. Infatti, da un lato LinkedIn si affida alla vendita di spazi pubblicitari e alla sottoscrizione diretta degli utenti che desiderano servizi aggiuntivi (*premium*).

Dall'altro produce reddito con *Hiring solutions*, che costituisce il 40% dei ricavi del SNS più importante del mondo del business. *Hiring solutions* è un servizio di reclutamento a pagamento che si rivolge a coloro che stanno cercando lavoro ma anche a chi non è *job-seeker* attivo.

Il fatto che i SNS più importanti si affidino a più fonti di ricavo non sorprende. Privilegiare una sola fonte di ricavo può essere rischioso. I SNS di successo tendono quindi a diversificare tra le possibili e già collaudate forme di ricavo e a svilupparne di nuove, basate sulla combinazione di servizi e contenuti specializzati con effetti di rete diretti e incrociati. Resta il fatto che la maggior parte dei ricavi dei SNS continua a provenire dagli inserzionisti (Cha, 2013) e che la monetizzazione diretta, mediante pagamento di un prezzo positivo da parte degli utenti, rimane un problema dei social media.

## 6. Conclusioni

È piuttosto comune imbattersi in suggerimenti e consigli riguardanti le caratteristiche che dovrebbe possedere un SNS per raggiungere il successo e produrre reddito. Di solito, simili considerazioni sono tratte dallo studio di storie “esemplari”, familiari al grande pubblico perché relative a una manciata di comunità virtuali di successo. Queste “storie” risultano appassionanti perché le barriere all'entrata nel nell'industria dei SNS sono basse (Beuscart e Mellett, 2008). Tuttavia, gli effetti di rete fanno emergere poche imprese: le informazioni sui “vincenti” abbondano, mentre i “perdenti” lasciano tracce digitali evanescenti. Sfortunatamente, saper molto dei leader non facilita necessariamente l'entrata di follower di successo nel nuovo sistema dei media. Dal punto di vista dell'analisi empirica e degli studi di industria, la digitalizzazione delle attività economiche e l'assenza di un registro degli esperimenti social online rende la raccolta dei dati particolarmente problematica.

Questo paper aveva due obiettivi. Primo, costruire un database dei social network apparsi online dal 1995 in avanti. Per farlo, il “nome” di un social network doveva essere associato alle caratteristiche di base del servizio offerto, ai metodi di finanziamento e alle relazioni con imprese esterne. La costruzione del database non è stata semplice, poiché le fonti di dati sono eterogenee e la stessa definizione di SNS può sollevare questioni complesse. Secondo, il paper voleva studiare se alcune caratteristiche dei SNS sono associate a una maggiore probabilità di sopravvivenza all'interno di questa particolare “in-

dustria". L'analisi empirica ha mostrato che i SNS rivolti a un pubblico specializzato e che rimangono indipendenti hanno una maggiore probabilità di proseguire le loro operazioni online. Inoltre, i SNS più "vecchi" fronteggiano un maggiore rischio di chiusura, a parità di altre condizioni.

Naturalmente sarebbe ambizioso considerare questi risultati generali e definitivi. Il campione comprende soltanto i SNS che, per varie ragioni, sono o sono stati "popolari", perché soltanto la popolarità permette di conoscere il SNS e le sue caratteristiche. Pertanto il campione è inevitabilmente viziato e l'interpretazione dei risultati empirici richiede molta cautela. Inoltre il campione include relativamente poche imprese.<sup>9</sup>

La ricerca futura dovrà quindi perfezionare il processo di raccolta dei dati, la certificazione dell'informazione raccolta e l'aumento delle dimensioni del campione. Inoltre, questo paper ha assunto che alcune variabili fossero uni-dimensionali, ma, come ha mostrato la sezione precedente, alcune strategie, apparentemente simili, nascondono differenze sostanziali che possono riflettersi sulla capacità di rimanere sul mercato. Naturalmente la definizione delle strategie dei SNS attraverso variabili qualitative, prevalentemente dummy, comporta un tradeoff tra la capacità descrittiva e la rilevanza dell'analisi econometrica. Questo aspetto merita attenzione date le ridotte dimensioni del campione.

Da una prospettiva antitrust, le esternalità dirette e incrociate all'interno dei social network consentono a poche imprese di dominare il mercato e favoriscono la creazione di monopoli o quasi monopoli (Haucap e Heimeshoff, 2014). Una concentrazione di mercato troppo elevata può danneggiare utenti e consumatori, e i mercati dell'informazione non fanno eccezione. Infatti, nonostante i mercati informativi siano inequivocabilmente diversi da quelli "tradizionali" come quello automobilistico o farmaceutico, essi devono comunque rispettare le norme sulla concorrenza. La Commissione Europea ha chiarito che anche i mercati informativi sono soggetti alle regole descritte negli articoli 101 e 102 del Trattato sul funzionamento dell'Unione Europea<sup>10</sup> (Fietkiewicz e Lins, 2016).

<sup>9</sup> Naturalmente, in assenza di informazioni sul numero di tentativi di lanciare un SNS, i termini "pochi" e "molti" possono suonare ambigui.

<sup>10</sup> Un esempio recente di applicazione della normativa antitrust ai mercati informativi è la fusione tra Facebook e WhatsApp, comunicata alla Commissione Europea nel 2014. La Commissione non ha bloccato la fusione, nonostante la posizione dominante delle due imprese in mercati "adiacenti" e alcune complementarità tra i servizi offerti (Fietkiewicz e Lins, 2016).

Il periodo durante il quale un'impresa si trova in posizione dominante è un aspetto cruciale per stabilire se l'esercizio del potere di mercato comprometta la concorrenzialità di un mercato. Infatti utenti, consumatori e inserzionisti non sarebbero danneggiati se un SNS esercitasse il potere di mercato per un breve periodo. Anzi, gli utenti potrebbero beneficiare degli effetti di rete e delle complementarità nei servizi che emergono all'interno del social network. Questo lavoro ha mostrato che i SNS "giovani" hanno una maggiore probabilità di sopravvivenza nel mercato rispetto a quelli più "vecchi" e con maggiore esperienza. Questo risultato ricorda la concorrenza schumpeteriana in cui le imprese entranti, per guadagnare quote di mercato e raggiungere una posizione di monopolio, investono in innovazioni che migliorano le condizioni dei consumatori, convincendoli a sostenere costi di cambiamento (*switching costs*) per abbandonare il vecchio network ed entrare in uno nuovo. Inoltre, il maggior tasso di sopravvivenza dei SNS specialistici può ridurre la pervasività delle piattaforme social generaliste e dominanti (come Facebook) e favorire la frammentazione del mercato, anche tenuto conto del comportamento *multi-homing* di molti utenti. Quindi i risultati di questo lavoro sembrerebbero ridurre le preoccupazioni di natura antitrust riguardanti il settore dei SNS, almeno in relazione agli abusi di posizione dominante (Messina, 2006).

Tuttavia lo stesso Schumpeter (1994 [1954]: 897 segg.) avvertiva che "esistono mezzi a disposizione dell'imprenditore di successo - brevetti, strategie, e così via per prolungare la vita della sua posizione monopolistica o quasi monopolista e per impedire ai concorrenti di avvicinarsi troppo...". Fusioni e acquisizioni rientrano certamente nelle "strategie" adottate da alcune imprese per mantenere una posizione dominante nell'industria dei SNS. Il campione utilizzato in questo lavoro include acquisizioni completate da giganti come Google, Amazon, Microsoft, Telefonica e Yahoo<sup>11</sup>. La maggioranza di questi protagonisti dell'economia di Internet ha tentato di lanciare una propria comunità virtuale. Acquisire un SNS con una certa popolarità è un sistema rapido per assicurarsi tecnologia di frontiera, risorse intangibili e fedeltà degli utenti. Un'acquisizione permette di conservare sia le funzionalità di successo di un SNS sia la massa di utenti registrati. Il completo e istantaneo trasferimento di clienti e consumatori da un'impresa all'altra non è osservabile nei mercati tradizionali, nonostante la fiducia e la fedeltà dei consumatori possano facilitare il processo di integrazione post-acquisizione. Inoltre, molte fusioni nell'industria dei

<sup>11</sup> Facebook ha completato circa 60 acquisizioni tra il 2005 e il 2016.

SNS possono rivelarsi anti-competitive: oltre alle acquisizioni orizzontali che hanno lo scopo di eliminare un concorrente, alcune operazioni sono condotte per evitare che un social network cada nelle mani dei concorrenti (Waller e Sag, 2015). Queste strategie sono difficili da identificare, specialmente quando la società target dell'operazione è di ridotte dimensioni. Infatti le autorità antitrust sono tenute a valutare le operazioni di concentrazione che si collocano al di sopra di date soglie dimensionali (relative all'intera operazione o al fatturato delle singole imprese coinvolte), poiché il controllo preventivo delle fusioni si concentra sugli effetti strutturali che l'operazione pianificata avrà sul mercato e sulle conseguenze in termini di esercizio di potere di mercato e di pratiche collusive. Tuttavia, il settore dei SNS è, da questo punto di vista, molto particolare. In primo luogo, le fusioni non coinvolgono molte risorse materiali, la cui riorganizzazione, nei mercati tradizionali, farebbe aumentare i costi dell'intera operazione. In altre parole, acquistare un SNS non richiede troppi sforzi organizzativi per l'impresa acquirente. In secondo luogo, l'impresa acquirente ha immediato accesso a una massa di utenti e ciò facilita la generazione di effetti di rete diretti e incrociati. Quindi anche se l'impatto delle fusioni sulla struttura di mercato è marginale, esse aumentano le chance di mantenere una posizione dominante anche in futuro. Queste circostanze, unite alla maggiore probabilità di cessazione delle imprese target, induce a valutare con molta attenzione la natura anti-competitiva delle fusioni nell'industria dei SNS. Questo aspetto, in sé importante, dovrebbe quindi inserirsi nel più ampio dibattito riguardante la compatibilità delle norme sulla concorrenza con il rapido sviluppo dei mercati informativi.

## Riferimenti bibliografici

- Albarran A. B. (2013), *The Social Media Industries*. New York: Routledge.
- Aldrich H., Auster, E. R. (1986), "Even dwarfs started small: liabilities of age and size and their strategic implications", *Research in Organizational Behavior*, 8, 165-186.
- Anand B. N., Shachar, R. (2009), "Targeted advertising as a signal", *Quantitative Marketing and Economics*, 7(3), 237-266.
- Armstrong A., Hage, J. (2000), *The real value of online communities*. Knowledge and communities, 85-95.
- Audretsch D. B., Feldman M. P. (1996), "Innovative clusters and the industry life cycle", *Review of Industrial Organization*, 11(2), 253-273.
- Audretsch D. B. (1995), *Innovation and Industry Evolution*. Cambridge (MA): Mit Press.
- Barnett W. P. (1990), "The organizational ecology of a technological system". *Administrative Science Quarterly*, 35(1), 31-60.

- Barron D. N., West E., Hannan, M. T. (1994), "A time to grow and a time to die: Growth and mortality of credit unions in New York City, 1914-1990". *American Journal of Sociology*, 100(2), 381-421.
- Basu K. (2006), "Merging brands after mergers". *California Management Review*, 48(4), 28-40.
- Beuscart J. S., Mellet, K. (2008), "Business models of the web 2.0: advertising or the tale of two stories". *Communications & Strategies*, Special Issue.
- Boyd D., Ellison N. B. (2007), "Social network sites: Definition, history, and scholarship", *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Bruderl J., Schussler R. (1990), "Organizational mortality: The liabilities of newness and adolescence", *Administrative Science Quarterly*, 35(3), 530-547.
- Bühler J., Baur A. W., Bick, M., Shi, J. (2015), *Big Data, Big Opportunities: Revenue Sources of Social Media Services Besides Advertising*. In *Open and Big Data Management and Innovation* (pp. 183-199), Springer International Publishing.
- Canzer B. (2006), *E-Business: Strategic Thinking and Practice: Strategic Thinking and Practice*. Cengage Learning: Boston.
- Cefis E., Marsili O. (2005), "A matter of life and death: innovation and firm survival", *Industrial and corporate change*, 14(6), 1167-1192.
- Cha J. (2013), *Business models of social networking sites*. In Albarran A. B. (Ed.), *The Social Media Industries* (pp. 60-85), New York: Routledge.
- Clark S.M., Gioia D. A., Ketchen D. J., Thomas J. B. (2010), "Transitional identity as a facilitator of organizational identity change during a merger", *Administrative Science Quarterly*, 55(3), 397-438.
- Clarke R. (2008), "Web 2.0 as syndication", *Journal of theoretical and applied electronic commerce research*, 3(2), 30-43.
- Cong Y., Du H. (2008), "Web syndication using RSS", *Journal of Accountancy*, 205(6), 48-50.
- Cressy R. (2006), "Why do most firms die young?" *Small Business Economics*, 26(2), 103-116.
- Das A., Kapi, S. (2012), "Explaining M&A performance: a review of empirical research", *Journal of Strategy and Management*, 5(3), 284-330.
- Deephouse D. L. (1999), "To be different, or to be the same? It's a question (and theory) of strategic balance", *Strategic management journal*, 20(2), 147-166.
- Dou W. (2004), "Will internet users pay for online content?" *Journal of Advertising Research*, 44(04), 349-359.
- Dwyer C., Hiltz S., Passerini K. (2007), *Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace*. AMCIS 2007 proceedings, 339.
- Enders A., Hungenberg, H., Denker H. P., Mauch S. (2008), "The long tail of social networking: Revenue models of social networking sites", *European Management Journal*, 26(3), 199-211.
- Esteve-Pérez S., Mañez-Castillejo J. A. (2008), "The resource-based theory of the firm and firm survival", *Small Business Economics*, 30(3), 231-249.
- Fackler D., Schnabel C., Wagner J. (2013), "Establishment exits in Germany: the role of size and age", *Small Business Economics*, 41(3), 683-700.
- Fietkiewicz K.J., Lins E. (2016), *New Media and New Territories for European Law: Competition in the Market for Social Networking Services*, in Knautz K., Baran K.S. (eds.), *Facets of Facebook: Use and Users*, 285-324.
- Fogel J., Nehmad E. (2009), "Internet social network communities: Risk taking, trust, and privacy concerns", *Computers in human behavior*, 25(1), 153-160.

- Fritsch M., Brixy U., Falck O. (2006), "The effect of industry, region, and time on new business survival – a multi-dimensional analysis", *Review of Industrial Organization*, 28(3), 285-306.
- Geroski P.A., Mata, J., Portugal P. (2002), *Founding Conditions and the Survival of New Firms*, Working paper 1-03, Banco de Portugal, Economic Research Department, Lisbon.
- Hajli M. (2014), "A study of the impact of social media on consumers". *International Journal of Market Research*, 56(3), 387-404.
- Hannan M.T., Carroll, G. (1992), *Dynamics of Organizational Populations: Density, Legitimation, and Competition*, New York: Oxford University Press.
- Hannan M.T., Freeman J. (1977), "The population ecology of organizations" *American Journal of Sociology*, 82(5), 929-964.
- Haucap J., Heimeshoff, U. (2014), "Google, Facebook, Amazon, eBay: Is the Internet driving competition or market monopolization?", *International Economics and Economic Policy*, 11(1-2), 49-61.
- Henkel J., Block J. (2013), Peer influence in network markets: a theoretical and empirical analysis", *Journal of Evolutionary Economics*, 23(5), 925-953.
- Hmedeh Z., Vouzoukidou N., Travers N., Christophides V., Du Mouza C., Scholl M. (2011), *Characterizing web syndication behavior and content*. In Bouguettaya A., Hauswirth M., Liu L. (Eds.), *Web Information System Engineering-WISE 2011*, Heidelberg: Springer Berlin, 29-42.
- Huang Z., Benyoucef, M. (2013), "From e-commerce to social commerce: A close look at design features", *Electronic Commerce Research and Applications*, 12(4), 246-259.
- Ismail, T.H., Abdou A.A., Annis, R.M. (2011), "Review of literature linking corporate performance to mergers and acquisitions", *The Review of Financial and Accounting Studies*, 1, 89-104.
- Kelly L., Ker, G., Drennan J. (2010), "Avoidance of advertising in social networking sites: The teenage perspective", *Journal of Interactive Advertising*, 10(2), 16-27.
- Kossinets G., Watts D. J. (2006), "Empirical analysis of an evolving social network", *Science*, 311(5757), 88-90.
- Krugman P.R. (1991), *Geography and Trade*, Cambridge (MA): MIT press.
- Kwoka J., Elhauge E. (2012), *Mergers that Eliminate Potential Competition*. Research Handbook on the Economics of Antitrust Laws, 90-120.
- Lambrecht A., Tucker, C. (2013), "When does retargeting work? Information specificity in online advertising", *Journal of Marketing Research*, 50(5), 561-576.
- Laudon K.C., Traver C.G. (2007), *E-commerce*. Pearson/Addison Wesley.
- Lee L.T. (2013), *Privacy and social media*. In Albarran A. B. (Ed.), *The Social Media Industries* (pp. 145-165). New York: Routledge.
- Leonard Barton D. (1992), "Core capabilities and core rigidities: A paradox in managing new product development", *Strategic management journal*, 13(S1), 111-125.
- Lin H.F. (2008), "Determinants of successful virtual communities: Contributions from system characteristics and social factors", *Information & Management*, 45(8), 522-527.
- Mata J., Portugal P. (1994), "Life duration of new firms", *Journal of Industrial Economics*, 42(3), 227-245.
- Messina M. (2006), "Article 82 and the New Economy: Need for Modernisation?", *The Competition Law Review*, 2(2), 73-98.
- Miller K.D., Fabian F., Lin S. (2009), "Strategies for online communities", *Strategic Management Journal*, 30(3), 305-322.



- Porter M.E. (1998), "Clusters and the new economics of competition", *Harvard Business Review*, 76(6), 77-90
- Prodhan G. (2012), *Twitter partners with Yandex for real-time search*, <http://www.reuters.com/>, scaricato il 24 Febbraio 2016.
- Ranger-Moore J. (1997), "Bigger may be better, but is older wiser? Organizational age and size in the New York life insurance industry", *American Sociological Review*, 62(6), 903-920.
- Rohn U. (2014), "Cross-Border Connectivity through Social Network Sites", *Journal of Social Media Studies*, 1(1), 35-52.
- Rohn U. (2015), *Social Media Business Models*. In *The International Encyclopedia of Digital Communication and Society*, 3 Volume Set. John Wiley & Sons.
- Schumpeter J.A. (1994 [1954]), *History of Economic Analysis*. London, UK: Routledge.
- Shriver S.,K, Nair H.S., Hofstetter R. (2013), "Social ties and user-generated content: Evidence from an online social network", *Management Science*, 59(6), 1425-1443.
- Stern I., Henderson A.D. (2004), "Within business diversification in technology intensive industries", *Strategic Management Journal*, 25(5), 487-505.
- Stinchcombe A.L., March J.G. (1965), *Social structure and organizations*. *Advances in strategic management*, 17, 229-259.
- Strotmann H. (2007), "Entrepreneurial survival", *Small Business Economics*, 28(1), 87-104.
- Subrahmanyam K., Reich S.M. Waechter, N., Espinoza G. (2008), "Online and offline social networks: Use of social networking sites by emerging adults", *Journal of applied developmental psychology*, 29(6), 420-433.
- Taylor, D.G., Lewin J E., Strutton D. (2011), "Friends, fans, and followers: do ads work on social networks?" *Journal of Advertising Research*, 51(1), 258-275.
- Tucker C.E. (2014), "Social networks, personalized advertising, and privacy controls", *Journal Of Marketing Research*, 51(5), 546-562.
- Varum C.A., Rocha, V.C. (2012), "The effect of crises on firm exit and the moderating effect of firm size", *Economics Letters*, 114(1), 94-97.
- Wagner J. (1994), "The post-entry performance of new small firms in German manufacturing industries", *Journal of Industrial Economics*, 42(2), 141-154.
- Walle, S.W., Sag, M. (2014), Promoting innovation, *Iowa Law Review*, 100, 2223-2247.
- Wang C.L., Zhang Y., Ye, L.R., Nguyen D.D. (2005), "Subscription to fee-based online services: what makes consumer pay for online content?", *Journal of Electronic Commerce Research*, 6(4), 304-324.
- Wilcox K., Stephen A.T. (2013), "Are close friends the enemy? Online social networks, self-esteem, and self-control", *Journal Of Consumer Research*, 40(1), 90-103.
- Zahavi T., Lavie D. (2013), "Intra-industry diversification and firm performance", *Strategic Management Journal*, 34(8), 978-998.
- Zeng M., Reinartz W. (2003), "Beyond online search: The road to profitability". *California Management Review*, 45(2), 107-130.



# **Firm growth, export behavior, and the age–size profile: evidence from the Italian manufacturing sector<sup>1</sup>**

MARCO GRAZZI E DANIELE MOSCHELLA

## **1. Introduction**

An important stream of literature within industrial economics has for long been interested in assessing the contribution to employment creation stemming from the different firm-size classes. In this respect, at least since Birch (1981), small firms have been considered as a much relevant source of job creation. The increasing availability of firm level dataset has further contributed to foster research on the issue, starting from the seminal works of Davis and Haltiwanger (1992) and Davis *et al.* (1996). These studies represented a relevant advancement for the understanding of employment and industrial dynamics, in that they confirmed, by means of new methodological and empirical tools, that smaller firms are major players in terms of job churning, hence contributing both to employment creation and destruction.

<sup>1</sup> A more comprehensive version of the article, which also includes analysis of nonemployer firms, is forthcoming on the *Journal of Evolutionary Economics* (see Grazzi and Moschella, forthcoming). Daniele Moschella received financial support by the Italian Ministry of Education, University, and Research under the SIR Programme (project code RBSI14JAFW). Without the unique backing of the Italian Statistical Office, and in particular Roberto Monducci, this all endeavour would not have been possible. The usual disclaimer applies.

A much related stream of literature, dating at least as back as to Penrose (1959), while still centered on the general issue of firm growth, was relatively more focused on the determinants of business growth rather than on industry or economy wide implications. Also this stream of literature has conspicuously benefited from the growing availability of firm-level dataset which provided the researchers the unprecedented opportunity to empirically investigate for a variety of potential determinants of firm growth, from innovation to financial performance, from relative productivity to export status and many other firm-specific factors (for a review, see Coad, 2009).

The recent economic slowdown and its consequences further complicated the task of the policy analyst and that of the economist seeking to identify the determinants of firm growth and the role of distinct categories of firms in contributing to aggregate growth. In most industrialized countries, over the last five years the growth rate of GDP per capita has been close to zero, if not negative (Fig. 1.1 from OECD, 2012). In such a scenario, also the most standard and reliable predictors of firm performance, such as productivity or export status, did not regularly allow to discriminate between high and low growth firms. Rather unexpectedly, higher than industry average productivity does not appear to translate into higher firm growth (for evidence on both Europe and US see Bottazzi *et al.*, 2010; Dosi *et al.*, 2015b). In this respect, also the long standing evidence according to which smaller firms tend to display higher growth rates has been challenged recently. Haltiwanger *et al.* (2013) and Lawless (2014), for example, show that once accounting for firm age, the inverse relationship between growth and size declines very markedly or even disappears.

The contribution of the paper to the literature on firm age and performance is twofold. First, by resorting to the uncensored population of Italian manufacturing employer firms we investigate the role of several firm characteristics in explaining the growth of companies. In this respect, the unprecedented availability of the universe of firms enables us to insulate against recurring issues due to censoring, truncation or selection bias. We use such data to investigate the property of the growth process for firms of different age classes. Second, building upon the literature on firm age and growth recalled above, we identify and focus our analysis on a category of firms, young exporters, which displays several of the characteristics of high growth businesses. The fact that productivity is not systematically related to firm growth suggests that such pure efficiency measure might not adequately capture the willingness to take up new business opportunities. In this respect, the setting up of a new firm which exports since its very inception, signals both the ex-

istence of the skills and capabilities required to engage in international trade and, at least as relevant, the willingness to look for growth opportunities not only domestically, but also abroad.

In this respect, we borrow from the entrepreneurship literature the concepts of “willingness to growth” (Davidsson, 1989) and that of “born-global” (Knight and Cavusgil, 2004). The entrepreneurship literature acknowledges the relevance of motivation, especially managerial motivation, in determining different patterns of firm growth even when accounting for different levels of human capital (Delmar and Wiklund, 2013). Having, for example, pursued a long education does not force anyone to pursue growth. However, if a manager wants their firm to expand, then it appears that having extensive education or rich experience is helpful for making such ambitions materialise (Davidsson and Wiklund, 2013). What we try to argue and show in this paper is that one might well observe an analogous relation with productivity: higher efficiency per se need not to translate into higher growth, however above average productivity *and* willingness to growth are more likely to result in actual firm growth.

We put into operation the concept of higher firm efficiency *and* willingness to growth by looking at the category of early exporters or born globals,<sup>2</sup> which are defined as business organizations that, from or near their founding, are engaged in international trade (Autio *et al.*, 2000; Zahra *et al.*, 2000; Knight and Cavusgil, 2004). The effectiveness of the export status as a signal for growth willingness is rather limited to the first years of activity of a firm: if among smaller firms, exporting is the exception, it becomes more common as the firm gets older and bigger (cf. Table 2 and 5 below).

We investigate the role of age on performance employing an integrated set of data merging information from the census of active Italian firms and data based on customs records covering all Italian exporting firms. Our findings reveal that, as could be expected, age impacts significantly on the propensity to export, with young firms less likely to be engaged in international trade. In this respect the status of “young exporter” is coherent with the firm level characteristics of willingness to grow and higher efficiency. In particular, with respect to the latter, following recent theoretical models and empirical results in the trade literature, we interpret the presence on international markets as evidence of higher productivity (Melitz, 2003; Bernard *et al.*, 2007). In gener-

<sup>2</sup> In this paper we will use interchangeably the terms born global and young (or early) exporters.

al, exporters report higher growth rates over all the age spectrum, although the difference between exporters and non exporters is more apparent among younger firms. More in detail, we find that, controlling for age, the negative size-growth relationship does not disappear, contrary to the recent evidence mentioned above. We also exploit the rich firm-product-country dataset for exporting firms to control for potential endogeneity issues and to further investigate the role of young exporters in foreign markets. Rather surprisingly, once accounted for size differences, the value of exports of born-global within each product-country combination is not smaller than that of older firms. Remarkably, in some years the value of exports is even bigger, and this is the result of born global being able to ship bigger quantities. Early exporters also appear to be well equipped to face exchange rates variations as their exports decrease less following a currency appreciation.

The paper is organized as follows. Section 2 describes the data. Sections 3 and 4 provide a first account of the age-size profile and the growth patterns of firms, Section 5 presents the results of the regression analysis. Section 6 concludes.

## **2. Data sources**

The analysis is based upon two firm-level datasets collected by the Italian statistical office (ISTAT), namely the Business Register known as “Archivio Statistico Imprese Attive” (ASIA) and Statistiche del Commercio Estero (COE). ASIA is the register of all active Italian businesses. It covers the period 1998-2006 and contains information on firms’ operations including the number of employees, total turnover,<sup>3</sup> geographic location of the firm, and much relevant for this work, firm’s age, defined as the year of incorporation. The COE dataset consists of all cross-border transactions performed by Italian firms and it covers the period 2000-2007. Further, the data can be disaggregated at the firm-product-country level, meaning that it is possible to know the value of export of each firm in each product-country destination in which it is active. Since data on physical quantity are also available, it is possible to compute the unit values of transaction within each product-country pair. Using the unique identification code of the

<sup>3</sup> Information on total turnover is available only in 2000 and 2003.

**Table 1. Observations by year and export status**

Year	All firms	Non exporters	Exporters	Exporters(%)
2000	1,217,251	1,107,791	109,460	8.99
2001	1,252,809	1,140,341	112,468	8.98
2002	1,335,538	1,219,673	115,865	8.67
2003	1,387,156	1,271,768	115,388	8.32
2004	1,418,365	1,305,327	113,038	7.97
2005	1,435,918	1,326,702	109,216	7.61
2006	1,494,419	1,380,913	113,506	7.59

**Note.** All firms, excluding self-employment. Exporters are defined as firms with strictly positive exports.

firm, we link the firm-level export data from COE to ISTAT's archive of active firms. The data collection and building process of the integrated database are described at length in *Grazzi et al. (2013)*.

Notice that the resulting dataset is not a sample but rather it covers the universe of Italian active firms and all international trade transactions of Italian firms over the period. In the remainder of the paper a firm is defined as a legal entity reporting a positive number of full-time equivalent employees during the calendar year. Hence the present analysis excludes self-employment.<sup>4</sup> In the end the dataset employed in the empirical analysis consists of 9,541,456 observations spanning from 2000 to 2006.

Table 1 reports the breakdown of our dataset by years and export status. The number of active businesses slightly increases over time, whereas the number of exporters, even if increasing in absolute terms, is decreasing as percentage of the total.<sup>5</sup> Export Participation over the universe of firms is rather low, around 9% in 2000, which is somewhat lower than in most other countries for which evidence is available (*Bernard et al., 2007*; International Study Group on Exports and Productivity, 2008). The reason is twofold. First, data from Structural Business Statistics are usually available only for firms above a certain size threshold. Given the known positive relationship be-

<sup>4</sup> See *Davis et al. (2009)* for a discussion of nonemployers' dynamics and its relationship with employer firms. In a different version of the article (see *Grazzi and Moschella, forthcoming*) we also include analysis of nonemployer firms.

<sup>5</sup> Note that we only consider firms with a positive number (or fraction) of employees, thus excluding self-employment. This is why we get a few less exporting firms than in *Bernard et al. (2015)* which employs the same set of data.

tween size and export status, left truncation of the size distribution provides an overestimation of the percentage of exporting firms. This is the case, for instance, of the dataset employed in Serti and Tomasi (2008) and Grazzi (2012) which report a rather high export participation (around 70% for manufacturing firms) for firms bigger than 20 employees. Second, and related, the firm size distribution of Italian business companies is even more left skewed than that of other countries, hence left truncation causes a bigger bias than for countries with a less skewed distribution of firm size.

Although we cannot report direct evidence on Italian firms, data on a wide set of countries from the World Bank Business Environment and Enterprise Performance Survey (BEEPS) document that on average, the ability to export indirectly (i.e. through export intermediaries) increases by one third the number of firms that can reach foreign markets with their goods (Grazzi and Tomasi, 2016). Given the disproportionate presence of small firms in the Italian economy (see more below) it is reasonable to expect that many small domestic manufacturers would resort to intermediaries to access to foreign destinations.

Table 2 reports the distribution of firms across economic activities and it also shows that imposing a size threshold has also a relevant impact on the

**Table 2. Observations by sector of economic activity in year 2000**

Sector	All firms					>20 employees				
	(I)	(II)	(III)	(IV)	(V)	(I)	(II)	(III)	(IV)	(V)
Manufacturing	291,156	23.92	40.42	23.83	87.29	33,094	51.58	44.60	70.11	92.27
Energy	3,832	0.31	1.10	8.74	0.14	581	0.91	1.47	20.48	0.14
Construction	191,764	15.75	8.46	1.02	0.26	5,203	8.11	4.28	8.40	0.22
Wholesale trade	109,557	9.00	6.65	22.89	9.04	5,165	8.05	4.72	66.21	4.48
Retail trade	144,715	11.89	6.58	4.76	0.74	2,734	4.26	5.23	27.91	0.48
Transport & TLC	47,084	3.87	9.93	3.09	1.22	3,849	6.00	13.43	12.18	1.30
Financial	171,871	14.12	15.54	1.61	1.18	7,396	11.53	17.88	7.56	1.03
Other services	257,272	21.14	11.32	0.62	0.12	6,136	9.56	8.38	4.09	0.07
Total	1,217,251	100	100		100	64,158	100.00	100		100

**Note.** (I) Number of firms; (II) distribution of number of firms (%); (III) distribution of employment (%); (IV) percentage of exporting firms within each sector (%); (V) distribution of export volumes (%).

distribution of firms across sectors of economic activity. When accounting for the universe of firms, the manufacturing sectors account for less than one quarter of the total number of firms and firms in other services, includ-



ing bars, restaurants and hotels are almost as numerous. On the contrary, when imposing the 20 employees threshold, a standard cutoff on many firm level dataset, one recovers the disproportionate share of manufacturing firms that is accustomed to observe. Given the wide variation in the organization of business across sectors of economic activity, in the remainder of the paper, we focus only firms in the manufacturing sectors. Manufacturing firms - as expected - also account for the largest share of exports of goods.<sup>6</sup>

### 3. Age, size, and export propensity

The literature has much investigated the relation between firm size and export propensity, showing robust evidence of a positive relationship (Wagner, 2001; Bernard *et al.*, 2007), which is well accounted by existing theories of sunk costs for export market participation (Melitz, 2003). Less attention has been devoted to the role of firm age in explaining firm propensity to export, mostly due to data limitations. This paper provides the first evidence on the subject exploiting the universe of Italian manufacturing firms.

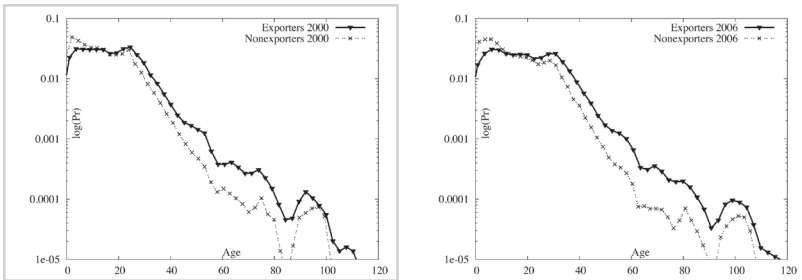
The descriptive evidence suggests that export market participation rates vary with age. Table 3 assigns firms to four age classes, respectively, less than five years from incorporation, between 5 and 10, between 10 and 20, and more than 20 years from incorporation (see Lawless, 2014 for a similar classification). Note that while the universe of firms is rather evenly distributed across age classes (and the same occurs for non exporting firms), exporters are more concentrated in the category of firms with more than 20 years, as more than one third of exporters belong to such class (not reported in the table). The export propensity is increasing within age categories, going from 15.59% of the first class to 31.34% of the last class. Overall, one notices that if being an exporter is already an exception with respect to the universe of manufacturing firms (24%, cf. Table 2) being a young exporter is even a much rarer event.

The relation between firm age and export status is also shown in Figure 1 by means of kernel density estimation for the years 2000 and 2006. The linear decay in the distributions over most of their support suggests that an exponential law would be an acceptable representation of empirical distributions, with departures

<sup>6</sup> For a comparative analysis of the role of manufacturers, vis a vis wholesalers, refer to Bernard *et al.* (2015).

Age class	All firms %	Exporters %
Age <5	21.15	15.59
Age 5-10	20.49	21.16
Age 11-20	28.42	23.97
Age 21+	29.94	31.34
Total	100.00	

**Note.** The second column reports the distribution of firms across the four age classes; the third column reports the percentage of exporters over the total number of firms for each age class. Pearson's chi-squared test of equal distribution of age classes across exporters and nonexporters:  $\chi_3^2=5200$ ; Probability = 0.



**Figure 1. Kernel density of the age distribution for exporters and nonexporters, in 2000 and 2006. Kernel densities are computed using an Epanechnikov kernel**

from the linear fit among the youngest and the oldest firms, in line with the evidence emerging from other studies (Coad, 2010; Coad and Tamvada, 2012; Barba Navaretti *et al.*, 2014). In particular, notice that the modal age among non exporters is 2 years in 2000, whereas it is 24 years among exporters: this implies that the departure from the exponential benchmark is negligible in the case of nonexporting young firms, while it is quite pronounced in the case of young exporting firms.<sup>7</sup> One possible explanation of this is the increase in net entry rate into export markets observed during the first years of a firm's lifecycle (see below). Finally, Figure 1 shows that the age distributions is almost unchanged over time.

<sup>7</sup> An exponential distribution would predict a modal age equal to the very youngest age group (Coad, 2010, p. 10).

Somewhat surprising, age has only a limited impact on the persistency of the export status, as shown by Table 4, which reports the annual transition matrix in and out of exporting for each age class. The probability to export in  $t+1$ , given that a firm was already exporting in the previous year, is around 83% in the first age class, and is only mildly increasing throughout the four classes. In general, the observed persistency in export status is quite high, in agreement with the available evidence from other countries (see Roberts and Tybout, 1997 for data on Colombia, Bernard and Jensen, 2004 for U.S. and Grazzi, 2012 for sample of medium and large Italian firms). Both the high persistency and the low entry rate into export markets (first row of Table 4) provide supporting evidence to the sunk costs hypothesis (Melitz, 2003) for engaging into international trade. In particular, the high persistency in the export status displayed in Table 4 shows that the explanation of exporting behaviour based on sunk costs is robust to age disaggregation.

As recalled above, firm size is known to be a relevant characteristic to predict the export status, since exporters are usually found to be bigger than nonexporters. In Table 5 we take advantage of information on both size and age to provide, for the year 2000, a complete descriptive account of the joint distribution of the variables we are mostly interested in: age, size and export status. In order to do so, we divide firms in four size classes: class 1 contains firms with employees  $>0$  and  $\leq 1$ ; class 2-9 contains firms with employees  $>1$  and  $\leq 9$ ; class 10-19 contains firms with employees  $>9$  and  $\leq 19$ ; class 20+ contains firms with employees  $>19$ . Such classification of firms across size classes is borrowed from Haltiwanger et al. (2013) and then adapted to the peculiarity of the Italian context by grouping together firms bigger than 20 employees which are rather few absolute numbers (35,058) and represent only the 12.04% of the population. Each cell of Table 5

**Table 4. Transition matrix in and out of exporting**

	Age <5		Age 5-10		Age 11-20		Age 21+	
	0	1	0	1	0	1	0	1
0	95.12	4.88	95.69	4.31	95.60	4.40	95.03	4.97
1	17.05	82.95	16.42	83.58	13.94	86.06	11.00	89.00

**Note.** 0 and 1 represent the status of nonexporter and exporter.

<b>Table 5. Distribution of firms by age, size class and export status in 2000</b>					
	<b>Size class</b>				
<b>Age Class</b>	<b>1</b>	<b>2-9</b>	<b>10-19</b>	<b>20+</b>	<b>All</b>
Age <5	18,504	32,860	6,584	3,646	61,594
	6.36	11.29	2.26	1.25	21.15
	6.84	14.28	27.84	49.70	15.59
	0.29	3.05	2.16	6.55	12.05
Age 5-10	13,613	32,289	8,551	5,199	59,652
	4.68	11.09	2.94	1.79	20.49
	7.18	17.32	34.92	59.03	21.16
	0.25	3.21	2.83	7.54	13.83
Age 11-20	16,276	42,383	14,421	9,671	82,751
	5.59	14.56	4.95	3.32	28.42
	5.86	16.87	37.21	65.88	23.97
	0.32	4.38	4.82	12.81	22.33
Age 21+	14,376	41,236	15,005	16,542	87,159
	4.94	14.16	5.15	5.68	29.94
	6.16	17.38	42.49	77.90	31.34
	0.30	4.36	5.04	42.09	51.79
All	62,769	148,768	44,561	35,058	291,156
	21.56	51.10	15.30	12.04	100.00
	6.50	16.54	37.16	68.85	23.83
	1.15	15.01	14.84	69.00	100.00

**Note.** Each cell reports, top to bottom, the number of firms, the percentage over the total (%), the percentage of exporting firms in that cell (%), the share of employment (%).

reports the number of firms, the percentage over the total, and the percentage of exporting firms in the cell. The first two size classes account for more than 72% of all firms, which are distributed across age classes more or less according to the aggregate distribution observed above, with most firms aged between 11 and 20.

Exporters are quite rare, in relative terms, among small firms: export propensity is around 6.5% in the first size class, and around 16.5% in the second class. In these two size classes, export propensity is quite stable across age classes.

The ratio of exporters over nonexporters increases dramatically among firms in the last two size classes, which contain less than 30% of all firms, but with an export propensity which goes from around 37% in the class 10-19, to around 69% in the class 20-max. Focus in particular on this latter size class, here export propensity is markedly increasing in age, going from around 50% among young firms (age <5), to around 78% among old firms (age 21+). Finally, as could be easily expected, even though the distribution of firms is much concentrated in the two classes of smaller size, the largest share of employment (69.00%) is accounted for by the firms bigger than 20 employees.

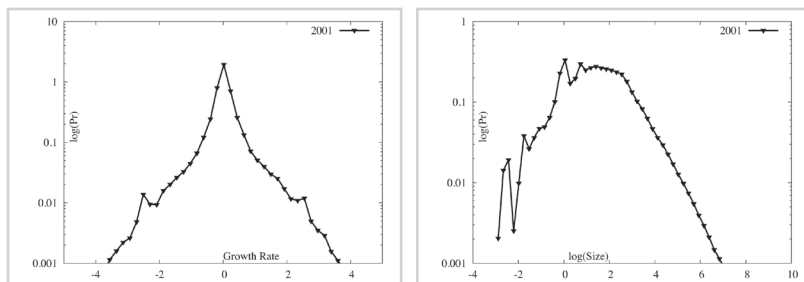
#### 4. Distributions of firm growth rates

In this section, we analyze the growth distribution of the different categories of firms (i.e. with respect to size, age and export status). In order to do so, we focus on the yearly growth rates of employment, as such variable is available for each year. Following the previous literature, firm growth rate is computed as the log-difference between two consecutive years:

$$g_{i,t} = \ln \text{Size}_{i,t} - \ln \text{Size}_{i,t-1} \quad (1)$$

Figure 2 (left) presents the distribution of growth rates for all firms in 2001, together with the size distribution (right) in the same year. The growth rates distribution displays the “tent-shape” characteristic which has been found in several other datasets (see, among others, Stanley *et al.*, 1996; Bottazzi and Secchi, 2006; Coad, 2007; Bottazzi *et al.*, 2011) and that confirms the existence of a large share of extreme growth events, of both positive and negative sign. This work also presents for the first time the size distribution of the whole population (i.e. no left truncation) of Italian manufacturing firms. The right tail of the distribution is quite close to the Pareto benchmark, i.e. a straight line of negative slope on log-log axes (Axtell, 2001). As expected, the right tail of the size distribution is much similar to that computed on the left-truncated sample available in previous version of the same dataset (Bottazzi *et al.*, 2007). The modal value for the number of employees is one full-time equivalent,

25,827 enterprises reported only one employee in 2001; whether on the left tail, 36,603 firms reported



**Figure 2. Kernel density of growth rates and firm size**

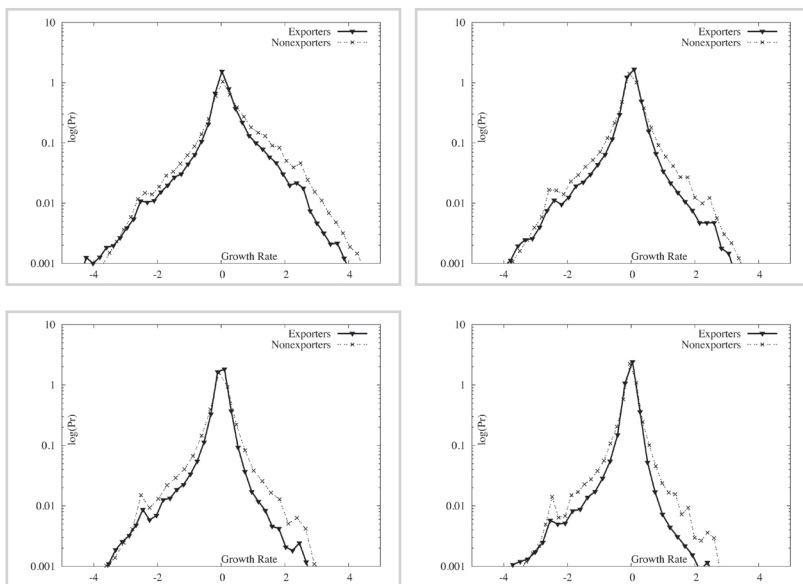
only a fraction of full-time equivalent. These figures confirm the known evidence of the Italian industrial structure characterized by an overwhelming presence of small-medium firms that faces many difficulties to expand (Dosi *et al.*, 2012). More in detail, performing the same analysis as in Hsieh and Klenow (2014) we find that Italian firms in the 30-34 age class are only around 3.5 times bigger than those of the 10-14 age class. This is a somewhat smaller than the 6.5 factor reported in Hsieh and Klenow (2014) which employed a more limited set of data.

As far as the differences in export status are concerned, Figure 3 presents the growth rate distributions for exporters and nonexporters by age class, pooled over the years. We define exporters based on the firm's export status at time  $t-1$ . Quite interestingly, across the four age classes, the support of the distributions of growth rates for both exporters and non exporters, appears to shrink, especially on the right, with young firms experiencing high growth episodes more frequently than older firms. This is consistent with some recent empirical evidence from other countries (see Coad *et al.*, 2013 for Spain, and Barba Navaretti *et al.*, 2014 for a comparative perspective across France, Italy, and Spain) and broadly in tune with theoretical learning models of firm growth à la Jovanovic, which predict that younger firms have both higher and more variable growth rates (Jovanovic, 1982). It also appears that although the supports of the distributions for both exporters and non exporters is not much different, the growth rates of exporting firms are more concentrated around the modal value than non exporting firms.

Figure 4 shows, only for young firms, the distributions of growth rates of exporters and non exporters broken down by size class. Among firms born in the last five years, the bigger ones experience less frequently large growth

episodes (the right tail is steeper), whereas the probability of a downsizing becomes more likely (the left tail is not a straight line but more convex to the origin). Exporters display a distribution of growth rates more concentrated around the mode, and such difference is more apparent among firms of bigger size (10-19 and bigger than 20 employees).

We now move from graphical analysis to a formal test of the distributional equality between exporters and nonexporters. Table 6 reports the results of the distributional comparison between exporters and nonexporters for each age-size class combination, using the Fligner-Policello (FP) test (see Fligner and Policello, 1981), which does not require neither equal shapes nor symmetry for the comparisons of two distributions. Among firms less than 5 years old, the FP statistic is always positive and significant, meaning that the growth rate distribution of young exporters dominate the growth rate distribution of nonexporters. In the other age classes, the picture is less clear-cut. In general, we observe that, moving to older age classes, exporters dominate over nonexporters in fewer size classes. Among firms aged between 5 and 10 years, the statistics is positive and significant for firms in the first and third size class; in the last two



**Figure 3. Kernel density of employment growth rates for exporters and nonexporters, by age class. Pooled data over 2000–2006**

**Table 6. Fligner–Policello test of stochastic equality**

Age class	1	2-9	10-19	20+
Age <5	28.47 (0.00)	26.65 (0.00)	7.01 (0.00)	6.18 (0.00)
Age 5-10	2.52 (0.00)	-2.57 (0.00)	5.14 (0.00)	15.45 (0.84)
Age 11-20	0.98 (0.33)	-3.21 (0.00)	1.54 (0.12)	16.18 (0.00)
Age 21+	-0.17 (0.86)	-7.28 (0.00)	-2.28 (0.02)	7.32 (0.00)

**Note.** Distributional comparison between growth rates of exporters and nonexporters. Pooled data over 2000-2006. p-value in parentheses.

age classes, the difference is still positive only among big firms (21+). In the remaining cases, the statistics is either not significant or negative (four cases), meaning that nonexporters dominate over exporters.

## 5. Regression analysis

So far we have offered some compelling evidence on the role of firm size, age and the export status. Exporting is a rather rare event, around 9% overall the whole distribution, and around 24% among manufacturing firms, and even more so for young (and small) firms. However the FP test of stochastic equality shows that it is for such category, young firms, that being in the export status exerts the most apparent effect in telling apart the distribution of exporting and non exporting firms. In this section we verify these results in a multivariate regression framework.

### 5.1. Age, export propensity, and firm growth

In this section we use a multivariate regression framework to investigate how export activity and size are related to firm growth within each age class. Our dependent variable,  $g_{i,t}$ , is the growth rate of firm  $i$  at time  $t$ , and is defined, as before, as the log-difference between firm's employment in two consecutive years.

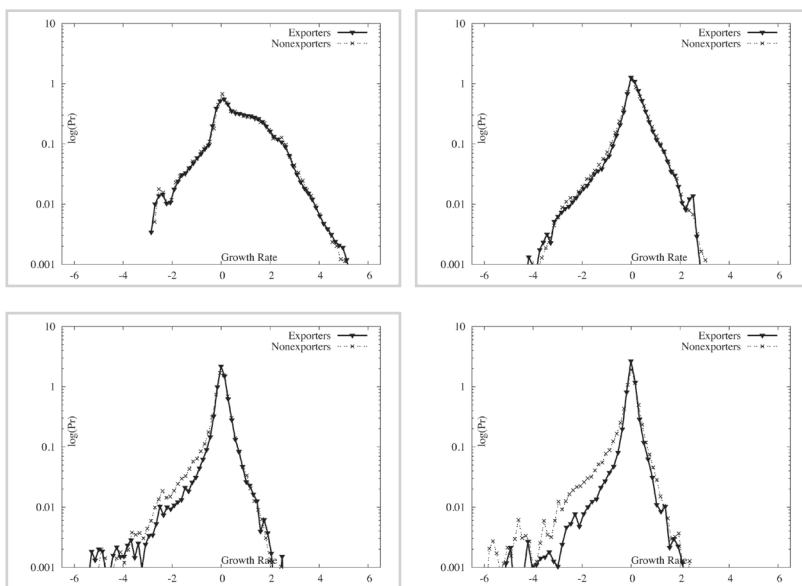
We relate  $g_{i,t}$  to the regressors through the following specification:

$$g_{i,t} = c + \alpha D_{i,t-1}^{\text{exp}} + \beta \ln \text{Size}_{i,t-1} + \gamma (\ln \text{Size}_{i,t-1})^2 + \epsilon_{i,t} \quad (2)$$



where  $D_{i,t-1}^{exp}$  is a binary variable taking value one if the firm exports, and zero otherwise,  $\ln Size_{i,t-1}$  is the (log) firm's employment, which also appears with its square  $(\ln Size_{i,t-1})^2$  to account for non linearity in the relationship (for similar specifications, see Evans, 1987; Lawless, 2014).

Results from OLS estimation of Equation (2) are presented in Table 7, with the first column pooling all firms and taking into account also the role of age ( $\ln Age_{i,t}$ ), and the remaining columns showing the results separately for each age class. Column (1) shows that, controlling for size, firm growth decreases with age and that exporters enjoy, on average, an increase in growth rate of 5.6%. This support our earlier finding that young firms grow more, and also underlines the additional advantage deriving from being an exporter. Columns (2)-(5) test whether the export advantage remains significant when the relationship is estimated within each age class, and the answer is positive: both young and old exporters growth more than nonexporters, controlling for size.



**Figure 4. Kernel density of employment growth rates for young (age <5) exporters and nonexporters, by size class. Pooled data over 2000-2006.**

Table 7. Growth, size and export status					
	(1)	(2)	(3)	(4)	(5)
	All firms	Age < 5	Age 5-10	Age 11-20	Age 21+
$D_{i,t-1}^{exp}$	0.056***	0.050***	0.061***	0.035***	0.029***
	(0.001)	(0.004)	(0.003)	(0.002)	(0.002)
$\ln\text{Size}_{i,t-1}$	-0.280***	-0.466***	-0.302***	-0.225***	-0.124***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
$(\ln\text{Size}_{i,t-1})^2$	0.055***	0.085***	0.063***	0.050***	0.023***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
$\ln\text{Age}_{i,t}$	-0.087***				
	(0.001)				
N	1,599,078	250,455	361,402	440,849	546,372
R <sup>2</sup>	0.155	0.282	0.126	0.084	0.038

**Note.** Dummies for years, 2-digits sectors, and geographical location (provinces) included. Robust standard errors clustered at the firm-level in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01

The coefficient on  $\ln\text{Size}_{i,t-1}$  is negative and significant, whereas that on  $(\ln\text{Size}_{i,t-1})^2$  is positive and significant. This implies that, contrary to the prediction of the Gibrat's law, growth and size are not independent, but firm growth decreases with firm size, and the negative effect of size is lower for older firms. It is worth noticing, moreover, that the magnitude of the coefficients are quite high, implying that an increase of one (log) employee is related to an average decrease in growth rate of around 46% for firms less than five years old.

These results continue to hold when we consider separate cross-sections for each year. In Table 8, we report the coefficients from the estimation of Equation(2) for all firms, and accounting for  $\ln\text{Age}_{i,t}$ , for each of the six years for which we can compute growth rates. It is apparent that the patterns did not change much over time; moreover, the average values of the coefficients across years are strikingly similar to those reported in Table 7, column (1).

Figure 5 plots the binned relationship between size and growth for exporters and nonexporters aged less than 5. In subfigure (a), we consider firms from all size classes. It is apparent that firms which reports full-time employees between 0 and 1 have growth rates that are many times larger than the growth rates of all other firms. This is due to the fact that for firms whose size is close to 0, very small increases in size correspond to very large percentage growth rates. In subfigure (b), we plot the same growth-size relationship ex-

cluding firms from size class 1. The relationship becomes much smoother, and does not seem to be dominated by a single group of firms.

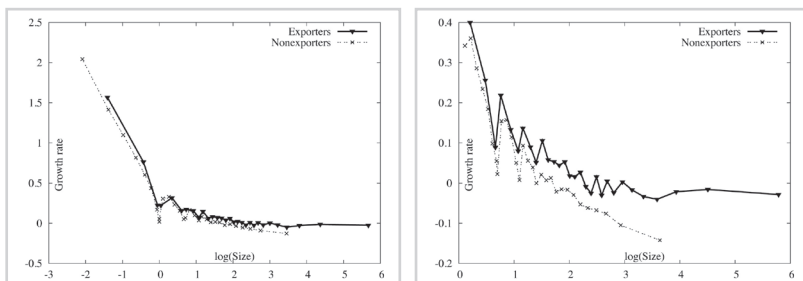
To check whether our results were driven by the micro firms (with employees between 0 and 1), we estimate Equation (2) after excluding firms that were in size class 1 (at time  $t-1$ ). Results are reported in Table 9. Looking at column (1), we observe that both the coefficient on  $\ln Age_{i,t}$  and  $D_{i,t-1}^{exp}$  are still significant and with the expected sign. From columns (2)-(5), we see that exporters grow more across the four age classes, with an advantage which is now markedly declining with age. Greater differences are observed for coefficient on  $\ln Size_{i,t-1}$  and  $(\ln Size_{i,t-1})^2$ . As before, we observe that the effect of size on growth is declining in firm age and size, with point estimates which are now much smaller.

These patterns show that, contrary to recent findings (see Haltiwanger *et al.*, 2013; Lawless, 2014), size impacts on firm growth rates even after taking into account the role of age, and that the Gibrat's hypothesis of independence is still rejected when we consider mature firms ( $Age > 20$ ) and exclude micro firms (employees between 0 and 1). (For an extensive survey of the Gibrat's law see Lotti *et al.*, 2003 and Coad, 2009, chapter 4.)

**Table 8. Growth, size and export status: all firms, by year**

	(1)	(2)	(3)	(4)	(5)	(6)
	2001	2002	2003	2004	2005	2006
$D_{i,t-1}^{exp}$	0.058***	0.068***	0.053***	0.054***	0.060***	0.040***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)
$\ln Size_{i,t-1}$	-0.277***	-0.279***	-0.321***	-0.278***	-0.291***	-0.203***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
$(\ln Size_{i,t-1})^2$	0.054***	0.054***	0.062***	0.056***	0.056***	0.041***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\ln Age_{i,t}$	-0.107***	-0.093***	-0.108***	-0.075***	-0.075***	-0.064***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
N	255,567	263,873	270,889	272,302	269,215	267,232
R <sup>2</sup>	0.168	0.153	0.224	0.139	0.164	0.082

**Note.** Dummies for 2-digits sectors and geographical location (provinces) included. Robust standard errors clustered at the firm-level in parentheses.



**Figure 5. Binned relationship between growth and size for young (age <5) exporters and nonexporters.**

## 5.2. Young exporters and exchange rate shocks

In this section, we take advantage of the finer level of detail that is available for exports over the years 2000-2006 (Section 2. Data sources) and we disaggregate total firm's exports in a given year, in the sum of exports to each product-country destination served by the firm in that year.

Following Bernard *et al.* (2007) and Bernard *et al.* (2015) a firm's total exports to a product-country destination can be decomposed into extensive (quantity) and intensive margins (unit values),<sup>8</sup>

$$\ln X_{fpc} = \ln \text{Quantity}_{fpc} + \ln \text{UnitValue}_{fpc} \quad (3)$$

where  $\ln X_{fpc}$  is the log value (in euro) of exports by firm  $f$  in product  $p$  to country  $c$ ,  $\text{Quantity}_{fpc}$  is the physical quantity of the good  $p$  exported to country  $c$  and  $\text{UnitValue}_{fpc}$  is the unit value of the transaction in that year. This of course results in a larger number of observations which we employ, first, to compare young exporters and older exporters within product-country destinations, and second, to study the sensitivity of young firms and older firms to annual exchange rate movements by considering export value, quantity and unit value.

Using exports to Extra-EU destinations, we estimate the following regression equation:<sup>9</sup>

<sup>8</sup> For a detailed description of the transaction level trade data and the product classification employed refer to Bernard *et al.* (2015).

<sup>9</sup> We focus on exports to Extra-EU destinations for several reasons. Most importantly, firm-level exports to the EU are not recorded for all exporters and these criteria have changed over time. Also, real exchange rate changes within the eurozone countries are driven entirely by changes in relative price levels.

**Table 9: Growth, size and export status: excluding firms in size class 1**

	(1)	(2)	(3)	(4)	(5)
	All firms	Age < 5	Age 5-10	Age 11-20	Age 21+
$D_{i,t-1}^{exp}$	0.036*** (0.001)	0.059*** (0.004)	0.051*** (0.003)	0.026*** (0.002)	0.008*** (0.001)
$\ln Size_{i,t-1}$	-0.060*** (0.001)	-0.235*** (0.005)	-0.070*** (0.003)	-0.03*** (0.002)	0.005*** (0.002)
$(\ln Size_{i,t-1})^2$	0.010*** (0.000)	0.030*** (0.001)	0.011*** (0.001)	0.005*** (0.000)	0.001*** (0.000)
$\ln Age_{i,t}$	-0.053*** (0.001)				
N	1,304,083	180,604	289,651	364,458	469,370
R <sup>2</sup>	0.019	0.040	0.014	0.010	0.008

**Note.** Dummies for years, 2-digits sectors, and geographical location (provinces) included. Robust standard errors clustered at the firm-level in parentheses. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01

$$\ln X_{fpc} = c + \alpha D_f^{young} + \beta \ln Size_f + d_{pc} + \epsilon_{fpc} \quad (4)$$

where  $\ln X_{fpc}$  denotes the logarithm of, respectively, the total value, quantity and unit value of the firms exports in the country-product pair,  $D_f^{young}$  is a dummy for firms with  $Age < 5$ , and  $d_{pc}$  denotes country-product fixed effects. The results in the first column of Table 10 show that young exporters have a substantially lower total value of exports relative to older exporters within product-country pairs. However, the difference in exports across firm types change sign after controlling for firm employment (column (2)): in four years, we do find that young firms are able to record higher exports than older firms. Columns 3-6 report the decomposition of the results as accounted by quantity and unit values. The difference in exports are mainly driven by differences in export quantities; in most cases, unit values are not statistically different for young and older exporters. Notice also that these results are largely unchanged when we control for (log) sales instead of (log) employment for the two years (2000 and 2003) in which this information is available. Moreover, the results continue to hold when we compare young exporters in 2006 with older firms that were already exporting for at least four years.<sup>10</sup>

<sup>10</sup> These additional results are available upon request.

These results give a first evidence about the different behavior of young and well established exporters. We do find that “born-global” firms are able to perform at least as well, and in many cases better than, more experienced firms in terms of exports and quantity sold on foreign markets. Although, we cannot further interpret such result here, this finding is for instance consistent with the evidence of Huergo and Jaumandreu (2004b); Huergo and Jaumandreu (2004a), who show that young firms have a higher propensity to innovate; the higher innovativeness of young firms might in turn impact on their competitiveness (Dosi *et al.*, 2015). Since the analysis presented above about the levels of exports does not allow to control for unobservable characteristics that might be specific to the firm (or to the firm-product varieties), we expand our analysis to correct for such potential source of bias.

In the following analysis, we control for firm, as well as for firm-product specific effects by studying the differential response of firms’ exports to an exogeneous shock, such as real exchange rates variations. In particular, following an emerging stream of empirical literature in international trade we investigate how different firms, in our case born globals and established exporters, respond to the same exogeneous shock within a product-country destination, and how such effects is accounted for by quantity and unit values variations (see among the others, Berman *et al.*, 2012; Chatterjee *et al.*, 2013).

The regression framework that we consider is

$$\Delta \ln X_{fpc,t} = c + \alpha D_{f,young} + \beta \Delta \ln RER_{ct} + \gamma \Delta \ln RER_{ct} * D_{f,young} + d_j + \epsilon_{fpc,t} \quad (5)$$

Table 10: Firm’s exports, quantity and unit value by product and country, by different type of firms						
	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln X_{fpc}$	$\ln X_{fpc}$	$\ln Quantity_{fpc}$	$\ln Quantity_{fpc}$	$\ln UnitValue_{fpc}$	$\ln UnitValue_{fpc}$
<b>2000</b>						
$D_f$ young	-0.123*** (0.017)	0.058*** (0.013)	-0.090*** (0.025)	0.083*** (0.023)	-0.033* (0.017)	-0.024 (0.018)
$\ln Size_f$		0.238*** (0.004)		0.227*** (0.007)		0.012** (0.005)
Country-Product FE	Yes	Yes	Yes	Yes	Yes	Yes
N	956,796	956,796	956,400	956,400	956,400	956,400
Adj R <sup>2</sup>	0.159	0.199	0.401	0.418	0.642	0.642
<b>2001</b>						
$D_f$ young	-0.127*** (0.017)	0.027** (0.013)	-0.119*** (0.026)	0.029 (0.025)	-0.007 (0.019)	-0.001 (0.019)

## Firm growth, export behavior, and the age-size profile

<b>2002</b>						
$D_f^{young}$	-0.136***	0.015	-0.139***	0.007	0.004	0.009
	(0.015)	(0.013)	(0.024)	(0.024)	(0.017)	(0.017)
<b>2003</b>						
$D_f^{young}$	-0.113***	0.041***	-0.115***	0.038	0.002	0.004
	(0.017)	(0.013)	(0.025)	(0.023)	(0.017)	(0.017)
<b>2004</b>						
$D_f^{young}$	-0.079***	0.059***	-0.085***	0.046*	0.006	0.013
	(0.020)	(0.016)	(0.027)	(0.026)	(0.017)	(0.017)
<b>2005</b>						
$D_f^{young}$	-0.089***	0.031	-0.114***	-0.006	0.024	0.036*
	(0.023)	(0.022)	(0.031)	(0.031)	(0.019)	(0.019)
<b>2006</b>						
$D_f^{young}$	-0.113***	-0.003	-0.128***	-0.026	0.016	0.024
	(0.020)	(0.019)	(0.027)	(0.026)	(0.019)	(0.019)

Table reports results of regressions at the firm product country level, using data on exports, quantity and unit value for years 2000-2006.  
Columns (2), (4) and (6) control for log of employment. We report coefficient on  $\ln Size_f$ ,  $N$ , and  $\text{adj } R^2$  only for 2000.  
Robust standard errors clustered at firm level in parenthesis. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

where  $\Delta \ln X_{jpc,t}$  is the change (log difference) in firm-level product-country export value, quantity or unit value,  $Df_{young}$  is a dummy for firms with  $Age < 5$ ,  $\Delta \ln RER_{ct}$  is the change in the log of the real bilateral exchange rate of the Italian currency which is computed with data from the International Financial Statistics database.  $RER_{ct}$  for each year is the product between the nominal Italian exchange rate expressed as the number of foreign currency units per home currency unit ( $ER_{ct}$ ) and the ratio of the domestic consumer price level and the consumer price index abroad ( $CPI_t/CPI_{ct}$ ).<sup>11</sup>  $\Delta \ln RER_{ct} * Df_{young}$  is their interaction, and  $d_j$  a set of fixed effects.

Results in Table 11 show that while the volume of exports to a given product-country destination (Col. 1 and 2) decreases for all firms following a currency appreciation, such reduction is cut by more than half (Col. 2) for young firms, as accounted by the interaction term. Such evidence is much surprising as such firms are new comers, and it also lends support to the category of

<sup>11</sup> We employ the Consumer Price Index as it provides data for many more countries than the Wholesale Price Index.

born globals as firms that are both highly efficient and willing to growth. This result is further reinforced by the analysis of quantities and unit values as the smaller response of firms' exports is actually the result of two opposite effects: a smaller reduction in the quantity sold (Col. 3 and 4), which more than compensate a higher reduction in unit values (Col. 5 and 6).

Overall, the evidence from Table 11 suggests that young exporters react to exogenous exchange rate shocks in such a way as to minimize the losses in terms of exports and quantity growth rates; and they are able to achieve this by lowering their prices more than other established exporters.

## **6. Conclusions**

In this paper we showed how size, age, and export status jointly concur to determine the different patterns of firm growth. In particular, we found that small, young and early exporting firms display a more pronounced attitude towards growth both on the domestic and on the international markets. Further, notwithstanding their much limited experience, born globals also appear to be well equipped to face unfavorable shocks, such as a real exchange rate appreciation.

The growth performance of early exporters supports the idea that the attitude towards growth opportunities is a meaningful and sometimes neglected determinant of firms growth, corroborating the robust evidence that growth processes are also shaped by behavioural factors (as for instance suggested in Zahra et al., 2000; Knight and Cavusgil, 2004; Davidsson and Wiklund, 2013). Indeed, this attitude of firms towards growth is already apparent from the very early stage of firm's life-cycle and it represents one of the 'idiosyncratic' covariates regarding, so to speak, the "identity cards" of individual firms, ideally revealing also their technological and organizational capabilities (more in Dosi and Grazzi, 2006). Being involved in international activities at a very early stage captures both the capabilities of firms to compete on foreign markets and also their intention to "grab" new growth opportunities.



Table 11: Exchange rates and firm's exports, quantity and unit value by product and country over time, by different type of firms

Annual Differences						
	(1)	(2)	(3)	(4)	(5)	(6)
	$\ln X_{f,t}$	$\ln X_{f,t}$	$\ln \text{Quantity}_{f,t}$	$\ln \text{Quantity}_{f,t}$	$\ln \text{UnitValue}_{f,t}$	$\ln \text{UnitValue}_{f,t}$
$D_{f,t}^{\text{young}}$	0.062*** (0.005)		0.063*** (0.005)		-0.000 (0.002)	
$\Delta \ln \text{RER}_{ct}$	-0.283*** (0.088)	-0.342*** (0.104)	-0.250*** (0.092)	-0.314*** (0.108)	-0.033*** (0.010)	-0.028*** (0.010)
$\Delta \ln \text{RER}_{ct} D_{f,t}^{\text{young}}$	0.053 (0.044)	0.186** (0.078)	0.096*** (0.048)	0.226*** (0.084)	-0.043*** (0.016)	-0.040* (0.021)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Product FE	Yes	No	Yes	No	Yes	No
Firm-Product FE	No	Yes	No	Yes	No	Yes
N	2,680,136	2,680,136	2,680,136	2,680,136	2,680,136	2,680,136
Adj R <sup>2</sup>	0.004	-0.005	0.003	-0.007	0.001	0.030

**Note.** Table reports results of regressions at the firm product country level, using data on exports, quantity and unit value between 2000 and 2006. The dependent and independent variables are defined as annual differences. Robust standard errors clustered at country-year level in parenthesis. Year dummies included. \* p<0.10, \*\* p<0.05, \*\*\* p<0.01

## References

- Autio E., Sapienza H. J., Almeida J. G. (2000), "Effects of age at entry, knowledge intensity, and imitability on international growth", *The Academy of Management Journal*, 43, 909–924.
- Axtell R. L. (2001), "Zipf distribution of US firm sizes", *Science*, 293, 1818–1820.
- Barba Navaretti, G., D. Castellani, F. Pieri (2014), "Age and firm growth: evidence from three European countries". *Small Business Economics*, 43, 823–837.
- Berman N., Martin P., Mayer T. (2012), "How do different exporters react to exchange rate changes?" *The Quarterly Journal of Economics*, 127, 437–492.
- Bernard A. B., M. Grazi, C. Tomasi (2015), "Intermediaries in International Trade: Products and Destinations", *The Review of Economics and Statistics*, 97, 916–920.
- Bernard, A. B., Jensen J. B. (2004), "Why some firms export". *The Review of Economics and Statistics*, 86, 561–569.
- Bernard A. B., Jensen J.B. Redding S.J, Schott P.K. (2007), "Firms in international trade". *Journal of Economic Perspectives*, 21, 105–130.
- Birch D.L. (1981), *Who creates jobs*. Public Interest, 65, 3–14.
- Bottazz, G., Cefis E., Dosi G., Secchi A. (2007), "Invariances and diversities in the evolution of Italian manufacturing industry". *Small Business Economics*, 29, 137–159.
- Bottazz G., Coad A., Jacoby N., Secchi A. (2011), "Corporate growth and industrial dynamics: evidence from French manufacturing". *Applied Economics*, 43, 103–116.
- Bottazzi G., Dosi G., Jacoby N., A. Secchi A., Tamagni F. (2010), "Corporate performances and market selection: some comparative evidence". *Industrial and Corporate Change*, 19, 1953–1996.
- Bottazzi G., Secchi A. (2006), "Explaining the distribution of firms growth rates". *RAND Journal of Economics*, 37, 235–256.
- Chatterjee A., Dix-Carneiro R., Vichyanond J. (2013), "Multi-product Firms and Exchange Rate Fluctuations", *American Economic Journal: Economic Policy*, 5, 77–110.
- Coad A. (2007), "A Closer Look at Serial Growth Rate Correlation". *Review of Industrial Organization*, 31, 69–82.
- Coad A. (2009), *The Growth of Firms: A Survey of Theories and Empirical Evidence*, Cheltenham, UK, Edward Elgar.
- Coad A. (2010), "Investigating the exponential age distribution of firms", *Economics - The Open-Access, Open-Assessment E-Journal*, 4, 1–30.
- Coad A., Segarra A., Teruel M. (2013), "Like milk or wine: Does firm performance improve with age?" *Structural Change and Economic Dynamics*, 24, 173 – 189.
- Coad A., Tamvada J. (2012), "Firm growth and barriers to growth among small firms in India", *Small Business Economics*, 39, 383–400.

## Firm growth, export behavior, and the age-size profile

- Davidsson P. (1989), "And After? a study of growth willingness in small firms", *Journal of Business Venturing*, 4, 211–226.
- Davidsson P., Wiklund J. (Eds.) (2013), *New Perspectives On Firm Growth*. Cheltenham, UK, Edward Elgar.
- Davis S. J., Haltiwanger J. (1992), "Gross job creation, gross job destruction, and employment reallocation". *Quarterly Journal of Economics*, 107, 819–863.
- Davis S. J., Haltiwanger J., Jarmin R. S., Krizan C., Miranda J., Nucci A., Sandusky K. (2009), *Measuring the dynamics of young and small businesses: Integrating the employer and nonemployer universes*. In *Producer Dynamics: New Evidence from Micro Data*, NBER Chapters, 329–366. National Bureau of Economic Research.
- Davis S.J., Haltiwanger J.C., Schuh S. (1996), *Job Creation and Destruction*. Cambridge, MA, MIT Press.
- Delmar F., Wiklund J. (2013), *The effect of small business managers' growth motivation on firm growth: A longitudinal study*. In P. Davidsson, J. Wiklund (Eds.), *New Perspectives On Firm Growth*. Cheltenham, UK, Edward Elgar.
- Dosi G., Grazzi M. (2006), *Technologies as problem-solving procedures and technologies as input-output relations: some perspectives on the theory of production*. *Industrial and Corporate Change* 15, 173–202.
- Dosi G., Grazzi M., Moschella M. (2015a), "Technology and costs in international competitiveness: from countries and sectors to firms", *Research policy*, 44, 1795-1814.
- Dosi G., Grazzi M., Tomasi C., Zeli A. (2012), "Turbulence underneath the big calm? The micro-evidence behind Italian productivity dynamics". *Small Business Economics*, 39, 1043–1067.
- Dosi G., Moschella D., Pugliese E., Tamagni F. (2015b), Productivity, market selection and corporate growth: comparative evidence across US and Europe. *Small Business Economics*, 45, 643-672.
- Evans D. S. (1987), "Tests of alternative theories of firm growth". *The Journal of Political Economy*, 95, 657–674.
- Fligner M.A., Policello G.E. (1981), "Robust rank procedures for the Behrens-Fisher problem". *Journal of the American Statistical Association*, 76, 141–206.
- Grazzi M. (2012), "Export and firm performance: Evidence on productivity and profitability of Italian companies". *Journal of Industry, Competition and Trade*, 12, 413–444.
- Grazzi M., Moschella D. (forthcoming). *Small, young, and exporters: New evidence on the determinants of firm growth*. *Journal of Evolutionary Economics*, forthcoming. DOI: <https://doi.org/10.1007/s00191-017-0523-7>.
- Grazzi M., Sanzo R., Secchi A., Zeli A. (2013), "The building process of a new integrated system of business micro-data 1989-2004". *Journal of Economic and Social Measurement*, 38, 291–324.

- Grazzi M., Tomasi C. (2016), "Indirect exporters and importers", *Review of World Economics* (Weltwirtschaftliches Archiv), 152, 251–281.
- Haltiwanger J., Jarmin R.S., Miranda J. (2013), "Who Creates Jobs? Small versus Large versus Young". *The Review of Economics and Statistics*, 95, 347–361.
- Hsieh C.-T., Klenow P.J. (2014), "The life cycle of plants in India and Mexico". *The Quarterly Journal of Economics*, 129, 1035-1084.
- Huergo E., Jaumandreu J. (2004a), "Firms' age, process innovation and productivity growth". *International Journal of Industrial Organization*, 22, 541 – 559.
- Huergo E., Jaumandreu J. (2004b), "How does probability of innovation change with firm age?" *Small Business Economics*, 22, 193–207.
- International Study Group on Exports and Productivity (2008), "Understanding Cross-Country Differences in Exporter Premia: Comparable Evidence for 14 Countries", *Review of World Economics* (Weltwirtschaftliches Archiv) 144, 596–635.
- Jovanovic B. (1982), "Selection and the evolution of industry", *Econometrica*, 50, 649–70.
- Knigh, G.A., Cavusgil S.T. (2004), "Innovation, organizational capabilities, and the born-global firm". *Journal of International Business Studies*, 35, 124–141.
- Lawless M. (2014), "Age or size? contributions to job creation", *Small Business Economics*, 42, 815–830.
- Lotti F., Santarelli E., Vivarelli M.(2003), "Does gibrat's law hold among young, small firms?", *Journal of Evolutionary Economics*, 13, 213–235.
- Melitz M.J. (2003), "The impact of trade on intra-industry reallocations and aggregate industry productivity", *Econometrica*, 71, 1695–1725.
- OECD (2012), *Compendium of Productivity Indicators*, OECD Publishing.
- Penrose E. . (1959), *The theory of the growth of the firm* (3rd ed.). Oxford, Blackwell.
- Roberts M. J., Tybout J.R.(1997), "The decision to export in Colombia: An empirical model of entry with sunk costs". *American Economic Review*, 87, 545–64.
- Serti F., Tomasi C. (2008), "Self-selection and post-entry effects of exports: Evidence from Italian manufacturing firms", *Review of World Economics* (Weltwirtschaftliches Archiv) 144, 660–694.
- Stanley M., Amaral L., Buldyrev S., Havlin S., Leschhorn H, Maass P., Salinger M., Stanley H. (1996), "Scaling behaviour in the growth of companies", *Nature*, 379, 804–806.
- Wagner J. (2001), "A note on the firm size - export relationship". *Small Business Economics*, 17, 229–237.
- Zahr, S A., Ireland R D., Hitt M.A. (2000), "International expansion by new venture firms: International diversity, mode of market entry, technological learning, and performance", *The Academy of Management Journal* 43, 925–950.

# Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica

ANNAMARIA NIFO, SABRINA RUBERTO, GAETANO VECCHIONE

## 1. Introduzione

Negli ultimi anni la letteratura economica ha dato grande attenzione all'ampio ricorso al multiaffidamento bancario. In quasi tutti i paesi, imprese anche relativamente piccole, prendono a prestito da più di una banca contemporaneamente, anche se la distribuzione del numero di relazioni bancarie per impresa varia molto tra i paesi. Ongena e Smith (2000), usando un dataset di 1079 grandi imprese di 20 diversi paesi europei, documentano che le single-bank relationships sono relativamente rare e che l'Italia – con un numero medio di 15 relazioni bancarie per impresa – è il paese in cui il fenomeno del multiaffidamento è più diffuso. Un dato confermato da Detragiache et al. (2000) che comparano campioni di piccole imprese operanti negli USA e in Italia mostrando che mentre il single banking è relativamente comune negli Stati Uniti dove il numero mediano di relazioni bancarie è 2 e il 55,5% delle imprese negozia con più di una banca, in Italia l'89% delle imprese ha relazioni bancarie multiple, con un numero mediano di relazioni per impresa pari a 5.

La letteratura economica ha largamente indagato sulle ragioni alla base delle diverse preferenze delle imprese sul numero di relazioni bancarie. Alcuni contributi hanno focalizzato l'attenzione sugli aspetti micro-economici alla base delle scelte delle imprese quali la dimensione, l'età, la

propensione ad innovare, la dotazione di capitale umano, l'investimento in ricerca e sviluppo, e così via. In sintesi, studi teorici ed empirici suggeriscono che imprese più grandi e più mature, maggiormente innovative e finanziariamente in difficoltà (Haroff e Korting, 1998a) ricorrono con più probabilità al multiaffidamento bancario, tipicamente associato ad imprese più rischiose (Foglia et al., 1998).

D'altro canto è stato anche ampiamente documentato che spesso fattori macroeconomici sono ugualmente importanti: per esempio la specializzazione produttiva regionale, la diffusione della tecnologia, il grado di competizione del mercato, e fattori istituzionali ampiamente riconosciuti come rilevanti nell'influenzare le preferenze delle imprese. In particolare, la letteratura ha largamente riconosciuto il ruolo delle istituzioni nell'influenzare i comportamenti delle imprese nei mercati finanziari (Chinn e Ito, 2006; Claessens e Leaven, 2003; Garrettsena et al., 2004; Andrianova e Demetriades, 2004), sebbene perlopiù attraverso analisi cross-country e riferite alla dotazione istituzionale del paese. In linea con i recenti sviluppi della letteratura, questo paper adotta un approccio che enfatizza in particolare il link tra qualità delle istituzioni locali, a livello provinciale, e preferenze delle imprese sul numero di relazioni bancarie.

Un ricco filone della letteratura ha riconosciuto l'influenza delle istituzioni locali e del contesto geografico in cui sono localizzate, sulle scelte delle piccole e imprese (Acemoglu e Dell, 2010; La Rocca et al., 2010; Cheng e Shiu, 2007; Guiso et al., 2004; Pollard, 2003; Demirguk-Kunt e Maksimovich, 1998, 1999) sottolineando come contesti caratterizzati da asimmetrie informative, mancanza di trasparenza, maggiore incertezza, corruzione, eccessiva burocratizzazione, scarsa sicurezza e debole controllo del territorio da parte delle forze dell'ordine, minore dotazione di infrastrutture, un insoddisfacente ambiente socio-culturale – tutti fattori tipicamente connessi ad una scarsa qualità istituzionale - generano contratti incompleti, incoraggiano comportamenti opportunistici ed accrescono il grado di rischiosità dei contratti favorendo il ricorso delle imprese ad un numero crescente di relazioni bancarie.

Affrontare la questione della scelta del numero di relazioni bancarie per impresa in Italia ha forti motivazioni nell'evidenza del durevole gap economico e istituzionale tra Mezzogiorno e il resto del paese. Le ampie differenze osservate tra le regioni italiane nei livelli di qualità istituzionale si sovrappongono con l'evidenza di altrettanto ampi divari in numerosi indicatori economici e sociali (Malanima e Zamagni, 2010; Giannola et al., 2016), fornendo evidenza della variegata natura del ritardo del Mezzo-

giorno e confermando che anche ad un livello sub nazionale, differenze nelle performance delle imprese possono essere spiegate sulla base di differenze istituzionali (Del Monte e Giannola, 1997; Scalera e Zazzaro, 2010; Erbetta e Petraglia, 2011; Nifo, 2011; Aiello et al., 2014). In particolare, contrariamente alla crescente integrazione del sistema finanziario nazionale, la sua efficienza a livello locale è molto diversa tra le regioni (Guiso et al., 2004; Giordano et al., 2013) e, sebbene si applichi lo stesso sistema normativo a tutto il paese, l'applicazione di tali regole differisce notevolmente a livello locale (Bianco et al., 2005).

E, mentre negli ultimi anni una crescente letteratura ha dedicato grande attenzione all'impatto della qualità istituzionale su vari indicatori di performance economica (Aiello, Pupo e Ricotta, 2014; Ganau e Rodriguez-Pose, 2016; Mannarino et al., 2016; Di Liberto e Sideri, 2015; Lasagni et al., 2015; Nerozzi et al., 2015; Raspe e Van Oort, 2011; Fazio e Piacentino, 2010), il ruolo dell'*institutional quality* nelle scelte finanziarie delle imprese è rimasto quasi inesplorato. Tra le rilevanti eccezioni, Sarno (2009) analizza le relazioni tra grado di *enforcement* a livello provinciale ed il funzionamento del sistema finanziario, confermando il ruolo delle istituzioni regionali e provinciali nel determinare le scelte delle imprese e lo sviluppo locale. Anche La Rocca et al. (2010) spiegano come lo sviluppo finanziario locale e le connesse differenze istituzionali influenzano le decisioni finanziarie delle piccole e medie imprese italiane. In linea con questa impostazione, Agostino et al. (2010) mostrano come migliori istituzioni locali creino un favorevole *business environment* e una struttura legale capace di promuovere una più efficace protezione del credito, che permette alle imprese di avere più facile accesso al credito e agli intermediari finanziari di essere più inclini ad erogarlo. Ad analoghe conclusioni giungono anche Ferri e Messori (2000) i quali mostrano come differenze geografiche nelle strutture produttive e socio-economiche tra le regioni italiane sono accompagnate da analoghe differenze nel numero di relazioni bancarie scelte dalle imprese: nel Mezzogiorno, dove prevalgono imprese e banche di minori dimensioni, emerge una più stretta e durevole relazione tra numero di relazioni bancarie e la struttura socio-economica della regione. Anche Cosci e Meliciani (2002) e Elsas (2004) trovano che un *business environment* più rischioso favorisce il ricorso delle imprese a relazioni bancarie multiple.

Il presente lavoro si inserisce in questo filone della letteratura, con l'obiettivo di stimare il peso della qualità delle istituzioni provinciali misurate dall'*Institutional Quality Index* (IQI) costruito da Nifo e Vecchione

(2014, 2015), come determinante del numero di relazioni bancarie scelto dalle piccole e medie imprese manifatturiere italiane. L'IQI è un indicatore composito derivato da 24 indici elementari raggruppati in 5 dimensioni delle istituzioni (corruzione, *government effectiveness*, qualità della regolamentazione, certezza del diritto, *voice and accountability*) che rappresenta una misura sintetica della qualità istituzionale nelle province e nelle regioni italiane.

Per condurre l'analisi econometrica, è stato costruito un panel sbilanciato di 5,137 imprese manifatturiere osservate nel periodo 2003-2006, per un totale di 16,460 osservazioni, utilizzando i dati di bilancio ricavati dalla nona e decima "Indagine sulle Imprese Manifatturiere" condotta da Uni-Credit-Capitalia, ed informazioni sul numero di sportelli bancari e sulla popolazione forniti rispettivamente da Banca d'Italia e ISTAT. Le stime sono state condotte con l'impiego di differenti metodi di stima: Probit, Poisson, Arellano e Bover (1995), Blundell e Bond (1998) System Generalised Method of Moments (*System GMM*), al fine di controllare per l'eterogeneità non osservata e la potenziale endogeneità.

Nelle differenti specificazioni dei modelli e controllando per le caratteristiche micro delle imprese e le variabili di contesto che possono condizionare le performance delle imprese, coerentemente con altri studi i risultati confermano che le istituzioni contano ed influenzano negativamente il numero di relazioni bancarie. In particolare le imprese localizzate in province del Mezzogiorno che sistematicamente mostrano una più povera dotazione istituzionale, registrano sempre un numero di relazioni bancarie relativamente più alto rispetto a quelle del Centro-Nord.

Il resto del paper è organizzato come segue. Nella Sezione 2 è presentata la metodologia usata per l'indagine empirica. In particolare, la sezione 2.1 presenta il modello; la Sezione 2.2 focalizza sulle variabile esplicative, i controlli (2.2.1) e l'indice IQI (2.2.2). La Sezione 2.3 illustra il dataset e alcune statistiche descrittive. La Sezione 3 presenta i principali risultati dell'analisi empirica e le conclusioni sono discusse nella Sezione 4.

## 2. Metodologia

Questa sezione è dedicata ad evidenziare i fattori che influenzano la scelta dell'impresa di multiaffidarsi, e specificamente ad isolare il ruolo della qualità istituzionale provinciale nel determinare questa scelta.



Per realizzare questo obiettivo, si sviluppa un'analisi econometrica, in cui la probabilità di ricorrere al multiaffidamento ed il numero di relazioni bancarie vengono poste in relazione con una serie di variabili esplicative, riconducibili alle caratteristiche individuali dell'impresa, alle caratteristiche del rapporto banca-impresa, alle caratteristiche del sistema bancario locale e alla qualità istituzionale. L'analisi mostra che un miglioramento istituzionale comporta una diminuzione della propensione delle imprese a multiaffidarsi, segnalando anche che la qualità istituzionale influenza negativamente il numero di relazioni bancarie. Per mirare propriamente alle problematiche sull'eterogeneità e potenziale endogeneità di alcuni regressori, si adottano alternativamente diversi metodi di stima.

### 2.1. Strategie di stima

La decisione dell'impresa di multiaffidarsi può essere analizzata attraverso molteplici modelli di stima. Prima di tutto, può essere vista come una scelta dicotomica (se multiaffidarsi o meno), appropriatamente modellata con un modello a risposta binaria. Alternativamente, considerando il numero di relazioni bancarie instaurate da un'impresa come una variabile di conteggio, un altro metodo di stima adatto potrebbe essere un modello per dati di conteggio, come quello Poisson. Inoltre, poiché la variabile dipendente tende ad essere inerte nel tempo (è probabile che il numero di relazioni bancarie avute in passato influenzi il numero presente delle stesse), il modello *System Generalised Method of Moments* (SYS-GMM) si propone anch'esso come un modello appropriato, permettendo anche di controllare per l'eterogeneità non osservata e per la presenza di variabili esplicative endogene (o predeterminate). Nel presente studio, si adottano tutti e tre i metodi sopra menzionati stimando le seguenti equazioni:

$$P(y_{it} = 1 | X) = \Phi \left( \alpha + \beta_1 IQI_{jt} + \gamma X'_{it} + \sum_s \delta_s S_s + \sum_j \gamma_j P_j + \sum_t \varphi_t T_t \right) \quad (1)$$

$$NBANK_{it} = \alpha + \beta_1 IQI_{jt} + \gamma X'_{it} + \sum_s \delta_s S_s + \sum_j \gamma_j P_j + \sum_t \varphi_t T_t + v_{ijt} \quad (2)$$

$$NBANK_{it} = \alpha + \beta_0 NBANK_{i,(t-1)} + \beta_1 IQI_{jt} + \gamma X'_{it} + \sum_s \delta_s S_s + \sum_j \gamma_j P_j + \sum_t \varphi_t T_t + v_{ijt} \quad (3)$$

dove gli indici  $i$ ,  $j$  e  $t$  si riferiscono rispettivamente alle imprese, alle province e al tempo. Nell'equazione (1), si adotta un modello Probit: la variabile dipendente  $y_{ijt}$  è una *dummy* che assume valore 1 se l'impresa  $i$  si situa nella provincia  $j$  al tempo  $t$  mantiene un numero di relazioni bancarie

maggiore o uguale di due (e o altrimenti), ed  $\Phi$  è la funzione di ripartizione della variabile casuale normale standardizzata.<sup>1</sup>

Nei modelli (2) e (3), la variabile dipendente NBANK è il numero di relazioni bancarie per impresa. Per stimare le equazioni (2) e (3), si adottano rispettivamente lo stimatore Poisson e *system* GMM (SYS-GMM) di Arellano e Bover (1995) e Blundell e Bond (1998).

Sul lato destro delle equazioni (1), (2) e (3), come principali variabili esplicative si considerano alternativamente l'indice di qualità istituzionale a livello provinciale (IQI), quello a livello regionale (IQI\_REG), e poi i sotto indici dell'IQI provinciale e regionale al posto degli indici complessivi. Il vettore X contiene le variabili esplicative di controllo presentate nelle successive sottosezioni. In tutte le equazioni, T, S e P rappresentano rispettivamente gli effetti fissi temporali, settoriali e provinciali, mentre per le equazioni (2) e (3),  $v_{ijt} = \eta_i + w_j + e_{it}$  è un errore composito, in cui  $\eta_i$  e  $w_j$  catturano rispettivamente le caratteristiche inosservate dell'impresa costanti nel tempo e gli effetti fissi provinciali, mentre  $e_{it}$  rappresenta la parte idiosincratca. I risultati delle stime delle equazioni (1), (2) e (3) sono commentati nella seguente sezione 3. Come si potrà osservare, appaiono robusti alla scelta del metodo di stima.

## 2.2. Le variabili esplicative

Le variabili esplicative impiegate nelle stime, forniscono informazioni riguardanti: i) le caratteristiche individuali dell'impresa e del rapporto banca-impresa, come dimensione, età, indebitamento, razionamento del credito, durata della relazione e percentuale di debito detenuto dalla banca principale; ii) le condizioni macroeconomiche locali, come il numero di filiali bancarie in rapporto alla popolazione totale; iii) la qualità istituzionale provinciale e regionale osservata in termini del valore sia dell'IQI che delle sue singole dimensioni.

### 2.2.1. Controlli

In questo paragrafo si illustrano brevemente le determinanti del multifidamento bancario precedentemente indagate in letteratura che entrano

<sup>1</sup> Si considerano multifidate tutte le imprese che mantengono un numero di relazioni bancarie maggiore o uguale di due, che approssimativamente corrisponde al decimo percentile della distribuzione della variabile numero di relazioni bancarie nel campione in esame.

come variabili di controllo nelle equazioni da stimare (racchiuse nel vettore  $X$ ), accanto alla *key variable* (IQI).

Tra le caratteristiche dell'impresa, si inserisce la dimensione dell'impresa misurata con il numero di dipendenti (EMP). Alcuni studi evidenziano un impatto positivo della dimensione dell'impresa sul multiaffidamento bancario. Questo perché, da un lato, le banche diversificando il rischio di credito potrebbero indurre le grandi imprese ad intraprendere molteplici rapporti bancari (Detragiache et al., 2000; Pelliccioni e Torlucio, 2007), e da un altro lato, le piccole imprese tenderebbero ad evitare il multiaffidamento a causa dell'esistenza di costi fissi della relazione di credito (Guiso e Minetti, 2004).

Come altra caratteristica dell'impresa, viene inclusa la variabile età (AGE), spesso considerata una *proxy* della trasparenza informativa. Le argomentazioni sull'effetto di questa variabile sono discordanti. Alcuni studi argomentano che le imprese mature, sopravvivendo alla fase critica di start-up e avendo un *track record*, sono meno opache e potenzialmente in grado di ottenere credito a basso costo da un maggior numero di banche (Diamond, 1991). Al contrario, altri studiosi affermano che essendo meno soggette ad una selezione avversa, le imprese mature preferirebbero mantenere un minor numero di relazioni bancarie (Detragiache et al., 2000).

Si considerano inoltre altre variabili: due variabili *dummy* indicatrici di innovazioni di prodotto e/o di processo ed innovazioni organizzative connesse ad innovazioni di prodotto e/o di processo realizzate dall'impresa (rispettivamente indicate come INPP, INORG), una variabile *dummy* (HT) indicatrice dell'appartenenza a settori Hi-Tech, ed un'altra variabile calcolata come il rapporto tra risorse intangibili e attività totali (INTAS). Secondo Elsas (2004), l'attitudine dell'impresa ad innovare è una *proxy* della trasparenza informativa. Le imprese più innovative tenderebbero a preferire *close lending relationship* per evitare la diffusione di informazioni private ai diretti concorrenti (Yosha, 1995). D'altro canto, potrebbero preferire di multiaffidarsi per evitare il problema *hold up*<sup>2</sup>. Per le imprese che operano in settori Hi-Tech e per quelle con una percentuale più elevata di attività immateriali rispetto a quelle totali, il multiaffidamento potreb-

<sup>2</sup> Uno stretto rapporto con la *main bank* nel corso di una *close lending relationship*, potrebbe comportare la "cattura" informativa dell'impresa (*hold up*) da parte della banca che traendo vantaggio dell'informazione esclusiva sull'impresa e facendo leva sul suo potere di contrattazione, potrebbe praticare tassi d'interesse più elevati di quelli corrispondenti al reale merito creditizio dell'impresa (Sharpe 1990, Rajan, 1992).

be essere più probabile anche a causa della propensione delle banche di effettuare una maggiore diversificazione del credito verso mutuatari rischiosi ed opachi (Pelliccioni e Torluccio, 2007).

Inoltre, si considera la variabile indebitamento dell'impresa (LEVER), calcolata come il rapporto tra passività finanziarie e la somma tra passività. Le imprese più indebitate tenderebbero ad essere multiaffidate sia per un problema di selezione avversa, più severo per loro che per altre imprese (Detragiache et al., 2000), sia per la scelta delle banche di finanziare piccole quote di debito al fine di diversificare i loro rischi (Cosci e Meliciani, 2002; Carletti et al., 2004).

Per quanto riguarda le variabili catturanti le caratteristiche del rapporto banca-impresa, vengono incluse: una variabile *dummy* CRED, *proxy* del razionamento del credito, la durata della relazione bancaria (DURAT) e la quota di indebitamento detenuta dalla banca principale (MAIN). Al fine di minimizzare il rischio di essere razionate dal credito, le imprese potrebbero essere propense a multiaffidarsi (Detragiache et al., 2000); la durata della relazione bancaria e la quota di indebitamento detenuta dalla banca principale potrebbero essere variabili rilevanti, tenendo conto che, da un lato, in un rapporto singolo le asimmetrie informative potrebbero essere mitigate, e dall'altro lato, che un forte potere contrattuale della *main bank* potrebbe comportare tassi di interesse più elevati (Sharpe, 1990; Rajan, 1992).

Infine, si tiene conto dello sviluppo del mercato locale del credito includendo le variabile BRANCH (numero di filiali bancarie sul totale della popolazione). L'impatto della variabile BRANCH è ambiguo: se la presenza di nuove banche nel mercato del credito provinciale induce migliori attività di screening e di monitoraggio, e quindi, una migliore raccolta di *soft information* da parte degli intermediari (Benfratello et al., 2008), il multiaffidamento potrebbe concretizzarsi, d'altra parte, è anche vero che una maggiore vicinanza con l'impresa potrebbe essere motivo di maggiore potere di mercato della banca affidante, con conseguente applicazione di tassi di interesse più elevati (*hold up problem*).

Come altre variabili esplicative di controllo, si considerano variabili che catturano le caratteristiche osservabili e specifiche dell'impresa. Innanzitutto, si inseriscono due variabili *dummy* indicatrici dell'appartenenza dell'impresa ad un gruppo o/e ad un consorzio (rispettivamente, GROUP e CONS). L'appartenenza dell'impresa ad un gruppo o/e ad un consorzio potrebbe comportare un minor numero di relazioni bancarie, grazie all'opportunità di ricevere credito dagli altri membri, o beneficiare di una *main bank* che finanzia tutte le

imprese del gruppo/consorzio (Detragiache et al., 2000). Secondariamente, si include la variabile *dummy* COOP al fine di verificare se le imprese cooperative detengono un numero più basso di relazioni bancarie, essendo generalmente finanziate da banche cooperative e popolari, con le quali intrattengono stretti rapporti bancari (Ferri e Messori, 2000; Cosci e Meliciani, 2005). Inoltre, le imprese internazionalizzate possono aver bisogno di maggior numero di affidamenti per gestire le loro transazioni all'estero. Quindi, viene inserita la variabile *dummy* EXP codificata 1 se l'impresa esporta i suoi prodotti all'estero e 0 altrimenti. In aggiunta, per verificare se le imprese che hanno maggior liquidità mantengono un minor numero di relazioni bancarie, si inserisce la variabile QUICK, definita come il rapporto tra attività correnti e passività correnti. Infine, tutte le stime includono *dummy* settoriali per controllare l'eterogeneità a livello settoriale (classificazione Ateco 2-digit).

Le variabili esplicative inserite nell'analisi econometrica sono elencate nella seguente Tabella 1, che riporta anche le principali statistiche descrittive. La Tabella 2 riporta la matrice di correlazione tra il numero di relazioni bancarie (NBANK), l'IQI provinciale e regionale e le singole dimensioni dell'indice provinciale.

**Tab. 1. Statistiche descrittive**

Variabili	Descrizione	Years	Obs	Mean	Std.	Min	Max
NBANK	Numero di relazioni bancarie per impresa	2003-2006	14433	4784	2.986	0	15
EMP	Numero di dipendenti dell'impresa	2003-2006	14862	45.399	45.124	0	250
AGE	Anno corrente- anno di fondazione (in anni)	2003-2006	14981	25.624	19.531	0	110
INPP	Dummy = 1 se l'impresa ha realizzato innovazioni di prodotto /processo, o altrimenti	2003-2006	15250	.583	.493	0	1
INORG	Dummy = 1 se l'impresa ha realizzato innovazioni organizzative, o altrimenti	2003-2006	15250	.172	.378	0	1
HT	Dummy = 1 se l'impresa appartiene ai settori Hi-Tech, o altrimenti	2003-2006	15254	.043	.203	0	1
INTAS	Attività immateriali/ Totale Attività (in %)	2003-2006	14994	2.367	4.331	0	25.45
LEVER	Passività finanziarie/(Passività finanziarie+ patrimonio netto)(in %)	2003-2006	14994	27.605	32.643	0	96.39
QUICK	Attività correnti/ Passività correnti	2003-2006	14990	1.075	.939	.233	21.57
GROUP	Dummy = 1 se l'impresa appartiene ad un gruppo, o altrimenti	2003-2006	15250	.172	.377	0	1
CONS	Dummy = 1 se l'impresa appartiene ad un consorzio, o altrimenti	2003-2006	15133	.038	.192	0	1
COOP	Dummy = 1 se l'impresa ha una forma giuridica cooperativa, o altrimenti	2003-2006	15107	.012	.111	0	1

Firm's characteristics

Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica

Bank-firm relationships		2003-2006	15245	.620	.485	0	1
EXP	Dummy =1 se l'impresa ha esportato parte dei suoi prodotti all'estero, o altrimenti	2003-2006	15245	.620	.485	0	1
CRED	Dummy =1 l'impresa avrebbe desiderato una maggiore quantità di credito al tasso d'interesse concordato dalla banca, o altrimenti.	2003-2006	12755	.059	.237	0	1
DURAT	Durata rapporto bancario con la main bank (in anni)	2003,2006	12054	15.999	11.422	0	53
MAIN	Quota di indebitamento detenuto dalla banca principale (in %)	2003,2006	9649	24.495	24.402	0	100
BRANCH	Totale numero di sportelli bancari per provincia/ popolazione provinciale in migliaia.	2003-2006	15254	6.433	1.473	2.193	10.49
RGDPC	GDP provinciale (per capita) in migliaia di €	2003-2006	15254	2021737	4033.258	9086.10	27414.37
IQI	Institutional quality index at the provincial level	2004-2006	14368	.711	.148	0	1
IQI_REG	Institutional quality index at the regional level	2004-2006	14368	.709	.138	.0973	.932
RULAW	IQI Dimension, Rule of Law at the provincial level	2004-2006	14368	.590	.164	0	1
GOVERN	IQI Dimension, Government at the provincial level	2004-2006	14368	.422	.133	0	1
REGUL	IQI Dimension, Regulatory Quality at the provincial level	2004-2006	14368	.620	.173	0	1
VOICE	IQI Dimension, Voice & Accountability at the provincial level	2004-2006	14368	.505	.218	0	1
CORR	IQI Dimension, Corruption at the provincial level	2004-2006	14368	.849	.142	0	1
Context characteristics							

Tab. 2. Matrice di Correlazione

	NBANK	IQI_REG	IQI	CORR	GOVERN	REGUL	RULAW	VOICE
NBANK	1							
IQI_REG	0.046	1						
IQI	0.041	0.926	1					
CORR	0.053	0.475	0.435	1				
GOVERN	0.026	0.617	0.699	0.153	1			
REGUL	0.050	0.637	0.627	0.510	0.333	1		
RULAW	0.011	0.353	0.349	0.325	-0.158	0.047	1	
VOICE	0.010	0.404	0.486	-0.094	0.405	0.372	-0.405	1

### 2.2.2. L'insititutional quality index (IQI)

Per valutare l'impatto della qualità istituzionale sul multiaffidamento, si inserisce l'*Institutional Quality Index* (IQI) recentemente proposto da Nifo e Vecchione (2014, 2015), costruito su base annuale e provinciale (NUTS3) e disponibile per 110 province italiane per gli anni che vanno dal 2004 al 2008<sup>3</sup>. Basato sul *framework* del *World Governance Indicator* (WGI) proposto dalla Banca Mondiale, l'IQI è strutturato su 5 dimensioni: il grado di libertà di stampa e di associazione (*Voice and Accountability*), la qualità del servizio pubblico e le politiche formulate ed implementate dalle amministrazioni locali (*Government Effectiveness*), le capacità del governo di promuovere e formulare efficaci interventi normativi (*Regulatory Quality*), la percezione relativa all'applicazione della legge in termini di adempimento del contratto, di diritti di proprietà, di forze di polizia, di attività della magistratura e di criminalità (*Rule of Law*), il grado di corruzione di coloro che esercitano funzioni pubbliche, sia in termini di guadagni illegali e proventi privati acquisiti a scapito della società (*Control and Corruption*)<sup>4</sup>. Ogni dimensione è a sua volta il risultato dell'aggregazione di numerosi indici semplici raccolti da fonti ufficiali ed indagini condotte da istituzioni pubbliche, private e non governative. Il risultato è un indice sintetico compreso tra zero ed uno, ottenuto da tecniche statistiche che prima

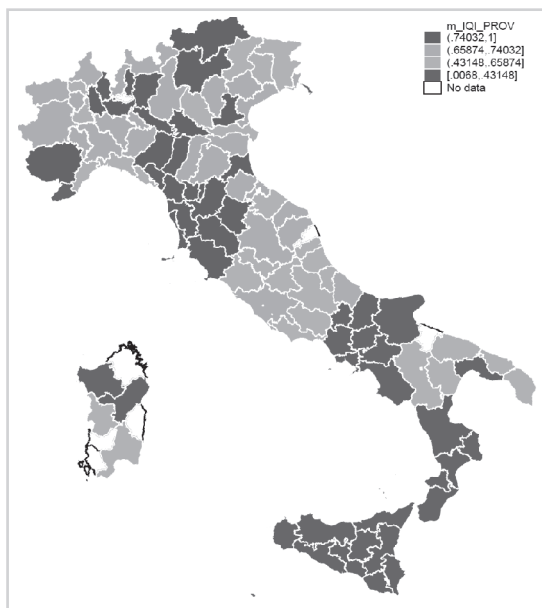
<sup>3</sup> Per dettagli sull'indice IQI, vedere Nifo e Vecchione (2014) mentre <https://sites.google.com/site/institutionalqualityindex/home> per scaricare il dataset.

<sup>4</sup> Kaufmann D., Kraay A. e Mastruzzi M. (2010), *The worldwide governance indicators: methodology and analytical issues*, Policy Research Working Paper Series 5430, The World Bank.



normalizzano, poi pesano ed infine sintetizzano ben 24 indicatori elementari<sup>5</sup>. La Figura 1 mostra i valori medi annui dell'IQI dal 2004 al 2008 per macroaree e l'ampiezza dei divari. Il gap Nord-Sud che, come è noto, attiene ad un'ampia gamma di condizioni socio-economiche, si ripropone drammaticamente anche per quanto riguarda la qualità delle istituzioni: tutte le regioni del Mezzogiorno si caratterizzano per livelli di qualità istituzionale più bassi del resto d'Italia.

**Fig. 1. IQI delle province italiane, valori medi 2004-2008.**



<sup>5</sup> Gli indici semplici sono 24 così suddivisi per ciascuna dimensione: 1) *Voice and Accountability*: Associazioni, Partecipazione votazioni, Libri pubblicati, Acquisti in libreria; 2) *Government Effectiveness*: Dotazione strutture sociali, Dotazione strutture economiche, Deficit sanitario regionale, Raccolta differenziata, Indice ambiente urbano; 3) *Regulatory Quality*: Apertura economia, Dipendenti pubblici locali, Densità imprenditoriale, Mortalità imprese, Qualità della vita imprese; 4) *Rule of Law*: Delitti denunciati, Tempi processi, Produttività magistratura, Economia sommersa, Evasione fiscale; 5) *Corruption*: Reati contro la PA, Golden-Picci Index, Comuni Commissariati.

### 2.3. Dati

I dati utilizzati in questo lavoro provengono da diverse fonti. La prima fonte dati è costituita dalla nona e decima *Indagine sulle Imprese Manifatturiere* italiane condotte da Unicredit - Capitalia. Ciascuna edizione dell'indagine è condotta su un arco temporale di tre anni: la nona indagine, condotta per il periodo 2001-2003 fornisce dati su 4,289 imprese, mentre la decima riporta dati su 4,126 imprese per il periodo 2004-2006. Queste indagini forniscono informazioni sia qualitative che quantitative, tra queste, l'anno di costituzione, l'appartenenza ad un gruppo, la dimensione, la forma giuridica dell'impresa, il numero e la durata delle relazioni bancarie<sup>6</sup>. Capitalia inoltre fornisce dati di bilancio sulle imprese intervistate, per lo stesso lasso di tempo delle *survey*. Incrociando i dati di *survey* con quelli di bilancio, si ottiene un *panel* non bilanciato di 5,137 imprese e 16,460 osservazioni, che costituisce il dataset impiegato nelle stime econometriche. L'analisi empirica si focalizza sulle PMI del settore manifatturiero italiano, per le quali i prestiti bancari costituiscono la maggior fonte di finanziamento (Banca d'Italia, 2007), per cui vengono eliminate 240 imprese dal campione, ossia quelle più grandi (con più di 250 lavoratori) e quelle quotate in Borsa.

Le informazioni sul numero di sportelli bancari e sulla popolazione, entrambe a livello provinciale, sono forniti rispettivamente da Banca d'Italia e ISTAT (Istituto Nazionale di Statistica). Infine, si utilizza l'*Institutional Quality Index* (IQI) recentemente proposto da Nifo e Vecchione (2014, 2015), descritto nel paragrafo precedente, per i dati sulla qualità istituzionale locale (provinciale e regionale).

### 3. Risultati

La tabella 3 riporta i risultati ottenuti stimando il modello (1), il (2) ed il (3) utilizzando rispettivamente gli stimatori Probit, Poisson e SYS-GMM. Per i modelli Probit e Poisson, sono riportati gli effetti marginali (per entrambi i modelli *pooled* e *panel*), in modo da avere una più immediata interpretazione

<sup>6</sup> Siccome le informazioni sul numero di banche (NBANK), la durata del rapporto con la banca principale (DURAT), e la quota di indebitamento detenuto dalla stessa (MAIN) sono disponibili solo per l'ultimo anno di ogni indagine, si assegna lo stesso valore ai due anni precedenti. In presenza di valori mancanti o inconsistenti, seguendo Gambini e Zazzaro (2010) e Agostino et al. (2012), si imputano valori idonei per DURAT, prendendo il valore riportato per l'ultimo anno della prima indagine (2000) e aggiungendo il numero 1 per 2001, il numero 2 per 2002 e così via.

Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica

dei risultati<sup>7</sup>. Per le regressioni SYS-GMM sono riportati i coefficienti stimati<sup>8</sup>. Tutte le regressioni sono condotte includendo gli effetti fissi temporali, settoriali e provinciali. Gli *standard-errors* sono clusterizzati a livello della singola provincia (NUTS3) e robusti rispetto alla potenziale eteroschedasticità degli errori.

I risultati della Tabella 3 colonna 1 mostrano la variabile d'interesse IQI (la qualità istituzionale a livello provinciale) negativa e statisticamente significativa per maggior parte dei modelli. Il segno negativo segnala che una migliore qualità istituzionale tende a diminuire sia la propensione delle imprese a multiaffidarsi sia il loro numero di relazioni bancarie.

Osservando le variabili esplicative di controllo, si riscontra che nella maggior parte dei casi, i coefficienti delle variabili EMP, AGE, LEVER, QUICK, HT, INPP e EXP sono statisticamente significativi per un livello di significatività dell' 1%. Per quanto concerne le variabili INTAS, INORG, MAIN, CRED e BRANCH, queste influenzano significativamente le variabili dipendenti solo in alcuni casi. Le altre variabili di controllo non sono statisticamente significative.

La colonna 2 della Tabella 3 mostra i risultati ottenuti quando si considera l'IQI a livello regionale (IQI\_REG). I risultati sopra discussi sono sostanzialmente confermati.

Si conducono separatamente test diagnostici per i modelli Probit, Poisson e SYS-GMM. Per il modello Probit si testa la correttezza della forma funzionale. Nel complesso, la forma funzionale risulta appropriata<sup>9</sup>. Per verificare l'ipotesi di equidispersione del modello Poisson, si effettua un test di verosimiglianza. I risultati del test indicano che non c'è evidenza di *over-dispersion*. I risultati dei test diagnostici relativi alle stime GMM mostrano che le assun-

<sup>7</sup> Per confrontare gli stimatori *pooled* e *panel*, si svolge un test di verosimiglianza. I risultati del test evidenziano l'appropriatezza degli stimatori *panel* (Probit e Poisson).

<sup>8</sup> Per applicare lo stimatore SYS-GMM ed eseguire il test di autocorrelazione del primo ordine tra i residui differenziati, si considera un campione di quattro anni assumendo che tra il 2003 ed il 2004 il valore dell'IQI rimanga invariato. Questo potrebbe essere accettabile considerando che "*the process of institutional change occurs slowly, and appreciable changes in institutional quality occur only in the medium to long-term*" (Nifo e Vecchione, 2014, pp. 6). Per coerenza, si utilizza lo stesso campione per le regressioni Probit e Poisson.

<sup>9</sup> Si testa la correttezza della forma funzionale del modello Probit usando metodi parametrici e semiparametrici.

	COLONNA 1 - IQI provinciale						COLONNA 2 - IQI regionale					
	PROBIT <sup>a</sup>		POISSON <sup>b</sup>		SYS-GMM <sup>b</sup>		PROBIT <sup>a</sup>		POISSON <sup>b</sup>		SYS-GMM <sup>b</sup>	
	pooled	panel	pooled	panel	pooled	panel	pooled	panel	pooled	panel	pooled	panel
<b>Institutions</b>												
IQI / IQI_REG	-0.106**	-0.062***	-1.220*	-0.183	-1.710***		-0.231***	-0.022	-2.283***	-0.327		-3.612**
	0.045	0.006	0.061	0.429	0.001		0.000	0.268	0.001	0.281		0.022
<b>Firm's characteristics</b>												
EMP	0.048***	0.036***	0.704***	0.137***	0.243***		0.048***	0.022***	0.704***	0.137***		0.540***
	0.000	0.000	0.000	0.000	0.002		0.000	0.001	0.000	0.000		0.000
AGE	0.039***	0.028***	0.253***	0.051***	-0.036		0.039***	0.016	0.253***	0.051***		-0.004
	0.000	0.000	0.000	0.000	0.551		0.000	0.008***	0.000	0.000		0.960
LEVER	0.002***	0.001***	0.026***	0.005***	0.008***		0.002***	0.001***	0.026***	0.005***		0.015***
	0.000	0.000	0.000	0.000	0.009		0.000	0.000	0.000	0.000		0.000
INTAS	0.002	0.0004	0.018*	0.004*	-0.005		0.002	0.0002	0.018*	0.004*		-0.005
	0.168	0.419	0.062	0.058	0.634		0.169	0.646	0.062	0.058		0.741
QUICK	-0.024***	-0.008***	-0.235**	-0.039***	-0.0311		-0.024***	-0.007***	-0.235**	-0.039***		-0.039
	0.001	0.000	0.013	0.000	0.633		0.001	0.001	0.013	0.000		0.678
GROUP	-0.012	-0.009	-0.091	-0.021	0.310		-0.012	-0.008	-0.092	-0.021		-0.001
	0.499	0.267	0.536	0.364	0.881		0.498	0.143	0.534	0.360		0.995
CONS	-0.003	-0.009	0.146	0.019	0.0641		-0.003	0.0004	0.148	0.019		-0.031
	0.930	0.433	0.600	0.687	0.700		0.931	0.974	0.597	0.685		0.882
HT	0.102***	0.060***	1.048***	0.151**	0.382***		0.102***	0.055*	1.047***	0.151**		0.616***
	0.006	0.007	0.000	0.019	0.007		0.006	0.053	0.000	0.019		0.008

Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica

INORG	0.023	0.010	0.197**	0.026	0.101	0.023	0.004	0.197**	0.026	0.0738
	0.247	0.310	0.048	0.185	0.452	0.247	0.409	0.048	0.185	0.660
INPP	0.028**	0.017***	0.266***	0.045***	0.194	0.028**	0.011**	0.266***	0.045***	0.279*
	0.024	0.002	0.005	0.008	0.132	0.024	0.019	0.004	0.008	0.085
EXP	0.034**	0.030***	0.461***	0.087***	0.237	0.033**	0.017**	0.461***	0.087***	0.391
	0.049	0.000	0.000	0.000	0.272	0.049	0.008	0.000	0.000	0.163
COOP	-0.040	-0.031	0.051	-0.010	-0.078	-0.040	-0.026	0.053	-0.009	0.129
	0.584	0.142	0.929	0.910	0.701	0.586	0.197	0.926	0.913	0.649
NBANK_1					0.732***					0.488***
					0.000					0.000
<b>Bank-firm relationships characteristics</b>										
CRED	-0.001	-0.0003	0.644***	0.097***	0.162	-0.001	-0.002	0.644***	0.097***	0.016
	0.979	0.978	0.000	0.001	0.365	0.979	0.742	0.000	0.001	0.944
DURAT	0.000	0.000	0.002	0.0001	0.00273	0.000	0.0001	0.002	0.000	-0.002
	0.867	0.985	0.756	0.640	0.668	0.865	0.826	0.753	0.636	0.861
MAIN	-0.000	-0.00003	-0.006***	-0.001**	-0.0037	-0.000	-0.0001	-0.006***	-0.001**	-0.0001
	0.418	0.642	0.001	0.016	0.298	0.417	0.290	0.001	0.016	0.960
<b>Context characteristics</b>										
BRANCH	-0.034	0.012***	-0.426	-0.039	0.0546*	-0.039	0.005	-0.420	-0.038	0.180
	0.576	0.000	0.194	0.714	0.079	0.492	0.031	0.216	0.719	0.350
Constant					0.814**					0.339
					0.029					0.752
EFFETTI FISSI PROVINCIALI	SI	NO	SI	SI	SI	SI	NO	SI	SI	SI

N	5687	5687	5687	5687	5687	5687	5687	5687	5687	6,381	5687	6,381
Numero di id										2,812		2,812
Log pseudolikelihood	-2105,878	-1180,414	-12256,36	-11780,62		-2105,5556	-1191,9989	-12431,423			-11780,36	
Likelihood-ratio test of alpha=0			1.11 0.146					1.11 0.146				
AB test for AR(1)										-8,547 0,000		-7,937 0,000
AB test for AR(2)										-0,967 0,333		-1,097 0,273
Hansen test										295,2 0,160		261,9 0,627
Difference-in-Hansen tests										30,00 0,224		52,88 0,088

I simboli \*\*\*, \*\*, \* indicano livelli di significatività dell'1%, 5% e 10% rispettivamente. La Tabella 1 riporta la descrizione delle variabili. In corsivo sono riportati i *p-values* dei test. <sup>a</sup> La variabile dipendente è una *dummy* codificata 1 se l'impresa mantiene un numero di relazioni bancarie maggiore o uguale a due, o altrimenti. <sup>b</sup> La variabile dipendente è il numero di relazioni bancarie per impresa. Per le regressioni Probit and Poisson sono riportati gli effetti marginali. Gli *standard errors* (non riportati) sono clusterizzati a livello della singola provincia (NUTS3) e robusti rispetto alla potenziale eteroschedasticità degli errori. Per evitare che valori estremi possano distorcere i risultati dello studio, le variabili sono winsorizzate all'1%. Per le regressioni Probit e Poisson, si prendono i lags di tutte le variabili potenzialmente endogene o predeterminate. Le variabili EMP e AGE sono espresse in logaritmi. Tutte le regressioni sono condotte includendo gli effetti fissi temporali, settoriali e provinciali. Si riportano i test Arellano-Bond AR(1) e AR(2), Hansen test e difference in Hansen test.

zioni chiave su cui si basa lo stimatore adottato sono verificate. In particolare, i test di autocorrelazione sostengono l'assunzione di mancanza di autocorrelazione negli errori in livello, segnalando una forte correlazione del primo ordine tra i residui differenziati e l'assenza di autocorrelazione di ordine più elevato (rispettivamente, Arellano-Bond *tests for autocorrelation in first* (AB test AR1) *and second differences* (AB test AR2)). Inoltre, il test di Hansen non rigetta in nessun caso l'ipotesi nulla della validità delle *overidentifying restrictions* e la statistica conosciuta come *difference in Hansen test* supporta sempre la validità degli strumenti aggiuntivi impiegati dallo stimatore SYS-GMM.

È interessante verificare se la relazione negativa tra la qualità istituzionale ed il multiaffidamento possa essere specificamente attribuita ad una o più dimensioni incluse nell'indice sintetico. Per studiare i possibili diversi effetti di ogni dimensione che compone l'indice IQI, si stimano cinque set supplementari di regressioni, usando di volta in volta una delle dimensioni dell'IQI come regressori al posto dell'indice complessivo. La Tabella 4 riporta i risultati ottenuti stimando i modelli (1), (2) e (3) rispettivamente con gli stimatori Probit, Poisson e SYS-GMM<sup>10</sup>.

<sup>10</sup> Per esigenze di spazio, nella Tabella 4 si presentano gli effetti marginali (per i modelli Probit e Poisson) ed i coefficienti (SYS-GMM) stimati per le singole dimensioni dell'indice IQI a livello provinciale e regionale.

**Tab. 4. L'effetto delle singole dimensioni dell'IQI sul multi affidamento bancario.**

	COLONNA 1 - sotto indici a livello provinciale						COLONNA 2 - sotto indici a livello regionale					
	PROBIT <sup>a</sup>		POISSON <sup>b</sup>		SYS-GMM <sup>b</sup>		PROBIT <sup>a</sup>		POISSON <sup>b</sup>		SYS-GMM <sup>b</sup>	
	pooled	panel	pooled	panel	pooled	panel	pooled	panel	pooled	panel	pooled	panel
GOVERN	-0.031	-0.059**	-1.360**	-0.152*	-1.420***	-0.117*	-0.032	-2.573***	-0.375	-3.626*		
	0.576	0.025	0.012	0.054	0.0002	0.069	0.346	0.000	0.206	0.059		
RULAW	-0.115***	0.027	-0.358	-0.011	-0.114	-0.212***	-0.023	-0.765*	-0.082	-0.628		
	0.002	0.104	0.328	0.945	0.601	0.000	0.246	0.090	0.695	0.336		
VOICE	0.007	-0.080***	0.349	0.054	0.382	-0.194*	-0.045*	0.824	0.063	-1.112		
	0.912	0.000	0.356	0.821	0.322	0.08	0.058	0.285	0.873	0.368		
REGUL	0.049	-0.019	0.364	-0.047	-2.187**	0.300	-0.014	0.168	-0.106	-0.519		
	0.523	0.392	0.641	0.883	0.014	0.138	0.624	0.928	0.876	0.483		
CORR	-0.055	-0.003	0.477	-0.012	-0.667	-0.056	-0.002	0.533	0.04	-1.135		
	0.366	0.853	0.211	0.954	0.271	0.346	0.934	0.151	0.842	0.220		
EFFETTI FISSI PROVINCIALI	SI	NO	SI	SI	SI	SI	NO	SI	SI	SI	SI	SI

I simboli \*\*\*, \*\*, \* indicano livelli di significatività dell'1%, 5% e 10% rispettivamente. La Tabella 1 riporta la descrizione delle variabili. In corsivo sono riportati i *p-values* dei test. <sup>a</sup> La variabile dipendente è una *dummy* codificata 1 se l'impresa mantiene un numero di relazioni bancarie maggiore o uguale a due, 0 altrimenti. <sup>b</sup> La variabile dipendente è il numero di relazioni bancarie per impresa. Per le regressioni Probit and Poisson sono riportati gli effetti marginali. Gli *standard errors* (non riportati) sono clusterizzati a livello della singola provincia (NUTS3) e robusti rispetto alla potenziale eteroschedasticità degli errori. Per evitare che valori estremi possano distorcere i risultati dello studio, le variabili sono winsorizzate all'1%. Per le regressioni Probit e Poisson, si prendono i lags di tutte le variabili potenzialmente endogene o predeterminate. Tutte le regressioni sono condotte includendo gli effetti fissi temporali, settoriali e provinciali.



Secondo i risultati della Tabella 4, la dimensione GOVERN (*Government Effectiveness*) è negativa e statisticamente significativa nella maggior parte dei modelli, mostrandosi la dimensione più rilevante nello spiegare il multiaffidamento. Le dimensioni RULAW (*Rule of Law*), VOICE (*Voice and Accountability*) e REGUL (*Regulatory Quality*) sono negative e statisticamente significative solo in pochi casi, mentre la dimensione CORR (*Corruption*) non è mai statisticamente significativa<sup>11</sup>.

La colonna 2 della Tabella 4 mostra i risultati ottenuti quando si considerano le dimensioni a livello regionale. I risultati discussi sopra sono sostanzialmente confermati.

I risultati mostrano che la qualità istituzionale influenza negativamente il multiaffidamento bancario. Questo accade probabilmente perché una buona qualità istituzionale è associata ad un ambiente dove banche ed imprese interagiscono positivamente, scambiandosi informazioni e promuovendo strette relazioni bancarie. In altre parole, le istituzioni potrebbero creare buone condizioni per mitigare le asimmetrie informative, permettendo ad imprese e banche di cogliere tutti i benefici derivanti dalle *close lending relationships*. Più in dettaglio, le dimensioni della qualità istituzionale che sembrano più rilevanti sono GOVERN, RULAW e VOICE.

Per quanto riguarda le dimensioni GOVERN (*Government Effectiveness*) e REGUL (*Regulatory Quality*), i risultati mostrano che la capacità amministrativa degli enti locali in termini di qualità delle politiche e dei servizi pubblici, diminuisce il numero di rapporti bancari e la propensione delle imprese a multiaffidarsi. Questo evidenzia l'impatto che gli organi di governo intermedi (quali istituzioni politiche ed amministrative locali) giocano in modo attivo e positivo, influenzando così le decisioni finanziarie delle imprese. Quindi, le più efficaci politiche pubbliche nel campo della sanità, della gestione dei rifiuti e dell'ambiente, dei trasporti ed dell'istruzione che influenzano il *business environment*, riducendo i costi di transazione e le asimmetrie informative (Kneller e Misch, 2010; Datta, 2008; Shirley e Winston, 2004), sembrano favorire stretti rapporti bancari.

I risultati della dimensione RULAW (*Rule of Law*), in linea con la principale letteratura teorica ed empirica che ampiamente riconosce il ruolo di "*Rule of Law*" nel promuovere lo sviluppo economico ed influenzare le scelte delle

<sup>11</sup> In dettaglio, GOVERN non è statisticamente significativa solo per la regressione Probit (*pooled*); RULAW è significativa solo per la regressione Probit (*pooled*); VOICE è statisticamente significativa per la regressione Probit (*panel*), mentre REGUL è significativa solo per lo stimatore SYS-GMM.

imprese (Ayres, 1998; Buvinic e Morrison, 2000; Islam, 2003; Dam, 2006; World Bank, 2006; Lorentzen et. al, 2008; Nifo et al., 2016), indicano che contesti istituzionali caratterizzati da un'alta incidenza di criminalità, evasione fiscale, economia sommersa, scarsa applicazione della legge e maggiori costi giudiziari, influenzano positivamente la propensione delle imprese a multiaffidarsi.

I risultati delle regressioni ottenuti inserendo la dimensione VOICE (*Voice and Accountability*), confermano il ruolo cruciale della partecipazione sociale sul *business environment* e, più in dettaglio, sulle scelte delle imprese (Powell e Owen-Smith 2004; Sorenson 2003; Tallman et al. 2004). In particolare, quando il problema di asimmetria informativa è grave, interazioni sociali favorevoli potrebbero rappresentare una forma indiretta di controllo per evitare comportamenti opportunistici e anti sociali, inducendo banche ed imprese a stabilire stretti rapporti bancari, essendo anche più facile per le banche ottenere informazioni qualitative sull'impresa e beneficiare del loro utilizzo.

Infine, quando l'IQI è sostituito con il sotto indice CORR (*Corruption*), i risultati non mostrano alcun effetto sul multiaffidamento. In particolare, il sotto indice CORR presenta il segno atteso in tutti i modelli (negativo), ma i coefficienti non sono mai statisticamente significativi. Secondo alcuni studi, una possibile spiegazione è che quando il livello di corruzione è molto simile in tutte le regioni italiane (De Rosa et al, 2010; Lasagni et al, 2015; Nifo et al, 2016), è difficile che piccole differenze possano essere associate ad eterogeneità a livello di impresa.

L'evidenza che si presenta sul caso italiano sembra confermare la validità dell'ipotesi di ricerca. L'analisi econometrica svolta in questo lavoro, controllando per le caratteristiche individuali delle imprese (dimensione, età, indebitamento, internazionalizzazione, appartenenza ai settori Hi-Tech, etc), per le caratteristiche banca-impresa (razionamento del credito, durata del rapporto e percentuale di debito detenuto dalla banca principale) e per le condizioni macroeconomiche a livello locale (sviluppo del mercato bancario locale), riconosce alla qualità istituzionale un ruolo significativo nel determinare il fenomeno del multiaffidamento<sup>12</sup>.

<sup>12</sup> Per sottoporre a verifica la robustezza dei risultati fin qui commentati, innanzitutto si stimano le equazioni (1), (2) e (3): a) inserendo la variabile RGDPC (PIL a livello provinciale), b) inserendo la variabile RGDPC, ma considerando soltanto le imprese localizzate al Nord d'Italia, dove lo sviluppo economico è più omogeneo. I risultati rimangono sostanzialmente inalterati. I risultati di quest'ultime stime sono disponibili su richiesta.

## 4. Conclusioni

Il lavoro ha proposto una stima dell'impatto della qualità delle istituzioni a livello provinciale sul numero di relazioni bancarie attive per un campione di imprese manifatturiere nel periodo 2003-2006. La misurazione di tale impatto è basata sull'indicatore IQI, un indicatore sintetico della qualità delle istituzioni derivato da 24 indicatori elementari raggruppati in 5 dimensioni (*Corruption, Government Effectiveness, Regulatory Quality, Rule of Law, Voice and Accountability*). I risultati sono in linea con gran parte della letteratura esistente che assegna un ruolo chiave al *business environment* e alla qualità delle istituzioni nella definizione del comportamento delle imprese.

Le stime, sviluppate con controlli relativi alle variabili più microeconomiche di carattere industriale, indicano che nelle aree con un più basso livello della qualità delle istituzioni si trovano le imprese con maggiori relazioni bancarie suggerendo quindi che i problemi tipici derivanti dal multiaffidamento bancario sono mitigate lì dove migliore è la qualità istituzionale. Questo risultato porta a concludere che in un contesto con alto capitale sociale, un sistema legale e finanziario efficiente, delle buone regole e delle buone politiche pubbliche le asimmetrie informative sono meno stringenti e più ampia è la possibilità di scambiarsi informazioni riducendo quindi alcuni dei problemi tipici derivanti dal multiaffidamento bancario: *the hold up problem*, *the soft budget constraint problem* o *the liquidity problem*.

Più nello specifico, i risultati principali del lavoro possono essere così sintetizzati: 1) migliori istituzioni locali sono un *driver* importante in quanto aumentano la propensione delle imprese a mantenere stabili relazioni bancarie; 2) considerando i sottoindicatori dell'IQI, la dimensione GOVERN, che cattura la capacità dei governi locali nell'implementare efficaci politiche pubbliche diminuirebbe il numero di relazioni bancarie delle imprese; la dimensione RULAW, che cattura la certezza di diritto, esercita un impatto significativo sulla decisione dell'impresa di multiaffidarsi; 3) interessante e per certi versi sorprendente rispetto alle conclusioni di altri studi: la dimensione CORR, relative alla corruzione, non sembra esercitare particolare influenza nelle decisioni di impresa.

La principale conclusione del lavoro suggerisce che la ricerca futura sul *relationship banking* dovrà considerare le possibili conseguenze di alternativi setting istituzionali che vanno a condizionare le decisioni delle imprese e degli agenti economici più in generale.

La presenza di spillovers connessi ad una buona qualità delle istituzio-

ni e al meccanismo di incentivo da essi attivato è uno dei canali principali attraverso il quale fattori di contesto impattano positivamente sul *business environment* ed il clima di competitività di intere aree geografiche, indicando al policy maker uno strumento strategico (esempio: riforme istituzionali specialmente nell'ambito della *Government Effectiveness*, la *Rule of Law* e la *Voice and Accountability*) per cogliere le migliori opportunità e promuovere sviluppo economico.

## Riferimenti Bibliografici

- Acemoglu D., Dell M. (2010), "Productivity differences between and within countries", *American Economic Journal: Macroeconomics*, 2(1), 169-188.
- Agostino M., La Rocca M., La Rocca T., Trivieri F. (2012), "Do local financial and legal systems affect SMEs capital structure?", *Economics Bulletin*, 32 (1), 260-271.
- Agostino M.R., La Rocca M., La Rocca T., Trivieri F. (2010), Local Financial Development and SMEs Capital Structure, WP SSRN, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1343136](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1343136)
- Aiello F., Pupo V., Ricotta F. (2014), "Explaining TFP at firm level in Italy. Does location matter?", *Spatial Economic Analysis*, 9 (1), 51-70.
- Andrianova S., Demetriades P. (2004), *Finance and Growth: what we know and what we need to know* in C. Goodhart (eds), *Financial Development and Economic Growth: Explaining the Links*, 38-65, Palgrave Macmillan: Basingstoke and New York.
- Arellano M., Bover O. (1995), "Another look at the instrumental variable estimation of error components models", *Journal of Econometrics*, 68, 29-5.
- Ayres R. (1998), "Crime and Violence as Development Issues in Latin America and the Caribbean" *World Bank Latin America and Caribbean Studies Viewpoints*. Washington, DC: World Bank.
- Banca d'Italia (2007), *Relazione Annuale per l'anno 2006*, Roma.
- Benfratello L., Schiantarelli F., Sembenelli A. (2008), "Banks and innovation: Microeconomic evidence on Italian firms", *Journal of Financial Economics*, 90, 197-217.
- Bianco M., Japelli T., M. Pagano M. (2005), "Courts and banks: effect of judicial enforcement on credit markets!", *Journal of Money, Credit and Banking*, 37, 223-244.
- Blundell R., Bond, S. (1998), "Initial conditions and moment restrictions in dynamic panel data models", *Journal of Econometrics*, 87(1), 115-143.
- Buvinic and Morrison (2000), "Living in a more violent world". *Foreign Policy*, 118, 58-72.
- Carletti E. (2004), "The structure of bank relationships, endogenous monitoring, and loan rates", *Journal of Financial Intermediation*, 13, 1, 58-86.
- Cheng S., Shiu C. (2007), "Investor protection and capital structure: international evidence", *Journal of Multinational Financial Management*, 17, 30-44.
- Chinn D.M., Ito H. (2005), "What matters for financial development? Capital controls, institutions and interactions", *NBER Working Paper Series*, 11370.
- Claessens S., Leaven L. (2003), "Financial development, property rights and growth", *Journal of Finance*, 58(6), 2401-2436.

## Qualità delle istituzioni e multiaffidamento bancario. Un'indagine empirica

- Cosci S., Meliciani V. (2002), *Multiple banking relationships: Evidence from the Italian experience*. Manchester School Supplement, 37–54.
- Cosci S., Meliciani V. (2005), *Multiple banking relationships and over-leverage in Italian manufacturing firms*, Money Macro and Finance Research Group Conference 2005, (87).
- Dam K. (2006), “The Judiciary and Economic Development”, *John M. Olin Law & Economics Working Paper*, 287(2), The Law School, University of Chicago.
- Datta S. (2008), *The Impact of Improved Highways on Indian Firms*. Unpublished manuscript. Available at <http://www.enterprisesurveys.org/ResearchPapers/>.
- De Rosa D., Gooroochurn N., Gorg H. (2010), “Corruption and Productivity: Firm-Level Evidence from the Beeps Survey”, *Kiel Working Paper*, 1632.
- Del Monte A. Giannola A. (1997). *Istituzioni economiche e Mezzogiorno*, La Nuova Italia Scientifica, Roma.
- Demirgüç-Kunt A., Maksimovic V. (1998), “Law, Finance, and Firm Growth”, *Journal of Finance*, 53, 2107-2137.
- Demirgüç-Kunt A., Maksimovic V. (1999), “Institutions, financial markets, and firms debt maturity”, *Journal of Financial Economics*, 54, 295–336.
- Detragiache E., Garella E., Guiso L. (2000), “Multiple versus single banking relationships: theory and evidence”, *Journal of Finance*, 55, 1133-1161.
- Di Liberto A., Sideri M. (2015), “Past dominations, current institutions and the Italian regional economic performance”, *European Journal of Political Economy*, 38, 12-41.
- Diamond D. (1991), “Monitoring and reputation: the choice between bank loans and privately placed debt”, *Journal of Political Economy*, Vol. 99, 689–721.
- Elsas R., Heinemann, Tyrell M. (2004). “Multiple but asymmetric bank financing: the case of relationship lending”, *CESifo Working Paper Series* 1251.
- Erbetta F., Petraglia C. (2011), “Drivers of regional efficiency differentials in Italy: technical inefficiency or allocative distortions?” *Working Paper No. 2*, CERIS-CNR.
- Fazio G., Piacentino D. (2010), “A Spatial Multilevel Analysis of Italian SMEs' Productivity”, *Spatial Economic Analysis*, 5(3), 299-316.
- Ferri G., Messori M. (2000), “Bank-firm relationships and allocative efficiency in northeastern and central Italy and in the South”, *Journal of Banking and Finance*, 24, 1067-1095.
- Foglia A., Laviola S., Marullo-Reedtz P. (1998), “Multiple banking relationship and the fragility of corporate borrowers”, *Journal of Banking and Finance*, 22, 1441-1456.
- Ganau R., Rodríguez-Pose A. (2016), *Industrial Clusters, Organised Crime and Productivity Growth in Italian SMEs*, Università degli Studi di Padova, manuscript.
- Garrettsena H., Lensink R., Sterkenb E. (2004), “Growth, financial development, societal norms and legal institutions”, *Journal of International Financial Markets, Institutions and Money*, 14, 165-183.
- Giannola A., Petraglia, C., Scalera D. (2016), “Net fiscal flows and interregional redistribution in Italy: A long-run perspective (1951-2010)”, *Structural Change and Economic Dynamics*, 39, 1–16.
- Guiso L., Minetti R. (2004). “Multiple creditors and information rights: theory and evidence from Us firms”, *CEPR Discussion Paper* 4278.
- Haroff D., T. Korting T. (1998a), “Lending Relationships in Germany: Empirical Results from Survey Data”, *Journal of banking and finance*, 22, 1317-1353.
- Islam R. (2003), “Institutional Reform and the Judiciary: Which Way Forward”, *World Bank Policy Research Working Paper*, 3134, 7-8.

- Islam R. (2003), "Institutional Reform and the Judiciary: Which Way Forward", *World Bank Policy Research Working Paper*, 3134, 7-8.
- Kaufmann D., Kraay A., Mastruzzi M. (2010), "The worldwide governance indicators: methodology and analytical issues", *Policy Research Working Paper Series 5430*, The World Bank.
- Kneller R., Florian M. (2011), "What Does Ex-post Evidence Tell us About the Output Effects of Future Tax Reforms", *ZEW Discussion Paper 11-029*, <http://ftp.zew.de/pub/zew-docs/dp/dp11029.pdf>. Last accessed June 12, 2015.
- La Rocca M., La Rocca T., Cairola A. (2010), "The influence of local institutional differences on the capital structure of SMEs: Evidence from Italy", *International small business journal*, 28(3), pp. 234-257.
- Lasagni A., Nifo A., Vecchione G. (2015), "Firm productivity and institutional quality: Evidence from Italian industry", *Journal of Regional Science*, 55(5), 774-800.
- Lorentzen P., MacMillan, J., Wacziarg, R. (2008), "Death and Development", *Journal of Economic Growth*, 13, 81-124.
- Malanima P., Zamagni V. (2010), "150 years of the Italian economy", 1861-2010, *Journal of Modern Italian Studies*, 15(1), 1-20.
- Mannarino L., Pupo V., Ricotta F. (2016), "Family firms and productivity: the role of institutional quality", *WP DESF*, No. 5
- Nerozzi S., Pipitone V., Ricchiuti G. (2015), *Social capital and firm's productivity in Italy: a multilevel approach*. AISRE Annual Conference, University of Calabria.
- Nifo A. (2011), "L'industria del Mezzogiorno: un'analisi empirica sui divari di produttività", *Economia e politica industriale*, 38(3), 105-126.
- Nifo A., Vecchione G. (2014), "Do institutions play a role in skilled migration? The case of Italy", *Regional Studies*, vol. 48(10), pp. 1628-1649.
- Nifo A., Vecchione, G. (2015), "Measuring institutional quality in Italy", *Rivista economica del Mezzogiorno*, N. 1-2, pp. 157-181, Il Mulino, Bologna.
- Nifo A., Scalera D., Vecchione G. (2016). "The rule of law and educational choices. Evidence from Italian regions", *Regional Studies*, DOI 10.1080/00343404.2016.1262945.
- Ongena S., Smith D.C. (2000), "What determines the number of number of bank relationships? Cross-country evidence", *Journal of Financial Intermediation*, 9, 26-56.
- Pelliccioni G., Torluccio G. (2007), *Il rapporto banca-impresa: le determinanti del multiaffidamento in Italia*, pp 43-77 chapter of the vol. 74 "Il rapporto banca-impresa in Italia", by Monferrà Stefano, Bancaria Editrice.
- Pollard J. (2003), "Small firm finance and economic geography", *Journal of Economic Geography*, 3, 429-452.
- Powell W.W., Owen-Smith J. (2004), "Knowledge networks in the Boston biotechnology community", *Organization Science*, 15, 5-21.
- Rajan R.G. (1992). "Insiders and outsiders: The choice between informed and arm's-length debt", *Journal of Finance*, 47(4), 1367-1399.
- Raspe O., van Oort F. (2011), "Firm Heterogeneity, Productivity, and Spatially Bounded Knowledge Externalities", *Socio-Economic Planning Sciences*, 45, 38-47
- Sarno D. (2009), *Sviluppo finanziario e crescita economica nel Mezzogiorno*, Franco Angeli.
- Scalera D., Zazzaro A. (2010), "L'economia del Mezzogiorno. Nuova politica regionale, crisi globale e federalismo fiscale", in M.C. Guerra and A. Zanardi (eds.), *La finanza pubblica italiana*, Rapporto 2010, 375-410.

- Sharpe S.A. (1990), "Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships", *Journal of Finance* 45(4), 1069-1087.
- Shirley C., Winston C. (2004), "Firm inventory behavior and the returns from highway infrastructure investments", *Journal of Urban Economics*, 55, 2, 398-415.
- Sorenson O. (2003), "Social networks and industrial geography", *Journal of Evolutionary Economics*, 13(5), 513-527.
- Tallman S., Jenkins M., N. Henry and S. Pinch. (2004), "Knowledge, clusters and competitive advantage", *Academy of Management Review*. 29(2), 258-271.
- World Bank Enterprise Survey. *Policy Research Working Paper Series* 6918, The World Bank.
- Yosha O. (1995), "Disclosure costs and the choice of financing source", *Journal of Financial Intermediation*, 4, 3-20.





## Profili degli autori

**Alberto Bianchi**, Dottore di Ricerca in Storia Economica e attualmente docente esterno di “Storia economica e statistica del turismo” presso il Campus Lucca e di “Storia del commercio e della navigazione” presso il Polo Universitario dei Sistemi Logistici di Livorno. Si è occupato nei propri studi di regolazione pubblica delle public utilities e di storia d’impresa nel settore chimico e farmaceutico nel corso del Novecento.

**Carlo Brambilla**, Professore Associato di Storia Economica presso il Dipartimento di Economia dell’Università dell’Insubria, dove insegna storia economica, dell’innovazione e del management. Si occupa principalmente di storia bancaria, di rapporti banca/industria e finanziamento industriale, di storia d’impresa e corporate governance.

**Domenico Buccella**, Ph.D., Assistant Professor presso la Leon Kozminski University, Warsaw, Poland, e External Lecturer presso la Bradford University School of Management, Bradford, UK. Interessi di ricerca: Industrial Organization; Contrattazione Collettiva; Economia dei sindacati; Commercio internazionale e Investimenti diretti all’estero; International business.

**Alessandra Chirco**, Professore Ordinario di Economia Politica presso l’Università del Salento, dove insegna Microeco-

nomia a livello di base e avanzato. I suoi interessi di ricerca sono nel campo delle microfondazioni della macroeconomia, della teoria dell'Organizzazione Industriale e delle relazioni tra eterogeneità degli agenti e struttura di mercato.

**Giuseppe Conti**, Professore Ordinario di Storia Economica presso il Dipartimento di Economia e Management dell'Università di Pisa. Insegna storia e politica monetaria, economia e storia delle crisi finanziarie e storia del management. I principali interessi di studio sono: storia del credito e della banca in età contemporanea, evoluzione delle strutture finanziarie e di governance di impresa e dell'economia.

Lorenzo Corsini è Ricercatore Senior presso il dipartimento di Economia e Management dell'Università di Pisa, dove è titolare del corso di Politica Economica. I suoi interessi di ricerca riguardano gli aspetti empirici e teorici del mercato del lavoro e dei sistemi di welfare.

**Simone D'Alessandro**, Professore Associato di Economia Politica presso il Dipartimento di Economia e Management dell'Università di Pisa dove insegna Macroeconomia ed Economia dello Sviluppo, Globalization and Economic Development e Happiness, Sustainability and Wellbeing (al dottorato regionale in Economics). I suoi interessi di ricerca riguardano la cooperazione, la distribuzione del reddito e della ricchezza, l'economia industriale, l'economia ecologica, la sostenibilità ecologica e la decrescita.

**Luciano Fanti**, Professore Ordinario di Economia Politica presso il Dipartimento di Economia e Management dell'Università di Pisa, dove insegna Economia Politica e Industrial Organization e Teoria dei Giochi. Tra i suoi principali interessi di ricerca la Dinamica Economica, l'Economia della Popolazione, la Teoria del Ciclo e della Crescita e l'Industrial Organization.

**Nicola Giocoli**, Professore ordinario di Economia Politica, Dipartimento di Giurisprudenza, Università di Pisa, insegna Economia Politica ed Analisi Economica del Diritto. Tra i suoi interessi di ricerca, la storia della teoria dei giochi e delle decisioni e l'economia e diritto dell'antitrust.

**Marco Grazzi**, Professore Associato di Politica economica presso l'Università di Bologna, ove tiene i corsi di Economia Industriale Avanzata, Economia Internazionale e Green Technology and Innovation. Interessi di ricerca: dinamica industriale, commercio internazionale, teoria dell'impresa

**Fabio Lavista**, Assegnista di ricerca presso il Dipartimento di Economia dell'Università degli studi dell'Insubria, insegna Storia Economica presso questa università, l'Università Bocconi e l'Università Cattolica del Sacro Cuore. Le sue principali aree di ricerca sono quelle della storia economica e d'impresa, in particolare: le culture manageriali, l'organizzazione d'impresa, le imprese a partecipazione statale e le politiche pubbliche.

**Andrea Mangani**, Professore Associato di Economia Applicata presso il Dipartimento di Scienze Politiche dell'Università di Pisa. Insegnamenti: Economia industriale, Economia internazionale. Interessi di ricerca: economia dell'informazione, proprietà intellettuale, mass media.

**Nicola Meccheri**, Professore Associato di Economia Applicata presso il Dipartimento di Economia e Management dell'Università di Pisa dove svolge i corsi di Macroeconomia, Economia dei mercati finanziari e Industrial Organization e teoria dei giochi. Le sue tematiche di ricerca riguardano la teoria dei contratti, Industrial Organization ed economia del lavoro.

**Mario Morroni**, Professore Ordinario di Economia Politica presso il Dipartimento di Scienze Politiche dell'Università di Pisa, insegna Economia politica e Economia dell'impresa. I suoi attuali interessi di ricerca si concentrano sull'economia dell'innovazione, sulle teorie dell'impresa, sull'economia dell'ambiente e sulle politiche monetarie e fiscali anticicliche.

**Daniele Moschella**, Ricercatore di Politica Economica presso la Scuola Superiore Sant'Anna, Pisa, ove insegna Organizzazione Industriale presso il corso di dottorato in Economia. Interessi di ricerca: dinamica industriale, economia dell'innovazione, commercio internazionale.

**Annamaria Nifo**, Ricercatrice di Economia Applicata presso l'Università degli Studi del Sannio. Insegna Economia Politica presso il Corso di Laurea magistrale in Giurisprudenza ed Economia Industriale presso il Corso di Laurea in Scienze Statistiche e Attuariali. I principali interessi di ricerca sono: qualità delle istituzioni e performance economiche; flussi migratori e capitale umano; trasformazioni strutturali e dualismi territoriali; organizzazione della produzione e crescita delle imprese industriali.

**Valeria Pinchera**, Professore Associato di Storia Economica presso il Dipartimento di Economia e Management dell'Università di Pisa, insegna storia economica e storia dei consumi. I suoi interessi di ricerca riguardano l'evoluzione economica della moda dal Novecento ad oggi, la storia dei consumi tra età moderna e contemporanea e l'economia dell'arte nei secoli dell'età moderna.

**Sabrina Ruberto**, dottoranda in Scienze Economiche ed Aziendali presso il Dipartimento di Economia, Statistica e Finanza "Giovanni Anania" dell'Università della Calabria. Si

interessa prevalentemente di multiple banking relationships, qualità istituzionale ed efficienza.

**Neri Salvadori**, Professore Ordinario di Economia Politica presso il Dipartimento di Economia e Management dell'Università di Pisa, dove insegna Economia Industriale e Advanced Microeconomics. Tra i suoi principali interessi di ricerca la Teoria della Produzione, la Teoria della Crescita, l'Industrial Organization, e la Storia dell'Analisi Economica.

**Marcella Scrimitore**, Professore Associato di Economia Politica presso l'Università del Salento, dove insegna Teoria dei Giochi, Microeconomia ed Economia Industriale. I suoi interessi di ricerca si collocano nell'ambito dell'Organizzazione Industriale, con particolare riferimento alla teoria della delega manageriale e alla teoria degli oligopoli misti.

**Raffaello Seri**, Professore Ordinario di Econometria presso il Dipartimento di Economia dell'Università dell'Insubria. Si occupa principalmente di econometria, statistica, matematica applicata e teoria delle decisioni.

**Rodolfo Signorino**, Professore Associato di Economia Politica alla Facoltà di Giurisprudenza dell'Università di Palermo, dove insegna Economia Politica. I suoi principali interessi di ricerca riguardano la storia e la metodologia dell'analisi economica.

**Massimiliano Vatiere**, Dottore di Ricerca in Law and Economics all'Università di Siena, è titolare della Cattedra "Brenno Galli" in Analisi Economica del Diritto all'Università della Svizzera italiana. I suoi temi di ricerca e insegnamento comprendono Analisi Economica del Diritto, Regulation e Concorrenza dei Mercati, Corporate Governance, ed Economia delle Istituzioni.

**Gaetano Vecchione**, Ricercatore di Economia Politica presso il Dipartimento di Scienze Politiche dell'Università degli Studi di Napoli Federico II dove insegna Integrazione Economica Internazionale. I suoi campi di ricerca sono: Institutional quality, skilled migration ed economia regionale.

**Giuseppe Vittucci Marzetti**, Professore Associato di Politica Economica, Dipartimento di Sociologia e Ricerca Sociale, Università di Milano-Bicocca. Discipline di insegnamento: Microeconomia, Macroeconomia, Econometria, Valutazione delle politiche pubbliche. Interessi scientifici: Analisi della produttività; Analisi delle reti; Economie di agglomerazione; Spillover tecnologici; Teoria della produzione.



Finito di stampare nel mese di dicembre 2017  
da Impressum - Marina di Carrara  
per conto di Pisa University Press



# EX LIBRIS



Αριστοτέλης

Questo e-book appartiene a  
rodolfo.signorino unipa.it 18021218-  
0258-0855-2847-16xcy42i3934