TESI DI DOTTORATO

Dipartimento Di Scienze Statistiche e Matematiche "Silvio Vianelli"

CAUSAL MODELS FOR MONITORING UNIVERSITY OF PALERMO ORDINARY FINANCING FUND

Modelli causali per il monitoraggio del Fondo di
Finanziamento Ordinario dell'Università di Palermo

Salvatore Marcantonio

Tutor: *Prof.ssa Antonella Plaia*

Coordinatore Dottorato: *Prof. Marcello Chiodi*

**Dottorato di Ricerca in "Statistica, Statistica Applicata e Finanza Quantitativa", XXIII ciclo - 2011
Settore Scientifico Disciplinare: SECS/S01 - Statistica, SECS/05 -Statistica Sociale**

Università degli studi di Palermo

UNIVERSITÀ
DEGLI STUDI
DI PALERMO

DSSM
DIPARTIMENTO DI
SCIENZE STATISTICHE
E MATEMATICHE
PALERMO

John 8,32

Then you will know the truth, and the truth will set you free.

To Judea Pearl for his Science and his Mercy

Acknowledgements

Contents

11

List of Figures

List of Tables

17

Chapter 1

Introduction

Recently iterated decreasing government transfers and an even more increasing proportion of budget allotted based on competitive performances took Italian Universities started struggling with competition for funds, in particular for the University Ordinary Financing Fund (FFO).

Three years ago, the Ministry of University and of Scientific and Technological Research outfitted a set of indicators to assess the quality of the educational offer and outcomes of formative processes and the quality of scientific research (Article 2, Law 9, January 2009, No 1). Although this system has been criticized by an authoritative and independent assessment organization [1], about 50 universities participate to the annual allocation funds under this scheme. Moreover, the share of FFO allocated in this way is constantly increasing.

The University of Palermo has decided to initiate statistical studies to monitor the FFO indicators in order to increase the budget. These statistical models are aimed at informing several stakeholders on:

- which are the variables responsible for the indicators, what are their present and past values and how they relate with national references.

---

[1]http://cronaca.anvur.it/

This is the first step by which subjects having the institutional duty to affect indicator values become aware of what is the state of the art;

- what could be the future values of these variables, look what are the components of strength and weakness in order to looking at measures for the correction of weaker performances;

- one or more strategies to be taken to increase the value of the indicators, also comparing different time and realization costs.

All this aims can be achieved by both elementary statistical techniques such as tables and graphics useful to show crude quantitative results and, intuitive trends, and with more complex models able to dealing with time, in particular with short term forecast, and to work with a small amount of empirical evidence (because this is only the second year allocation so there are few past observations available). Models should be theoretically equipped with the distinction between predicting under observation and predicting under intervention, in order to provide correct answers to the distinct tasks of pure out of sample extrapolation and policy making, e.g acting on the system of rules governing relationships among variables and changing it. They should be also capable of encoding not only information arising from empirical data, but also from extra knowledge, such as expert opinions, in order to quickly adapt to new possible scenarios and keep a genuine uncertainty about a priori information.

Chapter 2

The Italian University Funding System

The University Ordinary Financing Fund (FFO) is the primary entry of the universities budget. It is governed by several laws:

- Ministerial Decree No.655 December, 21 2010: Decree criteria for allocation of Ordinary Financing Fund (FFO) to Universities for the year 2010;

- Article 2, Law 9, January 2009, No 1: from 2009 onwards, in order to promote and support quality improvement activities of the public universities and improve the effectiveness and efficient use of resources, a portion of at least 7% of FFO (...), with progressive increases in following years, will be allocated by taking into account:

  a Quality of the educational offer and educational process results;

  b Quality of scientific research;

  c Quality, effectiveness and efficiency of campuses.

- Article 2, Law of 24 December 2007, No 244: (...) the Ministry of University and Research has established a fund with a budget of Euro 550 million for the year 2008, Euro 550 million for the year 2009 and Euro 550 million for the year 2010 (...). The allocation of resources is subject to the adoption of a program plan taking place by January 2008. This will be approved by Decree of the Minister of Universities and Research, in consultation with the Minister of Economy and Finance, in consultation with the Conference of Italian University Rectors (CRUI). This plan aims to:

  a raise the overall quality of the university system and the efficiency of the universities;

  b strengthen the incentive mechanisms for efficient and appropriate use of resources, with limitation of employer costs for the benefit of research and teaching;

  c accelerate the financial balance between the universities on the basis of binding parameters, uniform and realistic assessments of future costs and, in order to make the parameters effective, the ratio between staff costs and FFO should not exceed 90%;

  d redefine the university debt;

  e allow the rapid adoption of a system plan, with interventions including: appropriate monitoring and verification tools which are activated by the Ministry of University and Research, combiner with the Ministry of Economy and Finance, in consultation with the CRUI. This is aimed to influence the actual payment of resources to a more formal agreement with each university in objective with the plan.

- Article 5, Law 24, December 1993, No 537: The fund for the financing of the universities includes a basic fee, to be divided between the universities in proportion to the amount of state transfers and

costs incurred directly by the public universities 1993 budget, and a additional fee, to be divided on the basis of criteria determined by the Minister of University and Scientific and Technological Research, in consultation with the National University Counsel and the Conference of Rectors, relating to standard costs per student and the objective of research, taking into consideration the size, structural and environmental conditions. Since 1995, the basic fee for the financing fund of the universities, will be progressively reduced. The additional fee will be increased, by at least, an equal amount. The fee helps to balance the system for funding the initiatives carried out in accordance with development plans (...). The basic fee was calculated taking into account the 2009 allocation. This fee is reduced by about 80% for each university, taking into account the total budget(...). The criteria and indicators for the allocation of the quota referred to in Law 9, January 2009, No.1 are: The amount of 720 Ml€, is allocated by 34% (244.80 Ml€) on the basis of indicators A1-A4 and the remaining 66% (475.20 Ml€) based on indicators of B1-B4.

For our purposes it is important to know that although the proportion of FFO allocated according to a set of indicators changes year by year (7% in 2010, 10% in 2011, 13% in 2012), throughout this thesis it will be fixed at the 2010 level, 720Ml€, because it permit direct economic comparisons, in terms of gain or lost, of indicator values across years.

Funds are allocated according to the following table:

Quality of the educational offer and outcomes of formative processes				
Dimention	Name	Description	Weight	Amount to be allocated
Demand	A1	Students regularly anus. a. 2008/09 that have achieved at least 5 credits in 2009, broken down by group A, B, C, D and weighted with a specific weight (4 for group A, 3 for group B, group C for 2 and 1 for the Group D). For this indicator are applied three corrections: a) sustainability of the training (KA); b) the local context (KT); c) the strategic importance of the course of study (KR) (application pending for 2010).	0.5	122,4Ml€
Demand	A2	CFU ratio for acquired in 2009 and credits required for students enrolled in the anus. A. 2008/09, broken down by groups of course. For the calculation of the specific value is related with the median of the reference group.	0.50	122,4Ml€
Demand	A3	percentage of graduates employed three years after graduation. The indicator for the application is suspended pending the completion of the Registry 2010 National Graduates	0	0
Demand	A4	indicator of quality of teaching evaluated by students. The indicator was suspended in 2010 pending review of surveys currently in use.	0	0
Quality of the scientific research				
Dimention	Name	Description	Weight	Amount to be allocated
Research	B1	Percentage of positively evaluated researchers in projects PRIN 2005-2008, "weighted" for the success rate of the scientific area	0.35	166.32Ml€
Research	B2	Weighted average rates of participation in projects FIRB program "Futuro e Ricerca" weighted with their success rates, computed separately for two lines funding and normalized to national values.	0.15	71.28Ml€
Research	B3	Allocation coefficient of resources assigned to VTR 2001-03 (CIVR) Areas	0.30	142.56Ml€
Research	B4	The average percentage of: funding and successfully acquired by the universities projects in the Seventh Framework Programme - European Union - CORDIS; financing by the European Union; funding from other abroad public institutions.	0.20	95.04Ml€

Table 2.1: Indicators definition, weights and amount to be allocated on 2010

In this thesis only indicators $A1$, $A2$ and $B1$ are going to be monitored. Indicator $B3$ is a quantity computed years ago so it doesn't need to be forecast. Due to lack of local data, indicators $B2$ and $B4$ are not going to be considered.

Notice that all indicators are expressed as a percentage, so their values are directly comparable each other, e.g. $A2 = 2.8$ vs $B1 = 1.4$ means the quality of the outcomes of formative processes is twice the quality of the scientific research.

2.1 Methodology

Since the aim of this Thesis is predicting future quantities, the most suitable statistical technique for this task is time series analysis. It has been embedded into the Bayesian framework for two main reasons:

- for each quantity to be forecast, there are only a few observations available, minimum one maximum five. This means that usual asymptotic frequentist theorems cannot be applied, and information from empirical data, the likelihood, have to be enforced with extra knowledge, the prior distribution;

- time dependent observations call for flexible models able to adapt rapidly to system shocks or to external interventions. This is usually obtained through models with (time) varying parameters, a basic feature included in the Bayesian modelling.

The next section will provide a brief introduction to Bayesian framework and in particular, to its application to time series.

2.2 Introduction to Bayesian framework

2.2.1 Fundamentals

The well known Bayes theorem, (Bayes, 1763), (de Laplace, 1829), has been interpreted in two distinct ways: as a crude computation technique (Congdon, 2006), (Box and Tiao, 1992) and as a statistical philosophy (Roberts, 1994), (Raiffa and Schlaifer, 1968). Bernardo and Smith (2000) call this distinction the "Bayesian formalism" and "Bayesian thinking" dichotomy. The former being, the way probabilities are transformed, predictions are made and parameters are estimated. The latter being the ways people informally revisit, change and doubt opinions, on the basis of new evidence (Lindley, 2010). Theorem states that a probability of an event E to take a

generic value e, given that another event H takes a generic value h, depends on the probability of $E = e$ before any knowledge about $H = h$ is acquired, and on the probability of $H = h$ given by the evidence that $E = e$:

$$P(E = e|H = h) = \frac{P(E = e)P(H = h|E = e)}{P(H = h)} \quad \forall h \in H,\ e \in E$$

The probability on the left side of the equal sign is called posterior probability. The two probabilities in the numerator are called, respectively, the prior and the likelihood probabilities. $P(H = h)$ is a normalization constant (it does not depend on $E = e$) computable from the numerator. It is useful to notice that neither $E = e$ or $H = h$ have actually happened. If $E = e$ had happened, then $P(E = e|H = h)$ would be 1. If $H = h$ had happened, then it would be part of the sample space Ω[1].

Usually, the prior probability represents the available information about $E = e$ before $H = h$ happens and is sometimes called base knowledge. The likelihood represents empirical information arising from observational and/or experimental data[2]. The posterior probability contains up to date knowledge.

The philosophical aspect of the Bayesian paradigm involves prior probability interpretation: $P(E = e)$ is the probability before $H = h$ occurs, in empirical sciences, this means prior to conducting the experiment. As a consequence $P(E = e)$ is not estimable from data[3]. The two main sources of

---

[1]A probability space consists of three parts: A sample space, Ω, which is a non-empty set, sometimes called the sample space, whose members can be thought as a potential result of a random experiment. F is a sigma-algebra containing zero or more outcomes of Ω whose elements are called events. P is a measure of the probability in F, that is an assignment of a probability to events,(Kolmogorov, 1992).

[2] In some sense H restricts the set of possible values for E between the two boundaries, $P(E = e|H = h) = P(E = e)$, that is in the case of independence, and $P(E = e|H = h) = 1$ (or 0) whenever $E = e$ is a logical consequence of $H = h$ (or its negation $\bar{h}$). For all other intermediate cases, $H = h$ makes some e values of E inconsistent and it assigns the residual probability mass to the remaining values proportional to their prior probability.

[3]We are dealing with something different from the law of total probability $P(E = e) = \sum_h P(E = e|H = h)P(H = h)$, which is simply a constraint among probabilities. If we take into

this knowledge are expert judgements or previous similar analysis results. The first source embraces a probability calculus as a branch of philosophical thinking, known as subjectivism (de Finetti, 1989), (de Finetti, 1990), but it leaves the unsolved problem of transforming opinions into numbers, namely the elicitation task (O'Hagan, 2006). The second source is more pragmatic, for example an extensive bibliographical search could help in locating previous results as a good starting point for the prior probability. But, from a theoretical point of view, it poses again the same problem in that those results need their own prior probability: it is a sequence ad infinitum. Influence of the prior on the posterior is called informativeness. What prior to choose? There are two distinct families:

- non informative priors codify the notion of "no prior information available", e.g. uniform distribution means every hypothesis has equal probability. Sometimes, non informative priors are used as a reference level in a sensitivity analysis;

- informative priors codify some kind of knowledge, or assumption, about the quantity of interest. The degree of uncertainty on a parameter is formalized by a prior probability distribution on that parameter. sometimes for mathematical convenience (Good, 1965), it is useful to adopt a particular prior distribution on the basis of the likelihood in a way that prior and posterior have the same distribution, and the updating applies only to its parameters with considerable computational efficiency. This is called conjugacy analysis.

Moreover, applying bayes algorithm backwards, the prior can be inferred by the posterior and likelihood, e.g $P(E) = \frac{P(E|H)P(H)}{P(H|E)}$.

The prior distribution has no meaning in the frequentist approach and it is not used.

The likelihood $P(H|E)$ gives diagnostic information about how likely $H = h$, is in the light of $E = e$. The two approaches adopt it differently: Frequentists

account time $P(E = e)$ has to be evaluated before $H = h$ happens.

maximize it, whereas, Bayesians average it out using the prior values as weights.

For those who link probability and causation,(Suppes, 1970),(Galavotti, 2005),(Pearl, 1996), (Salmon and Wesley, 1998), *H* stands for the event cause, and *E* stands for the event effect.

Before closing I would like to digress briefly. I am sure, at least once in our lifetime, we were all Bayesians. I am referring to the case of the octopus Paul, who during the 2010 world champion soccer race, correctly guessed all seven of the Germany scores. Applying frequentist approach to interpret this phenomenon, gives two kinds of problems: for interpreting "Germany wins" as an event, Paul is supposed to continue guessing results until the success rate converge to a constant. Frequestists, I guess, would sustain non-scientific nature of the phenomenon, and consequently, it does not make sense to compute the probability. But, this argument does not concern frequencies, as it does Paul. Frequencies do not have anything in common with the nature of events generated them. The second problem regards the application of the binomial distribution within the frequentist framework, which leads to an estimation of the probability of success equal to one. Consequently, given that evidence and assuming temporal persistence, one will have sacred faith on Paul's art of divination in guessing further world champion races. By the way, Paul is now dead.

Obviously, none of us will have such faith in a fish. In fact Bayesian thinking, accounts for this common sense through the prior distribution on the octopus guessing capacity. If one thinks, whatever prevision Paul would ever make, and whatever the score will actually be, his opinion will remain the same, evidence (likelihood) won't have an impact on the posterior. But even though, for some reason, one agrees on a minimal divination capacity, the posterior will never have a value of one. That means, contrary to frequentist, bayesian approach starts from uncertainty and rests into uncertainty, also in the extreme case of full favourable certainty evidence.

2.2.2 Setting up the analysis: inference and prediction

Usually in statistics phenomena are represented by random variables, that is functions directly interpretable as probability of events. Parametric inference is a branch of statistics that assumes the functional form to be known up to some constants called parameters. The inductive task substantiates in inferring parameters values from (a sample of) observations of the random variables. In the Bayesian framework parameters are treated as random variables with their own distributions. In the most simple technique, called conjugate analysis, parameter posterior distributions take the same form of the prior although with updated parameters. It follows a brief description of the distributions used in the next models.

2.2.3 The normal distribution

One of the most exploited distribution for its good statistical properties is the normal distribution (Johnson *et al.*, 1994). It takes the form:

$$X \sim N(\mu, \tau) \triangleq P(X = x; \mu, \tau) = \frac{1}{\sqrt{2\pi\tau^{-1}}} exp[-\frac{\tau(x - \mu)^2}{2}]$$

The teoric moments are: $E[X]=\mu$, $V[X]=\sigma^2$. Sometimes the variance is substituted with the precision $\tau = \frac{1}{\sigma^2}$.
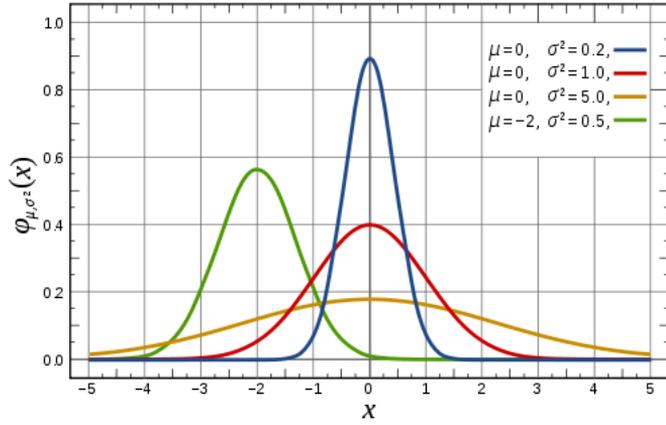
Figure 2.1: Normal probability mass function

Conjugate analysis prescribes the following parameter distributions:

$$\mu|\tau \sim N(\mu_{mu}, \tau_{mu})$$

$$\tau \sim \Gamma(\alpha, \beta)$$

$\{\mu_{mu}, \tau_{mu}, \alpha, \beta\}$ are assumed to be known. Γ is the Gamma function (Johnson et al., 1994)($\alpha = k$ is called shape parameter and β is called rate parameter):

$$\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta x}$$

$$x \geq 0, \ \{\alpha, \beta\} > 0, \ \Gamma(\alpha) = (\alpha - 1)!$$

The teoric moments are: $E[\tau] = \frac{\alpha}{\beta}, \ V[\tau] = \frac{\alpha}{\beta^2}$.
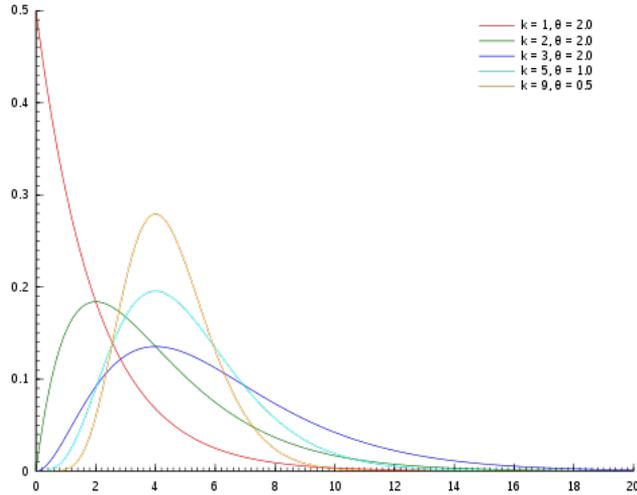
Figure 2.2: Gamma probability mass function

The status of "no prior information"is formalized with N(0,0.001) and
$\Gamma(0.001, 0.001)$.
Posterior parameters take the form ($\bar{x}$ is the empirical mean, n the sample
size):

$$\mu_{post}|\tau, \{x_1, ...x_n\} \sim N(\frac{\tau^{-1}\mu_{mu}n^{-1} + \tau_{mu}\bar{x}}{\tau_{-1}n^{-1} + \tau_{mu}}, (\tau n + \tau_{mu})^{-1})$$

$$\tau_{post}|\{x_1, ...x_n\} \sim \Gamma(\alpha + \frac{n}{2}, \beta + \frac{\sum_i(x_i - \mu)^2}{2})$$

2.2.4   The binomial distribution

A Bernoulli experiment is a statistical experiment that has the following
properties:

- The experiment consists of n repeated trials;

32

- Each trial can result in just two possible outcomes, one is called success and the other failure. Often events are replaced with indicator variables which take the value 1 whenever the outcome is success, and take the value 0 whenever failure happens;

- The probability of success, θ, is the same in every trial;

- The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

The binomial distribution (Johnson *et al.*, 2005) formalizes the answer to the question: what is the probability to have x successes in n Bernoulli trials. The formula is:

$$P(X = x; n, \theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \quad x = 0, ..., n, \quad \theta \in [0, 1], \quad n > 0$$

The teoric moments are: $E[X] = n\theta$, $V[X] = n\theta(1 - \theta)$.
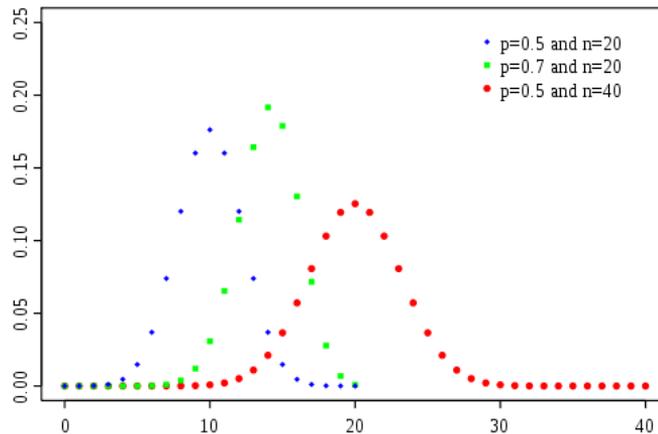
Figure 2.3: Binomial probability mass function

33

When θ is unknown the conjugate analysis establishes it to take the form of a Beta distribution:

$$P(\theta = t) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du}$$

The teoric moments are: $E[\theta] = \frac{\alpha}{\alpha+\beta}$, $V[\theta] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.
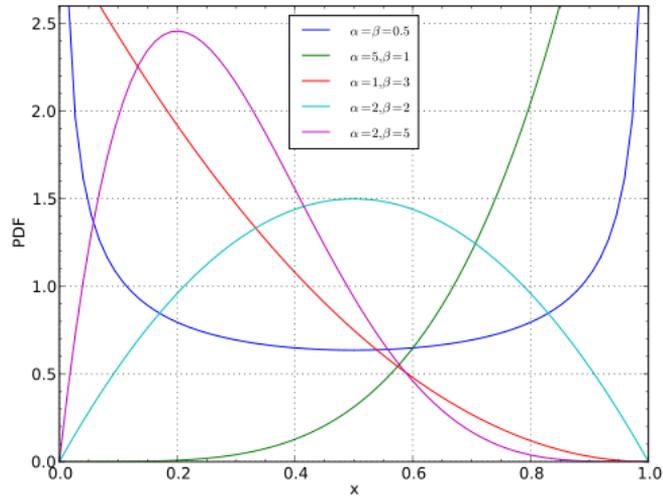
Figure 2.4: Beta probability mass function

The updated parameters are $\alpha' = \alpha + x$ and $\beta' = n + x + \beta$. The notion of "no prior information"(on β) is formalized with parameters $\alpha = \beta = 1$ which lead the Beta distribution to be equal to a uniform.

The model can be generalized to include a logistic regression, which linearises and normalises the relationship. The new model is:

$$X \sim Bin(\theta, n), \quad logit(\theta) = log(\frac{\theta}{1-\theta}) \sim N(\mu, \tau)$$

In this way all the arguments already mentioned about the normal model apply to the binomial.

2.2.5 Time series

One of the principal aims of this analysis, is to choose an appropriate model for predicting target quantities on the basis of their past observations. Such models are called autoregressive (Box and Jenkins, 1976), (Shumway and Stoffer, 2006),(Brockwell and richard A. Davis, 2002).

In the simplest autoregressive model, the value of the outcome variable depends linearly upon its immediate predecessor and upon a stochastic error term. The model, called $AR(1)$, typically takes the form:

$$Y_t = \alpha + \beta Y_{t-1} + \varepsilon_t \; ; \; t = 1, 2, ..., T$$

where α represents the mean level of the outcome, and β stands for the autocorrelation between two successive observations. The error term ε is usually assumed to be a white noise, $\varepsilon_t \sim N(0, \tau)$, with $\tau = \frac{1}{\sigma^2}$ constant across all time points t, and $Cov(\varepsilon_s, \varepsilon_t) = 0 \; \forall s \neq t$. These assumptions lead to a normal distribution for $Y_t \sim N(\alpha + \beta Y_{t-1}, \tau)$, independent from past observations given the immediate predecessor, $Y_t \perp\!\!\!\perp Y_{t-k(>1)}|Y_{t-1}$.

Usually, this model relays on the assumption of first and second order stationarity, namely $E(Y_t) = \mu$ and $V(Y_t) = \sigma^2 \; \forall t$, and if data shows temporal trend, it has to be removed before the analysis starts, by detrending techniques or successive differences. Stationarity entails the threshold $|\beta| < 1$.

In the Bayesian context, parameters are treated as random variables. Commonly $\beta \sim N(\mu_\beta, \tau_\beta)$ and precision $\tau \sim \Gamma(a, b)$, where $\{\mu_\beta, \tau_\beta, a, b\}$ are known parameters. Moreover, in the Bayesian context there is no need imposing any constraints, the proportion of coefficients exceeding the threshold can be verified a posteriori (Congdon (2006), pag. 282).

Finally, the Y series needs to be initialized before data are collected. Usually, this is treated as an extra parameter with fixed effect $Y_0 \sim N(\mu_0, \tau)$, the

effect of that choice decrease with series size. In short, the model is:

$$M^{AR1} = \begin{cases} Y_t|\mu_t, \tau \sim N(\mu_t, \tau), \ t = 1, 2, .., T \\ \mu_t|\tau = \beta Y_{t-1} \\ \beta \sim N(\mu_\beta, \tau_\beta) \\ \tau = \frac{1}{\sigma^2} \sim \Gamma(a, b) \\ Y_0 \sim N(\mu_0, \tau) \end{cases}$$

Autoregressive models with fixed parameters are often effective in the sample period, but time-varying parameters models (West and Harrison, 1997) have perhaps a greater fit in modelling short term forecasts when parameters evolve through time. Both of them will be used, the first because of its parsimony in estimating few parameters, and thus, more easily communicable to non expert auditorium. The second, because of its high flexibility and closeness to reality. Previous time series above assumes fixed β coefficient. Actually, relationships between variables are likely to vary over time.

One class of models dealing with that variation and with good statistical properties is Dynamic Linear Models (DLM) (West and Harrison, 1997),(Harrison and West, 1991),(Campagnoli *et al.*, 2009). The most simple case of DLM is a time-varying first order autoregressive model, TVAR(1), in which β_t follows a random walk:

$$M^{TVAR1} = \begin{cases} Y_t \sim N(\mu_{Y_t}, \tau_{Y_t}), \\ \mu_{Y_t} = \beta_t Y_{t-1} \\ \beta_t \sim N(\beta_{t-1}, \tau_{\beta_t}) \\ NIP : (\beta_0, Y_0) \sim N(0, .001), \ (\tau_{Y_t}, \tau_{\beta_t}) \sim \Gamma(0.001, 0.001) \\ IP : \beta_0 \sim N(\mu_{\beta_0}, \tau_{\beta_0}), \ Y_0 \sim N(\mu_{Y_0}, \tau_{Y_0}) \\ \tau_{Y_t} \sim \Gamma(rate_{\tau_{Y_t}}, shape_{\tau_{Y_t}}), \ \tau_{\beta_t} \sim \Gamma(rate_{\tau_{\beta_t}}, shape_{\tau_{\beta_t}}) \end{cases}$$

In this model β depends on t through a Gaussian random walk. Priors regard both the initialization and the precision of observations (Y_0, , τ_{Y_t}) and parameter (β_0 , τ_{β_t}).

2.3 Introduction to causal modelling

Sometimes knowing that two events are associated each other is not sufficient to ask questions we are interested in. This is especially the case when one event is interpreted as a cause and the other as the effect. This is because association is a symmetric relationship while causation is asymmetric, therefore a single association relationship gives rise to two distinct causal interpretations: a casual relationship from the cause to the effect and a diagnostic relationship from the effect to the cause. Here a brief example: to be pregnant (BP) and pregnancy test result (PT) are two associated events, it is very easy to build a contingency table and compute the conditional probability P(TP=yes|BP=yes) or the opposite P(BP=yes|PT=yes), both of them different from the marginal probability. But thinking causally, only the former probability makes sense because being pregnant is a cause the test to be positive, but not the other way round, that mean making the test positive doesn't cause any pregnancy. Formally P(BP=yes|PT=yes)=P(BP=yes). So, what is the interpretation of the P(BP=yes|PT=yes)? the interpretation relies on the diagnostic thinking: observing PT=yes gives information not about its acting as a cause, but about the truth of the cause BP=yes.

Pearl's methodology (Pearl, 2009) formalizes such difference using the (standard) notation of conditional probability for describing observational relationships ("given that you see"), while using a (new) notation of do() operator at the right side of the conditional bar (|) for describing causal effects ("given that you do"). Causal analysis requires more knowledge of the data generation process than observational analysis, see the Simpson's paradox (Pearl (2009), Chp 6), or the problem of the thousands of prisoners (Pearl (1988), p. 60).

Such methodology combines features of structural equation models with the explicit use of latent variables Bollen (1989), Lee (2007) with potential-outcome Rubin (2005) and graphical models (Pearl (1995), Pearl (2009), Spirtes *et al.* (2001)). It gives an operational definition of intervention,

37

with use of a new notation, $(do(X = x))$, and criteria for their identification (*do-calculus*, Pearl (1995)). Establishment of criteria for identifying independence implicit in the model, d-separation Pearl (1988), *back-door* (Pearl, 2000) and *front-door* (Pearl, 1995) criteria.

In Section 2.3.1, we introduce the structural equation model and their use in the pioneering work of geneticist Wright (1921). In Section 2.3.2, we introduce the terminology of the language of graphs. Paragraph 2.3.3 develops the concept of intervention, two operational criteria for its identification and a set of inference rules for its calculation. Finally, paragraphs **??** introduces the counterfactual analysis.

2.3.1 Structural equation models

The usual statistical modelling provides a formalization of the relationship between response variable and regressor set through a *single* function:

$$Y = f(X_1, ..., X_k, \varepsilon)$$

where Y is the variable of interest, ε represents the random component (under the usual conditions $\epsilon \sim N(0, \sigma^2)$, $(X_1, ..., X_k)$ is the vector of regressors and the function f represents the deterministic component. If regressors can be considered as random variables, the model predicts the conditional expected value for each (fixed) realization of the vector of regressors $E[Y|X_1 = x_1, ..., X_k = x_k]$. The systematic and random components are uncorrelated by assumption. When f is linear it assumes the well-known form:

$$Y|(X_1 = x_1, ..., x_k X_k =) \sim N(\beta_0 + \beta_1 x_1 + ... + \beta_k x_k, \sigma^2)$$

The structural equation model (SEM) differs from this paradigm as it is endemically multi-relational: there is no single a priori response variable, but each variable is defined by the direct relationship with a subset of the

38

other variables in the model:

$$X_i = f_{x_i}(X_1, ..., X_{i-1}, X_{i+1}, .., X_j, \varepsilon_i) \quad j = 0, ..., k \quad \forall i = 1, ..., k$$

This modelling paradigm is much more flexible than the previous one. The status of a variable (dependent / independent) is not a property of the variable itself, but it depends on the relationship with the others: each variable may have both the meaning of the cause (independent variable) in one equation and effect (dependent variable) in another equation. It explicitly models relationships between regressors (if none, it is the ordinary case). It also distinguishes several types of effects, such as the direct effect between variables, $(X_1 = f_{X_1}(X_2, \varepsilon_{X_1}))$, or mediated effects (if $X_2 = f_{X_2}(X_3, \varepsilon_{X_2})$ then X_1 depends on X_3 through X_2).

These models answer the question: how can you express mathematically the common claim that the symptoms do not cause disease? The first attempt to answer this question comes from the geneticist Wright (1921), who used a combination of equations and graphs to communicate causal relations. For example:

$$y = \beta x + \varepsilon_y$$

where x is the level of disease, y the level of symptoms and ε_y represents all other factors that could affect symptoms (for fixed levels of the disease). This equation does not express the causal relationship contained in the statement above because the algebraic equations are symmetric relations in the sense that you could write:

$$x = \frac{1}{\beta}y - \frac{1}{\beta}\varepsilon_y$$

and this equation could be misinterpreted as "symptoms cause disease". To express the directionality of the process, Wright coupled the equation with a figure, later called "path diagram", where arrows start from (presumed) causes and end into (presumed) effects, and moreover, absence of

an arc indicates no direct relationship between variables, and therefore the absence of an arc from Y to X is interpreted as the absence of Y among the causal factors that influence X.

Figure 2.5 can be thought as a visual aids to inspect the "causal direction":
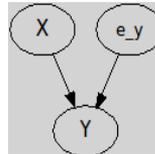
Figure 2.5: graph of a simple regression equation

Basically, the diagram encodes the assumptions of causal effects of X on Y, and the lack of causal influence of Y on X, whereas the equation above encodes quantitatively the strength of that relationship, e.g. β, called path coefficient, measures the causal effect of X on Y.

2.3.2 The language of graphs

Since they appeared in the early twentieth century, structural equation models were accompanied by graphical representations (path-diagrams) that provide a concise representation of the assumptions involved in the qualitative model.

In the eighties, these diagrams have been interpreted as probabilistic models (Bayesian networks) and then in the nineties, as a tool for a formal causal inference (causal networks), that is as predictive tools for effects of external intervention.

In this section, it is introduced the terminology used in the language of graphs.

The term "*dependency*" in a graph, usually represented by a link or connection between variables, can refer to a mathematical, causal or statistical dependence. In a graph, a connection is represented by *arcs*, or *links*, and

variables are represented by *node* or *vertices*. Two variables connected by an arc are *adjacent* , as well as two arcs that meet at a node. If the arc is an *arrow*, the nodes from which it comes are the *parents* (Pa), and the one in which it ends is the *child*, (Ch).

In causal diagrams, an arrow represents a "direct effect" of the parent on the child, although this effect is direct only in relation to a certain level of abstraction since variables lying on that link could be missing from the graph.

A variable with no parents is called *exogenous* or *root* and is determined only by variables outside the graph (and therefore not under study), otherwise it is called *endogenous*. A variable with no children is called *leaf*.

A *path* or *chain* is a sequence of adjacent arcs. A path is a *directed* path of arrows drawn entirely from the same direction ($A \rightarrow B \rightarrow C$). If there is a direct path from X to Y, then X is an *ancestor* (An) of Y and Y is a *descendant* (De) of X. In the diagrams, direct paths represent causal pathways from the starting variable, called cause, to the variable of arrival, called effect. A graph is called *direct* if all arcs are arrows, *acyclic* if there are no cycles, namely direct routes that depart from a variable and end in itself. If both features hold it is called DAG (Direct Acyclic Graph) and represents a complete causal structure as all sources of dependence are expressed through causal connections.

A variable *intercepts* or *mediates* a path if it lays on the path (but not at the end). Variables that intercept direct routes are called *intermediate*. In this case, it can be two types of paths, *serial* ($A \rightarrow B \rightarrow C$) or *divergent* ($A \leftarrow B \rightarrow C$). A variable is called *collider* and the path *convergent* if the path enters and exits through the tip of the arrow variable (e.g. $A \rightarrow B \leftarrow C$). A path is *open* or *unblocked* if it is formed by serial or divergent paths and is *closed* or *blocked* if it contains at least one collider node. So a path not containing colliders is *open* or *active* (e.g. $A \leftarrow B \leftarrow C \rightarrow D$) while a path with collider is *closed* or *inactive* (e.g. $A \leftarrow B \rightarrow C \leftarrow Q \rightarrow E$).

Two variables (or set of variables) are *d-separated* (Pearl, 1988) if there is no

open path between them (or each of them). Formally, a path p in a graph G is called d-separated by a set $\mathbf{Z}$ if and only if either:

1. p contains one of the following patterns: $I \to M \to J$, $I \leftarrow M \leftarrow J$, $I \leftarrow M \to J$ such that $M \in \mathbf{Z}$ or;

2. p contains the following pattern: $I \to M \leftarrow J$, such that neither $M \in \mathbf{Z}$ nor $De(M)_G \in \mathbf{Z}$.

The constraints imposed by a graphical model correspond to independence arising from d-separation, for example, the absence of an open path between A and E in $A \leftarrow B \to C \leftarrow A \to E$ constrains the two variables to be marginally independent. However, the reverse is not true, that the presence of an open path dependence does not imply independence, in fact, it could arise from the so-called violation of *faithfulness* (Spirtes *et al.*, 2001): $E \leftarrow A \to F$, E and F may be marginally independent if the dependencies $E \leftarrow A$ and $A \to F$ cancel each other (equal intensity and opposite direction, e.g. $E = \beta A + \varepsilon_E$, $F = -\beta A + \varepsilon_F$). However, for first approximation faithfulness is normally assumed.

A bidirected arc $\leftrightarrow$ represents two variables sharing the same hidden ancestors, e.g. $A \leftrightarrow B$ means that there is a third unobserved variable U with directed paths to A and B ($A \leftarrow U \to B$). Every set of equations induces a DAG, is by associating a node to a variable, and a set of link starting from the independent variables, either endogenous or exogenous, and ending into the dependent. Figures 2.6 and 2.7 show the induced DAG of an AR(1) and TVAR(1) models:
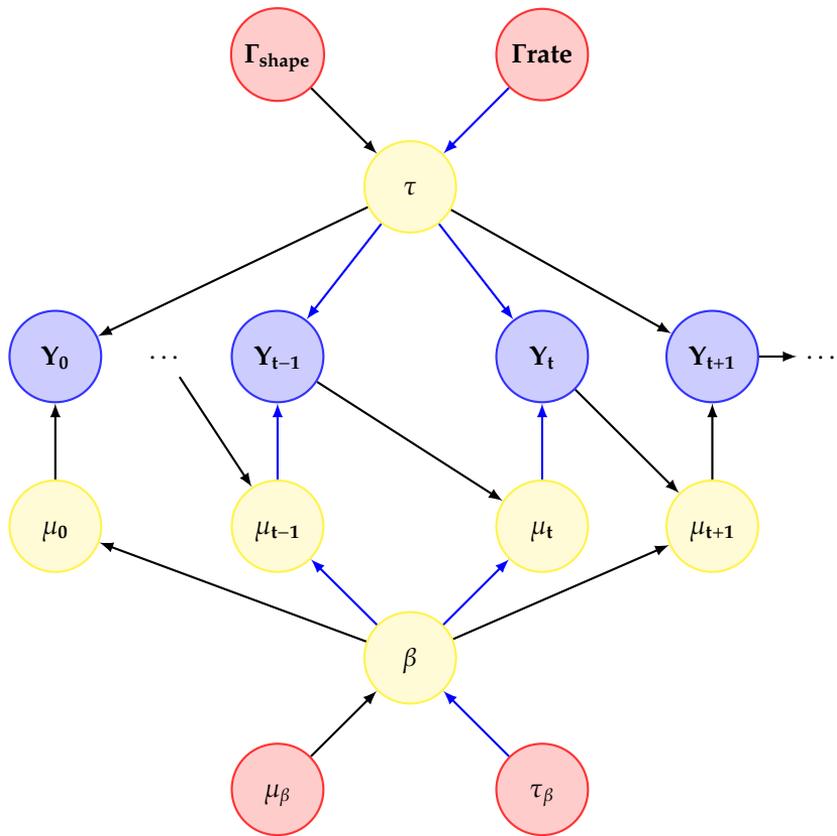
Figure 2.6: AR(1) inducing DAG
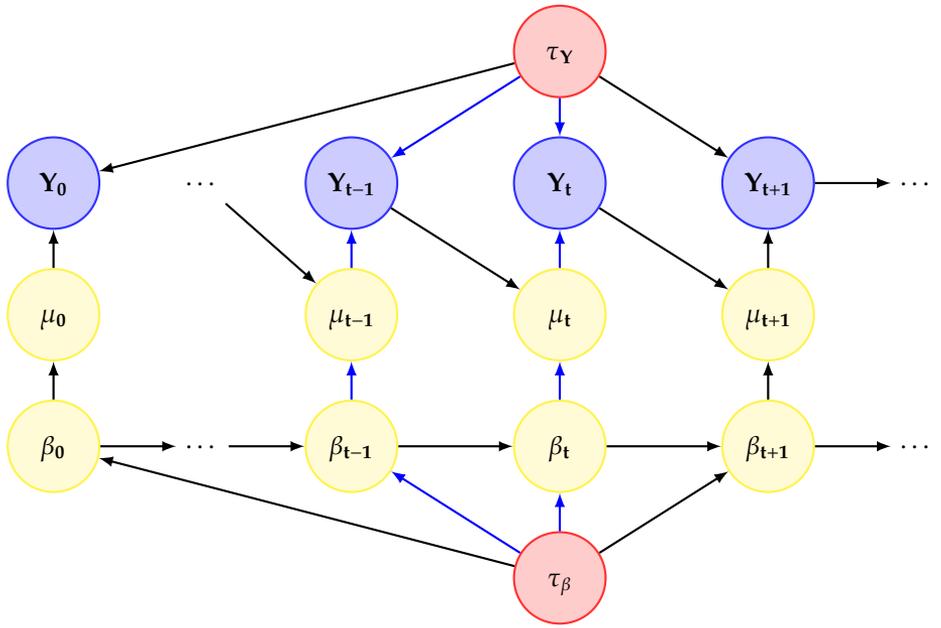
Figure 2.7: TVAR(1) inducing DAG

Finally, a personal comment: it has to definitely distinguish between genuine graphical analysis and instrumental graphical analysis. Usually in statistics, graphical analysis is associated either with a first exploratory stage (e.g. histograms, plots), or as merely representation of algebraical manipulation (e.g. a zero values in the normal covariance matrix out of principal diagonal, entails a missing link in a DAG), or later in order to verify some assumptions, such as the shape of the distributions of residuals. The methodology of Pearl, in contrast, uses the graphical language as primary and predominant: the model, although defined in terms of (structural) equations, is always put together with its graph. The d-separation criterion is a tool that helps in identifying, by simple inspection, the independences implied by the model, the concepts of causal and counterfactual effect are related to the graph in terms of eliminating some of its arcs. Identification of effects (both, total, direct and indirect (Pearl, 2001)), mediation (Pearl, 2012), transportability (Pearl and Bareinboim, 2011) are all defined in terms of *graphical* criteria. Computation, namely solving equations, comes afterwards.

2.3.3 Nonparametric models: operator do(x) as mathematical tool for interventions and counterfactuals

If the function f_{x_i} in $f_{x_i}(X_1, ..., X_{i-1}, X_{i+1}, .., X_j, U_i)$[4] is not specified, the structural model is called nonparametric[5]. Unlike the parametric linear model in which causal effects are defined algebraically by the regression coefficients, in nonparametric models, the effect is interpreted as the ability to change the function structure using the invariance properties of the structural equations.

---

[4]ε_i is substituted with U_i according with the original notation

[5]Here nonparametric does not refer to the nonparametric statistical estimation of the function f_{x_i}, but to what matters for drawing causal conclusions, that is the independences entailed in the graph G identified through the d-separation criterion, beyond any functional form.

A probabilistic causal model (PCM) (Pearl, 2009) is defined as a tuple M = $\langle U, V, F, P(u) \rangle$, where:

- $U = (U_1, ..., U_m)$ is a set of exogenous variables;

- $V = (V_1, ..., V_n)$ is a set of endogenous variables. These variables are functionally dependent on a subset of $U \cup V$;

- F is a set of functions such that each f_i maps a subset of $U \cup V \setminus \{V_i\}$ in V_i, and such that V is a function of U through F;

- $P(u)$ is a joint probability distribution of U.

For example:

$$z = f_Z(u_Z)$$
$$x = f_X(z, u_X)$$
$$y = f_Y(x, u_Y)$$

It follows the induced DAG:
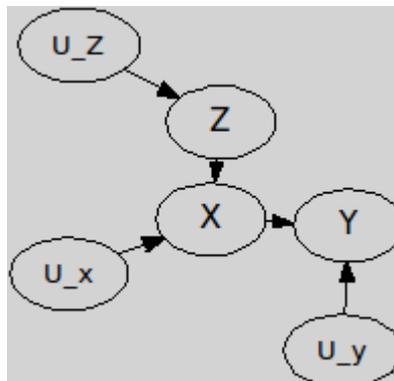
Figure 2.8: The induced DAG

Each of these equations represents a causal mechanism that determines the value of the variable on the left of the equal sign through those on

46

the right of the same. The absence of a variable on the right side of the equation encodes the assumption that the *Nature/Mechanism/Agent* ignores that variable in the process of determining the output variable.

This set of equations is called *structural*, if it is assumed that they are autonomous, that is, invariant to possible changes in the shape of the other functions. When $P(U)$ factorizes in jointly independent components the model is called Markovian. In this case, the following theorem holds: any joint distribution generated by a Markov model M can be factorized as:

$$P(X_1, ..., X_k) = \prod_i P(X_i|PA[X_i])$$

where $X_1, ..., X_k$ are endogenous variables in M and $PA[X_i]$ are the parents of X_i in the causal diagram associated to M. For the example above it follows:

$$P(X, Y, Z) = P(Z|U_Z)P(X|, U_X)P(Y|, U_Y)$$

where U_X, U_Y, U_Z are jointly independent but otherwise arbitrarily distributed. This factorization introduces more than one relationship between a graph G and a joint probability distribution $P(X_1, ..., X_k)$, e.g. the same joint probability distribution may arise from different graphs ($X \rightarrow Y \rightarrow Z$ and $X \leftarrow Y \leftarrow Z$ induce by the same joint probability because they entail the same set of independences).

The invariance feature allows for using structural equation for modelling causal effects (and counterfactuals). This is done using a new mathematical operator called *do(X = x)* (or simply *do(x)*), that simulates the consequences of a physical intervention (action) on the variable X by removing its defining functions from the model and replacing X as argument of other equations with a constant $X = x_0$, regardless of its *natural* value $f(Pa[X])$, and leaving everything else unchanged. Graphically, this resolves in removing the arcs ending in X and replacing X with the node x_0.

For example, to simulate an operation on the variable X, the new system of

equations is:

$$z = f_Z(U_Z)$$

$$x = x_0$$

$$y = f_Y(x_0, U_Y)$$

This operation, called wiping out equation in Strotz and Wold (1960), leads to the following corollary: a Markov model generated by a $do(X_j = x_j)$ is given by the truncated factorization $P(X_1, ..., X_{j-1}, ..., X_{j+1}, ..., X_k | do(x_j)) = \prod_{i|X_i \neq x_j} P(X_i | pa[X_i]) x_j$ where $P(X_j | PA[X_j])$ are conditional distributions pre-intervention. If M is the original Markov model, after the factorization associated with the truncated model it will be called submodel and denoted by M_{x_j}.
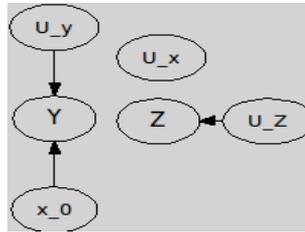
Figure 2.9: The induced DAG by the submodel M_{x_j} after an intervention $do(X_j = x_j)$

When the goal is to compute the effect of an intervention on a particular variable Y, the distribution $P(Y = y | do(x))$ is called *post-intervention*. In general, a post-intervention distribution can be defined as:

$$P_M(y | do(x)) \triangleq P_{M_x}(y)$$

In other words, the post-intervention distribution of the response variable Y is defined as the probability that the submodel M_x assigns to Y. A key aspect on causal analysis is the *identifiability* of causal effects: if, and how,

it is possible to calculate the post-intervention distribution $P(y|do(x_0))$ from observations governed only by the pre-intervention distribution $P(y, x, z)$. For example, in the previous graph, the value of $E(Y|do(x_0))$ is identifiable and is given by the conditional expectation $E(Y|X = x_0)$, regardless of the functional forms f_Y, F_X, F_Z and distribution $P(u)$:

$$E(Y|do(x_0)) = E(f_Y(x_0, U_Y))$$

the expected value of pre-entry is:

$$
\begin{aligned}
E(Y|X = x_0) &= E(f_Y(X, U_Y)|X = x_0) \\
&= E(f_Y(x_0, U_Y)|X = x_0) \\
&= E(f_Y(x_0, U_Y)) \\
&= E(Y|do(x_0))
\end{aligned}
$$

The identification problem is solved in Shpitser and Pearl (2008).

The do(X=X) operator, also called atomic intervention, is defined as the simplest type of external intervention (Pearl (2009), pag. 70). Some other type could be: to change the functional form, to delete a parent relationship, to add a parent relationship, to change the relationship (e.g. form child to parent or from parent to child), to add variable to the system, to delete a variable from the system. Notice that the do(X=x) intervention corresponds to performing a "delete a parent from the relationship" on all parents of X before setting X=x.

In this section, a distinction was made between identifying an effect and estimating it. This is because the two operations are distinct: to identify a causal effect means being able to derive from a quantity $P(Y|do(x))$ a quantity "*do*-free". Instead, to estimate means calculating the effect of its distribution on its average value or other characteristics, using any criteria, subjective, classical, frequentist or Bayesian. None implies the other. It is possible to identify an effect without being able to estimate it. For example, because the sample size is too small or because the goodness of fit is too low.

The do(X=x) operator can be thought as the simplest counterfactual sentence. Let us introduce counterfactuals first: I had an headache, I took the aspirin, headache is gone. Had I not taken the aspirin, would I be equally good?

I took the aspirin is the observed event, while had I not taken the aspirin is an hypothetical (not observed) event called counterfactual (Lewis, 1973). The formal definition is:

$$Y(x, u) \triangleq Y_{M_x}(u)$$

which reads, the sentence "the value that Y would assume in unit u, had X been x", is the solution for Y in the submodel M_x where the function for X is replaced by the constant x, being u the status of the exogenous variables (Balke and Pearl, 1994), or a specific unit in the potential outcome framework (Rubin, 1974).

Moreover, this hypotetical event is connected to the observed one through the consistency axiom (Robins, 1987), (Pearl (2009), pag 229):

$$X(u) = x \Rightarrow Y(u) = Y_x(u)$$

which reads, when the observed value of $X(u)$ is x, the counterfactual value of $Y_x(u)$ is equal to the observed value $Y(u)$.

Given a simple structural system:

$$\begin{cases} y = bx + e_1 \\ z = cy + e_2 \\ corr(e_2, x) = 0 \end{cases}$$

The empirical claim of the structural equation associated to this counterfactual is :

$$Y(x, z, u) = f(x, e(u))$$

for any set Z not intersecting X or Y. This reads as: had X and Z been x and z, respectively, Y would be $f(x, e(u))$ independently of z and of the other equation(Pearl (2009), pag 380). For more information about testable

50

counterfactuals refer to (Shpitser and Pearl, 2007).

In decision making situation, probability like P(future| do(action), see(context))[6] (Pearl (2009), pag 392) can be formulated in counterfactual form (see 3.6.1 for an example).

2.3.4 Observation vs manipulation vs counterfactual in a nutshell

This brief section aims to explain the difference between observation, manipulation and counterfactual with a trivial deterministic example.

In the beginning was the data generating model:

$$DGM \begin{cases} y := 3x \\ x := 2z \\ z := 4t \end{cases}$$

The system read as: y is generated according to x values, x is generated according to z values, z is generated according to t values. {x,y,z} are observed. {t} in not. {x,y} are endogenous, {z} is exogenous. X generates Y means, paraphrasing Hume (1976), "if the first object had not been, the second never would have existed".

Now let us assume to observe x=2. The new set of equations is:

$$OBS \begin{cases} y := 3 * 2 \\ 2 = 2z \\ z := 4t \end{cases}$$

Observing x=2 gives causal information about y, which is generated according to x, but gives also diagnostic information about z: the only z-generating value compatible with the generated value x=2 is 1. For the same reason it turns out t=1/4. The set of solutions is {y=6,x=2,z=1,t=1/4}.

---

[6]under the convention that the context is observed before the action is made.

Now let us assume to manipulate do(x=2). The new set of equations is:

$$INT \begin{cases} y := 3 * 2 \\ x := 2 \\ z := 4t \end{cases}$$

Literature says manipulating x:=2 is interpreted as wiping out the generating equation x:=2z and substitute it with the the equation x:=2. Such substitution gives only causal information about y, but no diagnostic information about z (the effect does not affect the cause). The set solution is indeterminate: {y=6,x=2,z=4t,t}.

Now let us assume the counterfactual {x=2,do(x=3)}, namely observing x=2 and after manipulating do(x=3). The new set of equations is:

$$COU \begin{cases} y := 3 * 3 \\ x := 3 \\ 1 = 4t \end{cases}$$

In the first equation y receives causal information from the manipulation (which overrides the former causal information from the observation x=2). In the second equation x is manipulated (as in the INT system). In the third equation z receives diagnostic information from the observation x=2 (as in the OBS system) while no information is received from the manipulation. The set solution is: {y=9,x=3,z=1,t=1/4}.

2.4 The skeleton

In unix-like systems it exists a directory named /etc/skel. It contains directories and files that will be copied into the personal user directory whenever a new user will be added to the system.

Following the same logic this section contains statistic prototypes will be

performed whenever a new indicator will be added to the set.

The goal of this thesis is monitoring variables responsible for FFO indicators. Monitoring has been interpreted as a fourfold concept translating in describing, analysing retrospectively, predicting and intervening on indicators. A detailed description follows.

To describe an indicator means to provide:

1. the description of the indicator, both in prose and in formal terms, through the system of structural equations it describes and the relative induced DAG;

   Here an example from $A2$ indicator:

   Indicator $A2$ aims to measure student abilities to earn the credits required each year by the curriculum. This capacity is measured by the ratio (R) between the actual CFU earned (EFF) and the teoric of CFU to be earned (TEO) in one academic year. The latter is typically 60 per year, the former a part of it. Similar to $A1$, the ratio is computed separately for the 4 degree courses groups. The different capabilities of credit acquisition between groups is taken into account by dividing such ratio by the national median value of each group ($Median(R)$ or MR). This results in a coefficient ranging around 1. Finally, this coefficient multiplies the number of actual credits. TEO and EFF are both function of the number of students (STD), $EFF \sim Bin(\theta, TEO)$, local and national variables, $\{S^*, I\}$ and $\{R^*, R\}$, could be correlated due to common national factors.

The structural equation system is:

$$A2 \begin{cases} TEO_X = 60STD \\ EFF_X \sim Bin(\theta_X, TEO_X) \\ R_X = 100\frac{EFF_X}{TEO_X}, \quad X = \{A, B, C, D\} \\ R_X^* = \{R_{1,X}, ..., R_{53,X}\} \\ MR_X = Median(\{R_X, R_X^*\}) \\ WCFU_X = \frac{R_X}{MR_X}EFF_X \\ I = WCFU_A + WCFU_B + WCFU_C + WCFU_D \\ Corr(S^*, I) \neq 0 \\ Corr(R_X^*, R_X) \neq 0 \\ S = \sum_{i=1}^{54} I_i = I + \sum_{i=1}^{53} I = I + S^* \\ A2 = 100\frac{I}{S} \end{cases}$$

As already mentioned in the section dedicated to the graph theory (section 2.3.2), every set of structural equations induces a graph in which every variable is represented by a node, and every equation is represented by a parent/child connection, from every node representing an independent variable starts an arrow ending at the dependent variable. Exogenous variables are shown in red, endogenous variables are in blue.

The induced DAG is represented in the next figure:
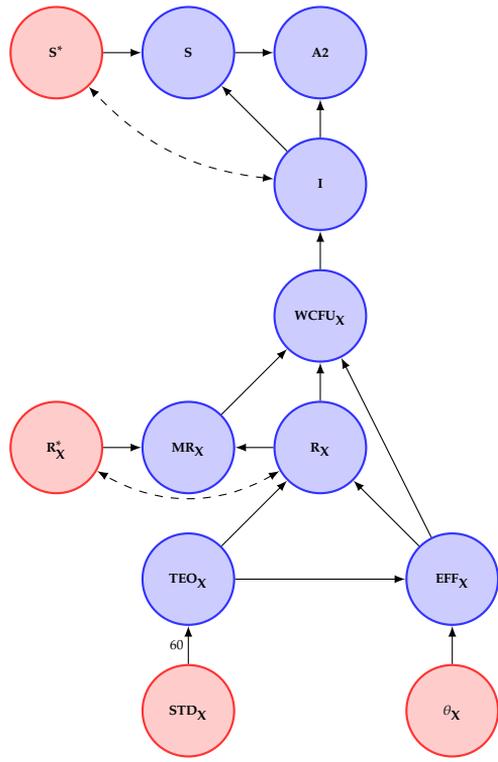
Figure 2.10: *A*2 induced DAG

2. to define the timetable of every variable involved in the indicator construction, that is for three consecutive years what time variable values will be available or need to be estimated. Next table continues with $A2$ example:

Year	T		T+1		T+2	
Quantity	RT	AT	RT	AT	RT	AT
EFF	2009	01/01/10	2010	01/01/11	2011	**01/01/12**
TEO	2008/09	01/01/09	2009/10	01/01/10	2010/11	01/01/11
MR	2010	12/10	2011	**12/11**	2012	**12/12**
S	2010	12/10	2011	**12/11**	2012	**12/12**
A2	2010	12/10	2011	**12/11**	2012	**12/12**

Table 2.2: A2 Reference Time and Available Time

3. to assign to every variable its observed value in order to obtain the 2010 indicator:

$$A2_{2010} \begin{cases} \text{EFF=\{63767,450706,428157,269522\}} \\ \text{TEO=\{105300,1110160,1231877,776298\}} \\ \text{R=\{60.56, 40.60, 34.76, 34.72\}} \\ \text{MR=\{59.7, 43.5,38.6,37.1\}} \\ \text{I=1122225.29} \\ \text{S=40656680.44} \\ \text{A2=2.76} \end{cases}$$

To perform a retrospective analysis means:

1. using the friendly what/if language, to implement counterfactuals aiming to discover past errors or delay in data entry, or to build hypothetical scenarios useful to discover what variables are more beneficial to intervene on;

Next example comes from $A1$ (see section 3.3 for $A1$ definition):

Query: In 2010 we observed $A1 = 3.66$, what value $A1$ could attain, were all regular students active (given that actually not all regular students were active)?

The question is answered in two steps: in the first, all variables are fixed to their observed values, obtaining (again) $A1 = 3.66$.
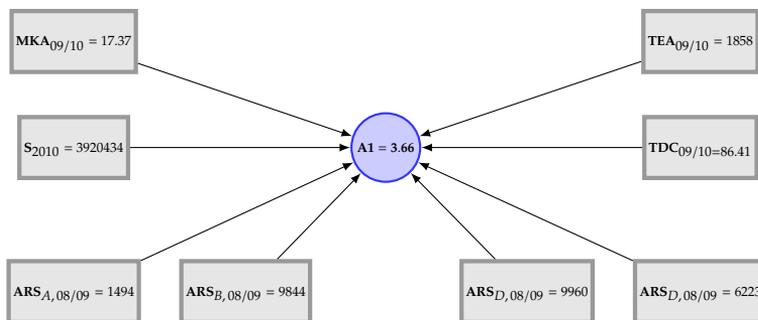
Figure 2.11: $A1$ marginalized DAG: observed values

In the second step, applying the wiping out equation operation Strotz and Wold (1960) Pearl (2000), variables are set according to the counterfactual antecedent, namely $ARS_A := RS_A$ and so on, then we compute again the $A1$ value getting the following answer: $A1$ would be 4.49.
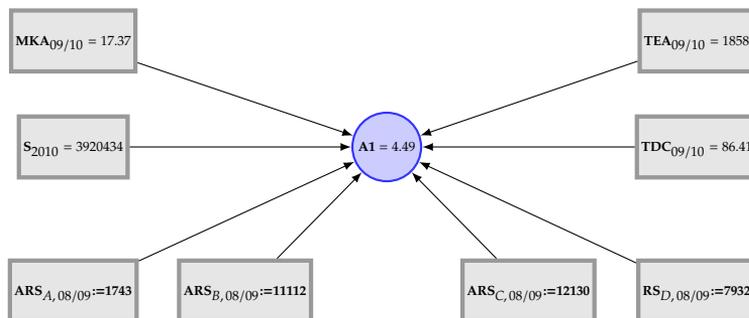
This toy example leads to the economic shocking conclusion that in 2010 inactive students cost 1M€, showing the essential need to reduce the great mass of students passing no exams in a years.

Nevertheless, counterfactuals are not observable, it is impossible to go backwards in time and make students active, and, moreover the computation above does not indicate a good interpretation for the future: it would be interpreted as the 2011 $A1$ value under the hypothesis of all variables remaining fixed at the 2010 value, except the active regular students. This is not feasible because all quantities are supposed to change between the years.

To perform a previsional analysis means:

1. to perfom a short term prevision, that is forecasting national variables (usually S the sum for all 54 local indicator values, and a median value MR of some local data, e.g. a ratio R) in order to make a prevision on one year ahead. The prevision policy is: whenever there is only one year observation available, the normal distribution is used because of its excellent mathematical properties, such as symmetry. When more than one year observation is available, time series analysis is used, in the context of Bayesian estimation, in particular AR(1)/TVAR(1). One goal is to test whether informative prior is needed in order to obtain a precise estimation.

    An example coming from $A1$: even though a causal diagram is depicted, being only one observation available on national data, it will used a normal approximation model.

    The causal diagram:
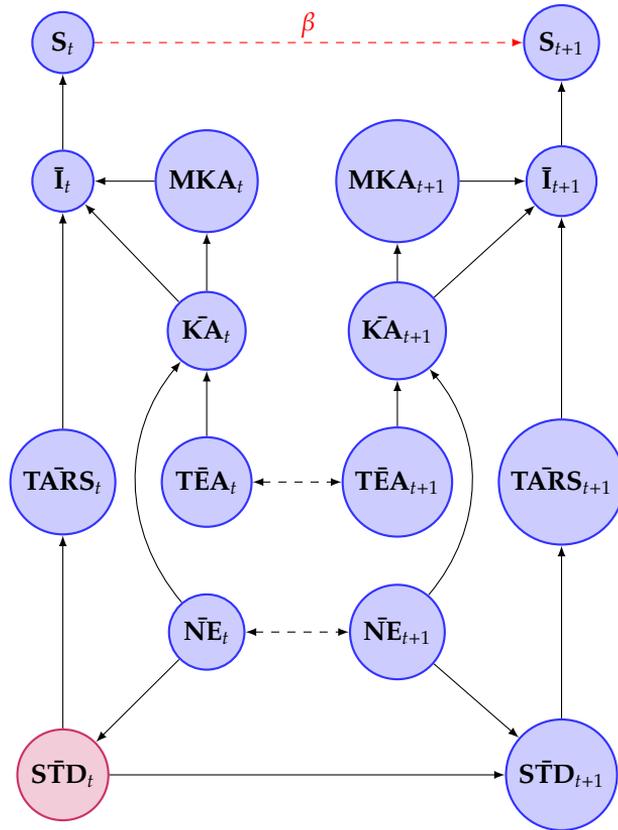
Figure 2.12: The spurious association between S_t and S_{t+1}

An example of the normal approximation model is:

$$S \begin{cases} S_{2011} \sim N(\hat{\mu}, \hat{\sigma}) \\ \hat{\mu} = k_1 S_{2010}^{obs} \\ \hat{\sigma} = k_2 \sigma(I_{2010}^{obs}) \\ \{k_1, k_2\} \in \{0.95, 1, 1.05\} \end{cases}$$

where S_{2010}^{obs} is the S observed value in 2010, and $\sigma(I_{2010}^{obs})$ is the empirical standard deviation among all 54 I values in 2010. $k_1 = k_2 = \{0.95, 1, 1.05\}$ are two coefficients useful for sensitivity analysis.

Without any empirical evidence, 5% is supposed to be rounded up because S as a sum of variables tends to balance increments and decrements of its own components.

An example of the output simulation table is:

Simulation table

n.iter=10000	S 2011				
Node	Mean	Sd	5%	50%	95%
0.95μ, 0.95sd	3588116	58391	3492952	3588202	3682550
0.95μ, sd	3587629	60623	3488732	3587611	3686982
μ, 0.95sd	3777204	57566	3682013	3777545	3872508
μ, sd	3776857	61075	3677620	3776782	3876240
1.05μ, sd	3965067	61221	3865571	3964356	4066440
μ, 1.05sd	3776653	63709	3671871	3777186	3880745
1.05μ, 1.05sd	3965929	64663	3860224	3966735	4071281

Table 2.3: $S_{2011} \sim N(k_1\hat{S}_{2010}, k_2\hat{\sigma} = \sigma(I_{2010}))$, $\{k_1, k_2\} \in \{0.95, 1, 1.05\}$

This simulation offers a grid of values inside which S should hopefully be. Summaries helps on deciding whether a value, e.g. 40000000, belongs to $(1.05\mu, \hat{\sigma})$ but not to $(0.95\mu, \hat{\sigma})$.

2. to perform a long term prevision, that is forecasting both national and local variables in order to make a two year ahead prevision on the indicator. Usually there are past observation so time series analysis are performed, usually AR(1) models. When data show trend on the parameters, time-varying AR(1) are used instead. Depending on the nature of the variable to be forecast, either binomial (also in the logit scale) or normal model are used. Every model is equipped with both non informative and informative prior distributions, in order to verify whether uninformative priors lead to the same forecasts of the informative ones, meanings data alone carry a sufficient amount of information to justify the weaker assumption.

Simulations are performed using jags software (Plummer, 2003) and rjags package (Plummer, 2011) of the R suite (R Development Core Team, 2011), in a winbugs style format. Tables of deviance come from (Plummer, 2008).

Structural equations, causal diagrams, table of deviance and table of estimates are provided. Here an example from A2 indicator.

The model is:

$$
M_{EFF}^{Bin} = \begin{cases} EFF_{i,t} \sim Bin(\theta_i, TEO_{i,t}), \ t = 1,2,3,4 \ \ i = \{A,B,C,D\} \\ NIP : \theta_i \sim Beta(1,1) \ \forall i \\ IP : \theta_A \sim Beta(58.67,44.34) \ \theta_B \sim Beta(1650,2536) \\ \theta_C \sim Beta(242.8,440.2) \ \theta_D \sim Beta(9023,16805) \end{cases}
$$

The induced DAG is:
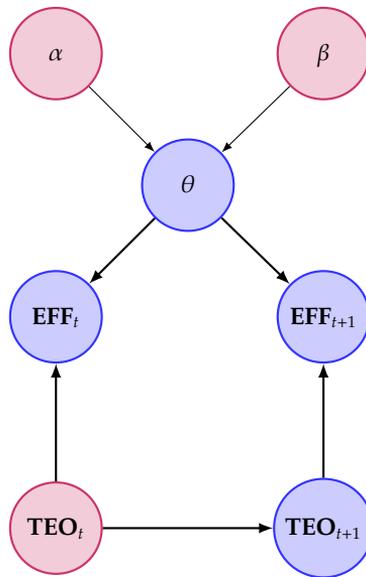
Figure 2.13: The spurious association between EFF_t and EFF_{t+1}

AR(1)	NIP/IP
Mean Deviance MD	18569
Penalty term P	4
Penalized deviance PD	18573

Table 2.4: EFF: fixed effects binomial model. NIP/IP deviance

The tale of estimates is:

n.iter	EFF - NIP/IP				
Node	Mean	Sd	5%	50%	95%
$A \ E\hat{F}F_{2011}$	88428.8	220.6	88066.0	88429.0	88792.0
$B \ E\hat{F}F_{2011}$	432306.5	568.3	431372.0	432307.0	433241.0
$C \ E\hat{F}F_{2011}$	378222.6	543.7	377328.0	378223.0	379116.0
$D \ E\hat{F}F_{2011}$	254550.4	450.4	253809.0	254550.0	255292.0
$A : \hat{\theta}$	0.569	0.001	0.567	0.569	0.570
$B : \hat{\theta}$	0.394	0.000	0.394	0.394	0.395
$C : \hat{\theta}$	0.355	0.000	0.354	0.355	0.355
$D : \hat{\theta}$	0.349	0.000	0.349	0.349	0.350

Table 2.5: EFF: fixed effects binomial model. NIP/IP estimation

Causal analysis deals with prediction under intervention upon variables on the system[7]. The best way to make this prediction is performing an actual intervention, predicting what the effect is going to be and, when it will be realized, evaluating the difference between prevision and observation. Here, a different task is performed: predicting the effect of an intervention without performing it. How is this possible? Two conditions are required: whether resting on some assumptions, called causal assumptions, are sufficient to narrow down to a model able to identify causal effect (Tian and

---

[7]"what happens when you chick something", in the Dawid informal definition http://videolectures.net/mlss09uk_dawid_caus/

Pearl, 2002) and whether a sufficent[8] amount of data is gathered. If these two conditions are satisfied, the interventional questions can be asked with pure observational data in which no intervention is taken.

To perform a causal analysis means:

1. policy evaluation. Using counterfactual to evaluate decisions like "is it useful increasing new enrolments"? (indicator $A1$, see section 3.6.1 for the whole story); Counterfactual diagram and computations are provided:
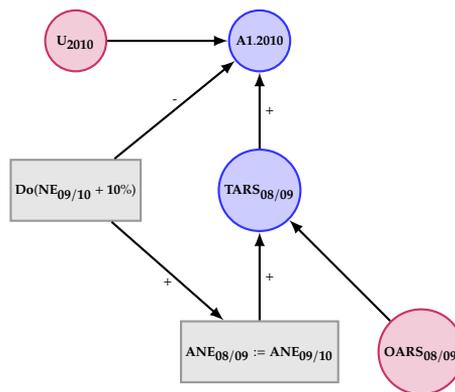
Figure 2.14: Counterfactual effect of $NE_{09/10}$ on $A1_{2010}$

In the light of these arguments, let us perform the computation:

- Observed $A1$ 2010=3.66;

- Direct effect (Pearl, 2001) $A1$ 2010[$NE_{09/10} := +10\%$] = 3.54;

- Counterfactual effect $A1$ 2010[$NE_{09/10} := +10\%, ANE_{08/09} := ANE_{09/10}$] = 3.69.

2. intervention policy. Building a strategy in order to get better results as "to equate, in an limited number of years, the local ratio R to the

---

[8]in the statistical sense, sufficient to make a precise estimation.

median value *MR* for every area."(it holds indicators *A2*/*B1*); Strategy rules and advancements are provided: Each year, every degree course compares *R* and *MR* and according to the difference defines its new target T, namely the new value it must reach:

C1) if $R \geq MR$ then $T \geq R$;

C2) if $0.95MR \leq R < MR$ then $T \geq MR$;

C3) if $0.90MR \leq R < 0.95MR$ then $T \geq 1.05R$;

C4) if $R < 0.90MR$ then $T \geq 1.1R$;

Next table shows for the next three years the percentage of degree courses in each class and the value of *A2*

Year	C1	C2	C3	C4	A2
2010	38.6	12.0	14.6	34.8	2.76
2011	40.3	11.4	8.9	39.4	2.81
2012	53.1	2.9	12.3	31.7	2.90
2013	56.0	3.1	13.7	27.1	2.97

Table 2.6: Percentage of target classes degree courses by area and year

The induced DAG is:
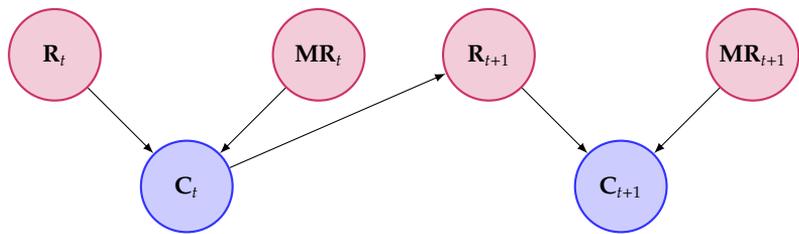
Figure 2.15: Act on EFF policy DAG

Chapter 3

Analysis

3.1 Analysis features

Before starting let me explain some of the features of this thesis. Since the principal scope is prediction, the most suitable statistical method is time series analysis. Nevertheless as far as it is possible the simplest statistical models (e.g. AR(1)) are used, moving all the complexity into the structure, meaning the set of relations among variables and the possible causal interpretation of that structure, namely the structure.

A warning to mention: due to a chronic lack of observations, massive use of a priori assumptions and ad-hoc imputations permitted by the Bayesian framework are used.

Much attention is paied to the substantive significance of involved quantities, working always with parameters having a well defined interpretation (even though not always in the real world as counterfactuals), and easily communicability to a non-technical audience.

At least I would like to mention a personal experience had when for the first time I made an estimation of $A1$ indicator, which I called "the thrill of the applied researcher": it is that feeling you get when a prediction is

subject to verifiability in the short: have I done something wrong? reality is out there, cold as a number, challenging you and your hard work, all your intuition, intelligence, assumptions, computations, experiences and pride. However I know that there is no way of learning unless doing, and sometimes lerning more by making mistakes then by going straight to the correct solution.

3.2 Data description

In order to make analysis both nationa and local students, degree courses and researchers dataset are enquired. It follows a detailed description.

3.2.1 Student dataset

Statistics about students are computed according to the Students National Registry[1] and the University of Palermo Student Registry. Available data cover academic years from 2006/2007 to 2010/2011 and contain:

- **ID**: Student Identifier, a unique number identifying students;

- **GE**: Gender of the student, male or female.

- **DCD**: Name of the Degree Course the student is enrolled on;

- **DCT**: Type of Degree Course: BA as Bachelor, MA as Master, MA-old as Master old classification, MALT as Long Term (4/5/6 years) Master;

- **FAC**: Faculty of the Degree Course the student is enrolled on;

- **G**: the group degree course belong to. All national Degree Courses are divided into 4 sets depending on the "standard cost per student". Every student is differently weighted according to the Degree Course group he is enrolled on: $A = 4, B = 3, C = 2, D = 1$;

---

[1]http://anagrafe.miur.it/index.php, The registry does not include students enrolled before 1999, so it quite underestimates actual student population.

- **NE**: New Enrolment, whether the student is enrolled for the first time on an Italian University;

- **YOE**: Year of Enrolment, first, second and so on up to 3rd out of course, other years are aggregated;

- **RS**: Regular Student, whether the student has been enrolled by a number of years less or equal to the legal Course duration;

- **AS**: Active Student, whether the student has earned at least 5 CFU[2] in the current C.Y.;

- **TY**: Type of the Student: active regular student (*RA*), inactive regular student (*RI*), out of course (*FC*). This variable, joint event of *RS* and *AS*, is explicitly required by one ministerial indicator;

- **C.Y.**: Calendar Year, from 2006 to 2010;

- **A.Y.**: Academic Year, from 2006/2007 to 2009/2010.

- **EFF**: numbers of CFU actually earned by the student in one C.Y.;

- **TEO**: max number of CFU to be earned by the student in one A.Y.;

3.2.2 Bachelor and Master Course dataset

Statistics on Bachelor and Master Courses have been computed according to National Regitry [3] and University of Palermo registry[4] containing:

- **DCD**: name of the Degree Course;

- **DCT**: type of Degree Course: BA as Bachelor, Master, Master old classification, Long Lerm Master;

---

[2]A university course credit, CFU; is a unit that gives weighting to the value, level or time requirements of an academic course

[3] http://offf.cineca.it/pubblico.php/ricerca/show_form/p/cercauniv

[4]http://offweb.unipa.it/

- **FAC**: the faculty Degree Courses belong to;

- **G**: the group degree courses belong to;

- **TEA**: number of teachers, teaching "base"and "characteristic"courses;

- **NE**: number of new enrolments in one *A.Y.*;

- **T**: a Threshold, namely a teoric number of new enrolments n one *A.Y.*, fixed by the Ministry;

- **A.Y.**: Academic Year, from 2006/2007 to 2009/2010.

3.2.3 Researcher dataset

Statistics about researchers have been computed according to National Interest Research Project (PRIN)[5] and the University of Palermo researcher Registry [6] containing:

- **ID**: teacher identifier, a unique number identifying teachers;

- **GE**: Gender of the teacher, male or female.

- **AR**: one of the 14 areas the researcher belongs to;

- **FAC**: the Faculty the teacher belongs to;

- **DEP**: the Department the teacher belongs to;

- **PA**: whether the teacher participated to a project PRIN;

- **PE**: whether the teacher was positively evaluated in the project PRIN;

- **GR**: whether the teacher was granted in the project PRIN;

- **C.Y.**: Calendar Year, from 2006 to 2010. For PA/PE/GR from 2006 to 2009;

---

[5]http://prin.miur.it/
[6]http://surplus.unipa.it/

3.3 Indicator A1

Indicator A1 is concerned with educational offer and encompasses three different aspects:

1. regularity of studies ($TARS_x$): regularity of studies is measured by the number of active (AS) and regular (AR) students enrolled on Degree Courses belongings to group $x \in G$;

2. educational offer sustainability (KA): educational offer sustainability is measured by the ratio between the number of regular teachers who cover "base "and "characteristic"courses (TEA) and the total teoric number of Degree Courses (TDC). The latter is the sum, for all bachelors and master courses, of the ratio between new enrolments (NE_i) and a predetermined threshold depending on the Degree Course (T_i). When the ratio is less than 1, it is forced to 1. The former ratio is then divided by the national median value (MKA). KA reflects the paradigm: the more crucial courses are covered by internal rather than adjunct teachers, the more reputable the Degree Course. KA ranges around 1;

3. local context (KT): local context is measured by a variable which is in inverse proportion of the regional family income. Such variable is a priori bounded on ± 10% the national average, that is even though the lowest regional family income is 50% less the mean, its KT can't be more then 10% higher. KT ranges around 1.

The formula is the following:

$$A1 = 100 * \frac{(KA + KT)TARS}{S} \tag{3.1}$$

Where S is a (normalizing) constant, namely the sum for all 54 participant Universities of the product $I = (KA + KT)TARS$.

The indicator main contributor is $TARS$, a dimension-dependent quantity,

71

reflecting university size. *KA* mirrors the law on "minimum requirements of teachers" (art. 9, comma 2, D.M. 22 oct 2004, n. 270) optimizing the number of internal teachers and number of students. *KT* is not a quality parameter at all. The national university evaluation agency blog considers such indicator as outlandish and incomprehensible [7].

3.3.1 A1 Structural equations and induced DAG

The above formula comes at the end of a set of structural equations as defined by the national model:

$$
A1 \begin{cases}
TARS = 4ARS_A + 3ARS_B + 2ARS_C + ARS_D \\
KT = f(\text{regional family income}) \\
TDC = \sum_i max\{\frac{NE_i}{T_i}, 1\} \qquad i = 1, 2, ..., \#(DC) \\
I_{KA} = \frac{TEA}{TDC} \\
I_{KA}^* = \{I_{(1,KA)}, ..., I_{(53,KA)}\} \\
MR_X = Median(\{I_{KA}, I_{KA}^*\}) \\
KA = \frac{I_{KA}}{MR_X} \\
I = (KA + KT)TARS \\
S = \sum_{i=1}^{54} I_i = I + \sum_{i=1}^{53} I_i = I + S^* \\
A1 = 100\frac{I}{S}
\end{cases}
$$

*A*1 can be easily verified as a deterministic function of a bunch of observed variables (ARS_X, NE_i, TEA, regional family income) and, to show the relationship,it is linear on ARS_X, *KT* or *TEA*.

There are some other relationships not explicitly mentioned in the definition system that may help to show how the indicator works. For example, ARS_A and ARS_B are essentially the same variable measured in different degree courses, therefore, it appears quite probable to assume they could have a common (hidden) cause U_{ARS} (the same with ARS_C and ARS_D).

---

[7]http://cronaca.anvur.it/2011/10/ordinare-il-fondo-per-il-finanziamento.html

There must be a probabilistic effect of NE on ARS because new enrolments are the base of active new enrolments which in turn are a part of ARS.

Finally, there could be a correlation between S^* and I or between I_{KA} and I_{KA}^* due to national policies common effect.

Due to these considerations, other equations have to be added to the above system:

$$
\begin{cases}
ARS_X = f(NE_X, \varepsilon) \\
Corr(ARS_X, ARS_Y) \neq 0 \quad \forall X \neq Y \\
Corr(S^*, I) \neq 0 \\
Corr(I_{KA}^*, I_{KA}) \neq 0
\end{cases}
$$

Next Figure shows induced DAG (nodes labelled starting with U represent unobserved variables, since our interest is focused on the University of Palermo, KT is treated as a constant and depicted as a square.). The d-separation rule guides in the recognition of the independences embodied in the DAG, e.g. $I_{KA} \not\perp\!\!\!\perp NE$, $I_{KA} \perp\!\!\!\perp NE|TDC$ or $NE \perp\!\!\!\perp A1$, $NE \not\perp\!\!\!\perp A1|TARS, TDC$.
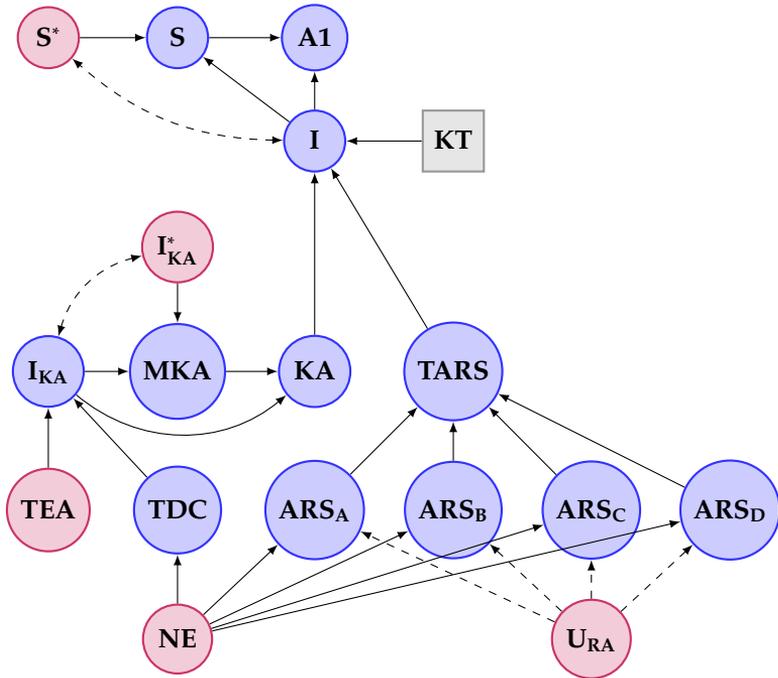
Figure 3.1: *A*1 induced DAG

In making analysis, it needs to know at any given time, which variable values are available, and which need to be estimated. For each variable involved in the indicator computation, Table 3.1 shows the reference time, namely the A.Y. or the C.Y. which quantities are referred to, and the available time, namely the date starting from, data are either public or accessible for computation. National quantities are labeled with the same year as $A1$.

Year	T		T+1		T+2	
Quantity	RT	AT	RT	AT	RT	AT
NE	2009/10	01/01/10	2010/11	01/01/11	2011/12	**01/01/12**
TEA	2009/10	01/01/10	2010/11	01/01/11	2011/12	**01/01/12**
ARS (RS)	2008/09	01/01/09	2009/10	01/01/10	2010/11	01/01/11
ARS (AS)	2009	01/01/10	2010	01/01/11	2011	**01/01/12**
MKA	2010	12/10	2011	**12/11**	2012	**12/12**
S	2010	12/10	2011	**12/11**	2012	**12/12**
A1	2010	12/10	2011	**12/11**	2012	**12/12**

Table 3.1: $A1$ Reference Time & Available Time

At present[8], all quantities, $A1$ included, are available for 2010. To make a prediction on $A1$ 2011, only national quantities, *MKA* and *S*, need to be predicted because all local quantities, $\{NE, TEA, ARS\}$ are readily accessible, whereas to make a prediction on $A1$ 2012 all variable values need to be forecast, one year ahead for local data and two years ahead for national data.

---

[8]draft copy of this thesis was submitted on 3/12/2011

3.3.2 A1 2010 results

In 2010, the University of Palermo achieved the following results:

$$A1\ 2010 \begin{cases} TARS = 4*1494 + 3*9844 + 2*9960 + 6223 = 61651 \\ KT = 1.09 \\ TDC = 86.406 \\ KA = \frac{\frac{1858}{86.406}}{17.37} = 1.24 \\ I = (1.24 + 1.09)61651 = 143.64 \\ S = 3920434 \\ A1 = 100*143.64/3920434 = 3.66 \end{cases}$$

$A1$=3.66 is the eighth among Italian universities[9]. It is equivalent to 4479,84M€. The $A1$ main factor $TARS$ is equal to 61,651, the eighth. Corrector KT is the highest and KA is the third, having the eighth highest numerator and the eleventh lowest denominator. 3.66 corresponds to 4.48M€. The correlation coefficient between $A1$ and the total number of students for all Universities is equal to 0.975, showing the great dependence of $A1$ from the size of the universities, as noticed by the National Agency for Evaluation of Research and University System[10].However individual differences between amount of FFO shared according to $A1$ or according to the number of students can be not negligible.

Although $A1$ fails its purpose, e.g. measuring efficiency, decreasing out-of-course students is the most beneficial way for improving the indicator value. Note that an increase of 0.1 is equivalent to an increase of 122.4M€.

---

[9]http://attiministeriali.miur.it/media/161927/i_assegnazione.pdf

[10] http://cronaca.anvur.it/2010/11/si-apprezza-una-buona-istruzione.html

3.4 A1 Retrospective analysis

Counterfactuals describe hypothetical phenomena in which events happen differently than in reality. With this in mind, counterfactuals are particularly appropriate for describing how things could have happened, e.g. what value *A*1 could attain, had *TARS*, taken a different value. Evaluating differences between reality and hypothetical gives a hints on the action to undertake for the future.

Following the method exposed in **??**, the following retrospective question sare asked:

Q1 what *A*1 would have been had *A*1 2010 computed today?

A1 *A*1 would be 3.65. This counterfactual (*A*1 2010 was actually computed one year ago) means there wasn't significant delay in data entry operation[11], 12,24M€. Otherwise changes in the databases would lead to changes in the indicator value;

Q2 what *A*1 would be had *KT* been equal to 1?

A2 *A*1 would be 3.51. The difference, 0.15, between the observed value and the counterfactual value is interpreted as the increment due to local context, 183,6M€. In this case, the counterfactual nature of computation is evident: had Sicily the same regional income of "Piemonte";

Q3 what would be *A*1 had *KA* equal to the national median value, *MKA*?

A3 *A*1 would be 3.30 The difference, 0.36, is interpreted as the increment due to the University of Palermo educational offer sustainability that is 440.64M€;

Q4 What A1 would be had TARS increased by 1%?

A4 *A*1 would be 4.02, this means every increasing of 1% in TARS would lead an increasing of 0.36 in *A*1, that is a gain of 440.64M€.

---

[11]assuming no delay in other Universities too

3.5 A1 Previsional analysis

Next task is to make a prediction on *A*1, both for one and two years ahead.

3.5.1 Short term prevision

For *A*1 2011 prevision, the only quantities need to be forecast are *MKA* and *S*. This operation is called Short Term Prevision.

Currently, there are 3 *MKA* observations, from A.Y. 2007/2008 to 2009/2010 and only one *S* observation, from *FFO* 2010 allocation.

Since local quantities are already known, it is possible to estimate *A*1 under the hypothesis that *S* and *MKA* attain the same value as the previous year. Table 3.2 shows one year back differences:

*A*1	ARS_A	ARS_B	ARS_C	ARS_D	*TEA*	*TDC*
2010	1494	9844	9960	6223	1858	86.41
2011	1448	8721	9322	6097	1667	80.96
Δ%	-3.08	-11.41	-6.41	-10.28	-11.46	-6.73

Table 3.2: *A*1 Local quantities variation

TARS decreased by 8.20%, and I_{KA} by 4.23%. With these values *A*1 2011 would be equal to 3.31, with a lost of 428,4M€.

One year forecasting: the *S* case

S is the sum of the local indicator *I* for all 54 Italian Universities taking section in the funding. 2011 being the second year application, only one observation is available, $S_{2010} = 3920434$. A simple model is built, assuming the normal approximation: S_t is treated as a function of t[12], centered

---

[12]*MKA* is fixed at his A.Y. 2009/2010 value, 17.37

on the last observation and with an ad-hoc standard deviation, namely the empirical standard deviation of *I*.In order to achieve a minimal sensitivity analysis, mean and standard deviation are multiplied by coefficients, $k_1 = k_2 = \{0.95, 1, 1.05\}$.

Figure 3.13 shows graphically how a simple AR(1) model on S_t is able to hide complex relation between its factors.
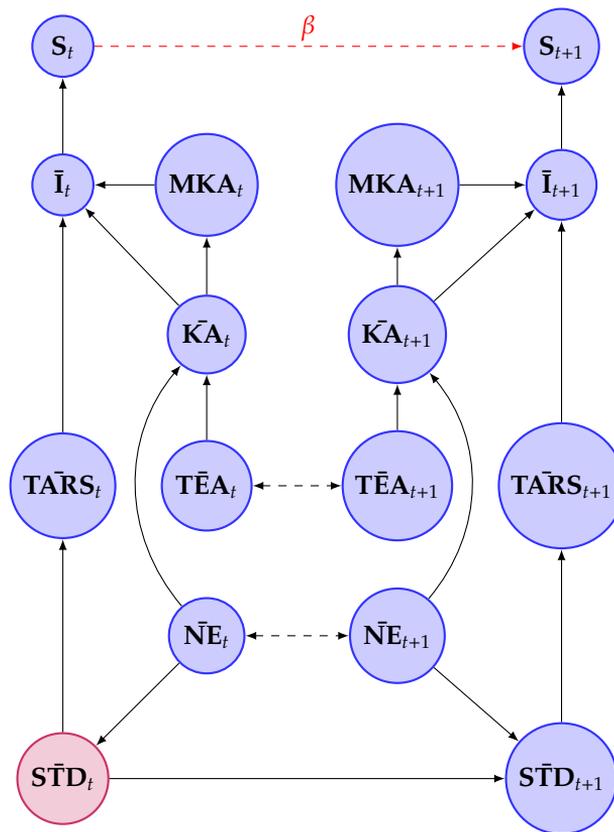
Figure 3.2: The spurious association between S_t and S_{t+1}

The model is:

$$S \begin{cases} S_{2011} \sim N(\hat{\mu}, \hat{\sigma}) \\ \hat{\mu} = k_1 S_{2010}^{obs} \\ \hat{\sigma} = k_2 \sigma(I_{2010}^{obs}) \\ \{k_1, k_2\} \in \{0.95, 1, 1.05\} \end{cases}$$

Where S_{2010}^{obs} is the S observed value in 2010, and $\sigma(I_{2010}^{obs})$ is the empirical standard deviation among all 54 I values in 2010. k_1, and k_2 are two coefficients useful for sensitivity analysis. Next table shows simulation values from the model above.

Simulation table

n.iter=1Ml	S 2011				
Node	Mean	Sd	5%	50%	95%
0.95μ, 0.95sd	3731763	58391	3636599	3731849	3826197
0.95μ, sd	3731276	60623	3632379	37312582	3830629
μ, 0.95sd	3920851	57566	3725660	3921192	4016155
μ, sd	3920434	61075	3724412	3920429	4116456
1.05μ, sd	4108714	61221	4009218	4108003	4210087
μ, 1.05sd	3920300	63709	3815518	3920833	4124392
1.05μ, 1.05sd	4109576	64663	4003871	4110382	4214928

Table 3.3: Simulation table for S_{A1} 2011

This simulation offers a grid of values inside which S should hopefully be. Summaries helps on deciding whether a value, e.g. 40000000, belongs to N(1.05μ , $\hat{\sigma}$) but not to N(0.95μ , $\hat{\sigma}$).

One year forecasting: the *MKA* case

The second quantity to estimate in order to predict $A1$ for 2011 is *MKA*, namely the national median of all *KA* correctors. Figure 3.3 shows the

causal diagram between two consecutive observations:
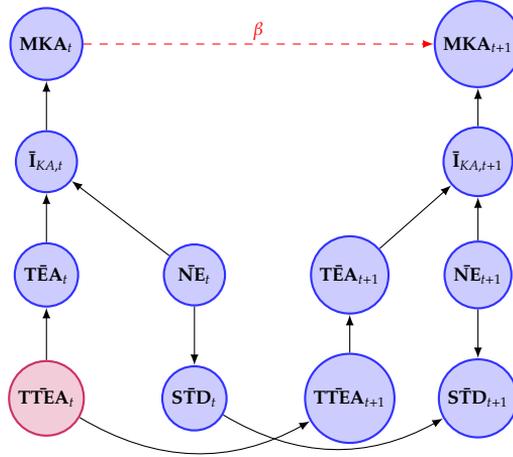
Figure 3.3: Spurious association between MKA_t and MKA_{t+1}

Table 3.4 shows three year time series on *MKA*, a linear positive trend is evident.

A.Y.	*MKA*
2007/08	13.55
2008/09	15.08
2009/10	17.37

Table 3.4: *MKA* 3 years time series

In order to make a prediction on *MKA* both AR(1) and TVAR(1) with non informative prior (NIP) and informative prior (IP) are used.

The model is:

81

$$M^{AR1} = \begin{cases} MKA_t \sim N(\mu_t, \tau), \ t = 1, 2, 3 \\ \mu_t = \beta MKA_{t-1} \\ NIP : \beta \sim N(0, .001), \ \tau \sim \Gamma(0.001, 0.001), \\ MKA_0 \sim N(0, 0.001) \\ IP : \beta \sim N(6.84, 1/1.43^2) \ \tau \sim \Gamma(21.91, 22.03), \\ MKA_0 \sim N(15.33, 0.4057) \end{cases}$$

$E[\beta_{IP}]$ and $\tau[\beta_{IP}]$ come from a linear regression (without intercept) coefficient and standard error of MKA_t on $t = (1, 2, 3)$, $E[MKA_0]$ and $\tau[MKA_0]$ are respectively empirical mean and precision. Rate and shape of the Gamma distribution are computed making use of the method of moments which allows for equating parameters as a function of empirical mean and variance. $E[\Gamma]$ is estimated taking the mean over three years of the mean of the 54 KA values, $V[\Gamma]$ is estimated taking the mean over three years of the variance of the 54 KA values. The followings are tables for the model deviance and, in a typical winbugs style, summaries from the posterior distribution of quantities of interest. One million samples were drawn.

AR(1)	NIP	IP
MD	7.53	7.73
P	122	0.98
PD	129.53	8.71

As expected, non informative model deviance is much higher, mainly for the penalty factor.

n.iter 1M	MKA - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011}$	19.20	34.76	17.70	19.70	21.66
$\hat{s}$	3.223	56.145	0.200	0.568	5.481
$\hat{\beta}$	1.106	2.587	1.019	1.134	1.247
	MKA - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011}$	19.831	0.874	18.401	19.826	21.277
$\hat{s}$	1.010	0.110	0.850	1.000	1.207
$\hat{\beta}$	1.142	0.050	1.059	1.141	1.225

Table 3.5: $MKA_{2010/11}$ AR(1) NIP/IP model

It is important to notice the $\hat{\mu}_{2011}$ mean is approximately the same with both priors (3.3 absolute percentage difference) after only three year observations, meaning such short observation time is sufficient to agree upon the *MKA* estimation point (usually the mean) starting from complete divergent prior states of knowledge. Therefore vague prior can be legitimate used.

It stands out to the reader the huge difference around estimators standard deviation and asymmetry in the distribution of $\hat{s}$ in the non informative case.

The second fitted model is TVAR(1) with a random walk on β_t. Again, both non informative and informative priors are used. Formally:

$$M^{TVAR1} = \begin{cases} MKA_t \sim N(\mu_{MKA_t}, \tau_{MKA_t}), \ t = 1, 2, 3 \\ \mu_{MKA_t} = \beta_t MKA_{t-1} \\ \beta_t \sim N(\beta_{t-1}, \tau_{\beta_t}) \\ NIP : (\beta_0, MKA_0) \sim N(0, .001), \ (\tau_{MKA_t}, \tau_{\beta_t}) \sim \Gamma(0.001, 0.001) \\ IP : \beta_0 \sim N(6.84, 1/1.43^2), \ MKA_0 \sim N(15.33, 0.4057) \\ \tau_{MKA_t} \sim \Gamma(21.91, 22.03), \ \tau_{\beta_t} \sim \Gamma(2.42, 2.2) \end{cases}$$

In the informative case, $E[\Gamma]$ is estimated taking the AR(1) point estimation, and $V[\Gamma]$ is estimated as $0.2 * E[\Gamma]$.

83

AR(1)	NIP	IP
MD	6.72	7.67
P	4618	1.03
PD	4624.72	8.70

Table 3.6: MKA_2011: TVAR(1) NIP/IP Deviance

n.iter 1Ml	MKA - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011}$	19.831	0.874	18.401	19.826	21.277
$\hat{s}_y$	2.597	15.831	0.040	0.346	9.805
$\hat{\beta}[1]$	1.036	4.224	0.420	1.119	1.725
$\hat{\beta}[2]$	1.038	1.729	0.962	1.114	1.276
$\hat{\beta}[3]$	1.065	1.741	0.985	1.151	1.278
$\hat{\beta}[4]$	1.015	4.182	0.420	1.143	1.752
$\hat{s}_\beta$	0.563	6.240	0.030	0.100	1.349
	MKA - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011}$	19.95	30.39	-28.02	20.02	68.10
$\hat{s}_y$	1.020	0.113	0.854	1.010	1.220
$\hat{\beta}[1]$	4.023	1.475	1.832	3.888	6.654
$\hat{\beta}[2]$	1.120	0.775	0.996	1.120	1.246
$\hat{\beta}[3]$	1.151	0.770	1.041	1.152	1.260
$\hat{\beta}[4]$	1.149	1.623	-1.613	1.153	3.921
$\hat{s}_\beta$	1.600	0.754	0.752	1.433	3.034

Table 3.7: TVAR(1) NIP/IP on MKA 2011

Results mime the previous model.

One year forecasting: *A*1 **2011 prevision**

*A*1 2011 relies on already available local quantities:

$$ARS=4*1448+3*8721+2*9322+6097=56696$$
$$KT=1.09$$
$$TDC=86.406$$
$$KA=\frac{1858/86.406}{17.37}=1.24$$

From the previous section, the estimations for *S* and *MKA* are (5° and 95° percentile in squared brackets):

$$\hat{S}=3920434 \;\; [3724412, 4116456]$$
$$\hat{MKA} = 19.83 \; [18.40, 21.27]$$

These values leads to an estimation for *A*1 2011 of:

$$A1_{2011} = 3.27 \; [3.01, 3.51]$$

corresponding to a reduction of 477,36M€[795.6M€, 183.6M€].

3.5.2 Long term prevision

Long term prevision corresponds to the second year prediction of national quantities, S_{2012} and MKA_{2012}, and to the first year prediction of local quantities, $TEA_{2010/11}$, $NE_{2010/11}$ and $ARS_{X,2009/10}$.

Two year forecasting: the S and MKA 2012 cases

Since only one year observation is available for S, the same normal approximation as for 2011 estimation will be used with the same parameters. The reason is that, having no clue about what is going on, whether S will increase or decrease, leaving prediction unchanged is preferred.

	Mean	p05	p95
S_{2012}	3920434	3724412	4116456

Table 3.8: S 2012 estimation and C.I.

For the MKA_{2011} second year estimation, the same AR(1) and TVAR(1) models with respect to MKA_{2010} will be used.
Let us start with the AR(1) model.
The following are deviance and summaries tables for AR(1) model:

AR(1)	NIP	IP
MD	6.71	7.68
P	5708	1.03
PD	5714.71	8.71

Table 3.9: MKA 2012: AR(1) NIP/IP Deviance

There is a huge model penalty term for non informative prior model.

n.iter 1M	MKA - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2012}$	22.35	5130.39	17.90	22.35	28.26
$\hat{s}$	2.691	31.657	0.201	0.584	6.227
$\hat{\beta}$	1.134	1.416	1.012	1.134	1.256
	MKA - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2012}$	19.83	0.87	18.40	19.82	21.26
$\hat{s}$	1.010	0.109	0.848	1.000	1.204
$\hat{\beta}$	1.141	0.050	1.059	1.141	1.224

Table 3.10: *MKA* 2012: AR(1) NIP/IP prevision

Results are quite different. The non informative model shows abnormal values for $\hat{\mu}_{2012}$ and its standard deviation as well as the standard deviation and the credible interval of $\hat{s}$. Whereas in the informative case, estimated quantities are the same as 2011 ($\hat{MKA} = 19.83$ [18.40, 21.27]).

Table 3.11 and 3.12 show results for TVAR(1) model:

AR(1)	NIP	IP
MD	3.24	8.56
P	542026	1.54
PD	542029.24	10.10

Table 3.11: *MKA* 2012: TVAR(1) NIP/IP Deviance

n.iter 1M	MKA - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2012}$	75,61	483,9	3.562	22.91	79.52
$\hat{s}_{mka}$	4.818	64.228	0.040	0.351	10.497
$\hat{\beta}_{2012}$	1.171	29.690	0.182	1.147	2.116
$\hat{s}_{\beta}$	0.891	22.843	0.030	0.104	1.654
	MKA - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2012}$	76.52	177.84	-18.20	27.86	319.68
$\hat{s}_{mka}$	1.020	0.112	0.855	1.010	1.218
$\hat{\beta}_{2012}$	1.153	2.479	-2.751	1.154	5.052
$\hat{s}_{\beta}$	1.590	0.737	0.759	1.425	2.970

Table 3.12: $MKA_{2011/12}$: TVAR(1) NIP/IP prevision

An huge NIP penalty factor deletes every comment on the results. Also with informative priors elicitation the model leads to poor estimations, standard deviations are too high and credible intervals too large (sometimes inadmissible like -18.20). Due to its robustness, the median is preferable to the mean.

One year forecasting: the *TEA* case

The first local quantity is going to be estimated is the number of teachers who taught "base"and "characteristics"courses during A.Y. 2011/2012[13], *TEA*, on the basis of the the total number of teachers, *TTEA*. The following table and graph show time series on both *TEA* and *TTEA*.

---

[13]actually this quantity is no longer a random variable because its valued has been determinate, but it is not yet in our possession and needs to be estimated.

A.Y.	TEA_t	$TTEA_t$	C.Y.
2006/07	1967	2194	2006
2007/08	1909	2140	2007
2008/09	1979	2195	2008
2009/10	1859	2112	2009
2010/11	1667	1988	2010
2011/12	-	1674	2011

Table 3.13: *TEA*, *TTEA* data
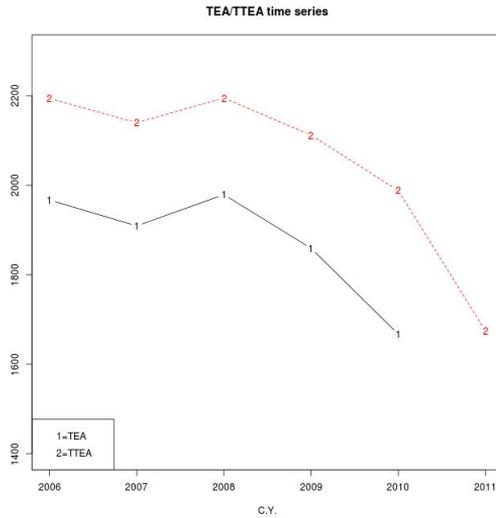
Clearly *TEA* follows the same pattern of *TTEA*, but the last year *TTEA* observation had a more rapid decreasing. The model takes care of that, conditioning on *TTEA* using a binomial model (with *TEA* as number of successes on *TTEA* trials) instead of a normal model.

The model is:

$$M^{BIN} = \begin{cases} TEA_t \sim Bin(\theta, TTEA_t), \ t = 1, 2, 3, 4, 5, 6 \\ TTEA_t = f(TTEA_{t-1}) \\ NIP : \theta \sim B(1, 1) \\ IP : \theta \sim B(141.1, 18.92) \end{cases}$$

Since the estimation regards $TEA_{2011/12}$ conditioned to the known value $TTEA_{2011}$, the second equation doesn't need to be explicated. Informative prior parameters come from empirical mean success rate, 0.881.
The induced DAG is the following:
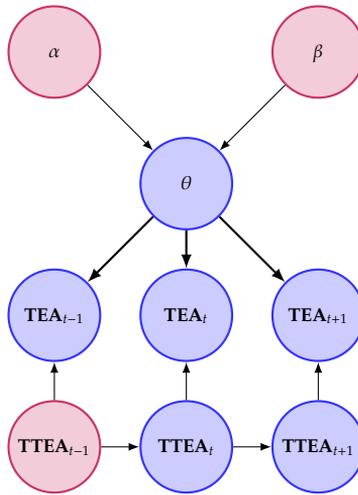
Figure 3.4: $TEA_{2010/11}$: binomial model

Both parametrizations return the same values, meaning non informative prior is legitimated, so they are displayed once.
Deviance table:

AR(1)	NIP/IP
MD	85.3
P	1.01
PD	86.3

Table 3.14: $TEA_{2011/12}$: Binomial model NIP/IP Deviance

n.iter 1M	TEA - NIP/IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011/12}$	1478.44	14.18	1454.00	1477.00	1500.00
$\hat{\theta}$	0.883	0.003	0.877	0.883	0.888

Table 3.15: $TEA_{2011/12}$: binomial model NIP/IP prevision

Allowing time varying θ leads to the following model:

$$
M^{TVAR(1)} = \begin{cases}
TEA_t \sim Bin(\theta_t, TTEA_t), \ t = 1, 2, 3, 4, 5 \\
TTEA_t = f(TTEA_{t-1}) \\
\text{logit}(\theta_t) = b_t \\
b_t \sim N(\mu_t, \tau) \\
\mu_t = \beta \mu_{t-1} \\
NIP : \{\beta, \mu_1\} \sim N(0, 1.0e - 6), \ \tau \sim \Gamma(0.001, 0.001)
\end{cases}
$$

Only non informative prior is used. The induced DAG is:
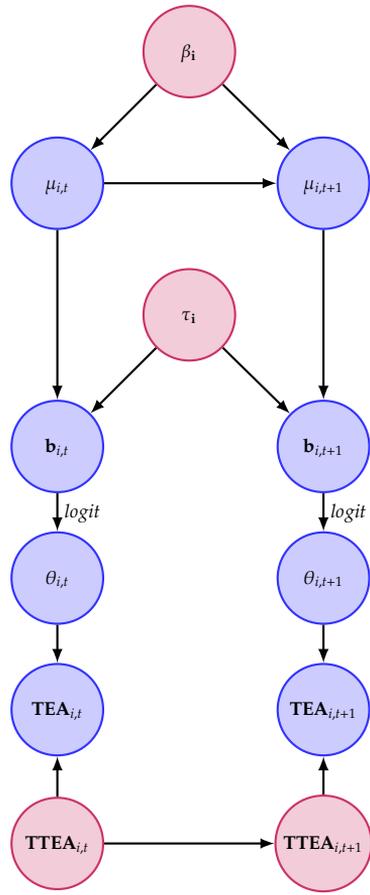
Figure 3.5: Induced DAG from TEA TVAR(1) model

Results:

AR(1)	NIP
MD	41.5
P	4.82
PD	46.32

Table 3.16: $TEA_{2009/10}$: TVAR(1) NIP/IP Deviance

n.iter 1M	TEA - NIP				
Node	Mean	Sd	5%	50%	95%
TEA	1407.36	89.11	1288.00	1416.00	1509.00
θ_1	0.898	0.006	0.888	0.898	0.908
θ_2	0.892	0.006	0.882	0.893	0.902
θ_3	0.898	0.007	0.887	0.898	0.909
θ_4	0.878	0.007	0.867	0.879	0.890
θ_5	0.842	0.008	0.828	0.842	0.855
θ_6	0.841	0.053	0.771	0.847	0.900
$\hat{\beta}$	0.943	0.071	0.891	0.946	0.999
s	0.206	0.196	0.070	0.163	0.460

Table 3.17: $TEA_{2009/10}$: TVAR NIP/IP prevision

The estimate value is lower than the fixed effect model, because of the θ_t negative trend.

One year forecasting: the *TDC* case

The second quantity to estimate is *TDC*, the total number of degree courses. It depends upon new enrolments and on a threshold ($TDC = \sum \frac{NE_i}{T_i}$). Data from *A.Y.* 2006/2007 to 2009/2010 come from the national registry of stu-

dents, whereas *A.Y.* 2010/11 comes from University of Palermo registry. The time series below shows a clear negative serial trend with a marked jump in 2009/10.

A.Y.	TDC
2007/08	134
2008/09	124
2009/10	86.41
2010/11	80.96

Table 3.18: *TDC* data

On average, during the last 4 years, *TDC* has been decreasing by 15% per year.
An AR(1) model with a normal approximation for TDC_t is used.

The model is as follows:

$$M^{AR1} = \begin{cases} TDC_t \sim N(\mu_t, \tau), \ t = 1, 2, 3, 4 \\ \mu_t = \beta TDC_{t-1} \\ NIP : \beta \sim N(0, .001), \ \tau \sim \Gamma(0.001, 0.001), \ TDC_0 \sim N(0, 0.001) \\ IP : \beta \sim N(-19.6, 0.0136) \ \tau \sim \Gamma(28.47, 0.2677), \ TDC_0 \sim N(140, 0.1537) \end{cases}$$

parameter values for the informative prior arise assuming a 95% range based on a linear model estimation (frequentist estimation), $E(TDC_0)$ is at a glance a values on the trend, $\tau(TDC_0)$ parameters arises from method of moments.

The true causal DAG shows a further relationship between enrolments and (teen) population (the first is a fraction of the second):
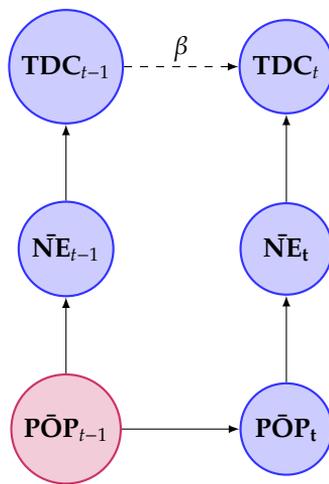
Figure 3.6: *TDC*: AR(1) model

Results:

	NIP	IP
MD	35.9	77
P	17.1	1.52
PD	53.0	78.52

Table 3.19: $TDC_{2011/12}$ AR(1) NIP/IP Deviance

n.iter 1M	TDC - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011/12}$	72.70	20.22	52.76	72.70	92.69
$\hat{s}$	27.924	38.863	9.278	19.250	70.123
$\hat{\beta}$	0.841	0.234	0.611	0.84	1.073
	TDC - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\mu}_{2011/12}$	72.988	1.284	70.886	72.984	75.104
$\hat{s}$	3.020	0.284	2.592	2.998	3.521
$\hat{\beta}$	0.845	0.015	0.820	0.845	0.869

Table 3.20: $TDC_{2011/12}$: AR(1) NIP/IP prevision

The two models estimate the same mean and slope but (as expected) different variability. The decreasing estimate depends on both new enrolments decreasing (as highlighted in 3.29) and on a continuing educational offer reformulation.

One year forecasting: the *TARS* case

The last quantity needed to be estimated is *TARS*, namely the sum, for each degree course group, of active and regular students (*ARS*).

TARS is the crucial variable for *A*1 determination. To stress the importance, two different estimating methods are used: the first employs the normal distribution, the second the binomial distribution.

Table 3.21 shows a four-year time series on *ARS* for each group. The next plot depicts an evident ongoing decreasing trend in all groups. During the last 4 years, group *A* decreased by 9.5%, *B* by 9.4%, *C* by 6.7%, *D* by 2.0%.

A.Y.	ARS_A	ARS_B	ARS_C	ARS_D
2006/07	1598	10070	10049	6651
2007/08	1576	10261	10460	6660
2008/09	1532	9981	10162	6342
2009/10	1448	8721	9322	6097

Table 3.21: *ARS* 4 year time series
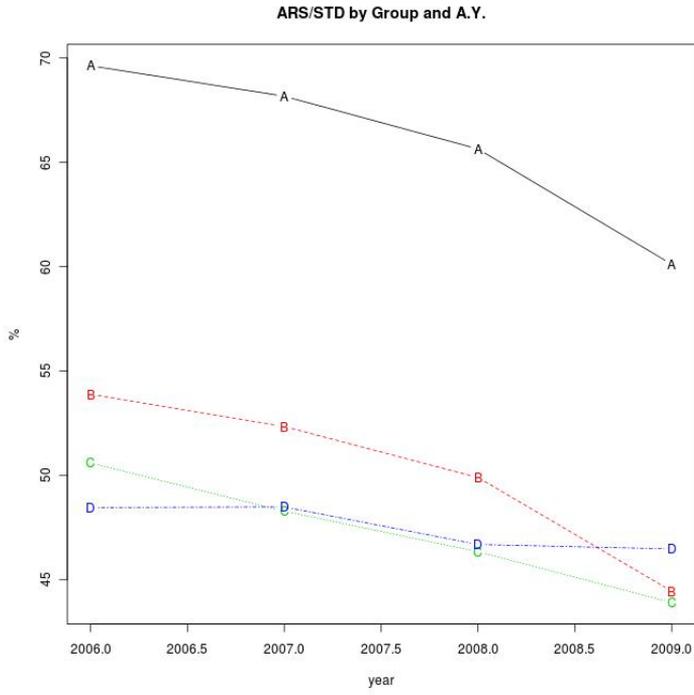
Figure 3.7: *ARS* time series for {A,B,C,D} groups

The first model is:

$$
M^{AR1} = \begin{cases}
ARS_{x,t} \sim N(\mu_{x,t}, \tau_x), \ t = 1,2,3,4; \ x = \{A,B,C,D\} \\
\mu_{x,t} = \beta_x ARS_{x,t-1} \\
NIP: \beta_x, ARS_{x,0} \sim N(0,.001), \ \tau_x \sim \Gamma(0.001,0.001) \\
IP: \beta_A \sim N(-49.4, 1/((9.89^2)*4)) \ \beta_B \sim N(-432.7, 1/((232^2)*4)) \\
\beta_C \sim N(-247.9, 1/((198^2)*4)) \ \beta_D \sim N(-198, 1/((49.1^2)*4)) \\
\tau_A \sim \Gamma(957.9, 0.6226), \ \tau_B \sim \Gamma(334.2, 0.03527), \\
\tau_C \sim \Gamma(761.7, 0.07619), \ \tau_D \sim \Gamma(1004, 0.156), \\
ARS_{A,0} \sim N(1538, 1/(49.71^2)), \ ARS_{B,0} \sim N(9758, 1/(526^2)), \\
ARS_{C,0} \sim N(9998, 1/(362^2)), \ ARS_{D,0} \sim N(6438.5, 1/(203.2^2))
\end{cases}
$$

Informative priors on β_X and $ARS_{X,0}$ come from frequentist linear model estimations, τ_x from the method of moments.

The causal diagram the following:
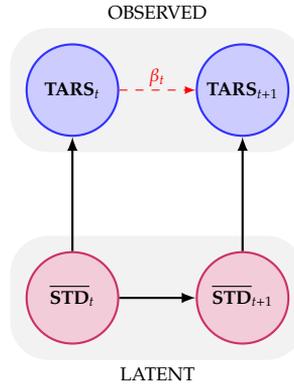
Figure 3.8: Spurious association between $TARS_t$ and $TARS_{t+1}$

	NIP	IP
MD	833	6229
P	4	4
PD	837	6223

Table 3.22: TARS: AR(1) NIP/IP Deviance

β estimates, and obviously means, are similar for both models, whereas as expected informative prior bring to lower variability, sometimes too low as for β's standard deviation.

n.iter 1M	ARS - NIP				
Node	Mean	Sd	5%	50%	95%
$A\ \hat{\mu}_{2012}$	1401.0	491.6	633.9	1401.3	2167.2
$B\ \hat{\mu}_{2012}$	8329	2908	3786	8330	12863
$C\ \hat{\mu}_{2012}$	9089	3063	4306	9091	13865
$D\ \hat{\mu}_{2012}$	5920	2063	2692	5921	9141
$A : \hat{s}$	850.9	358.5	480.8	766.7	1495.6
$B : \hat{s}$	5385	2271	3045	4853	9456
$C : \hat{s}$	5366	2263	3032	4837	9421
$D : \hat{s}$	3541	1494	2002	3192	6220
$A : \hat{\beta}$	0.968	0.339	0.438	0.968	1.497
$B : \hat{\beta}$	0.955	0.334	0.434	0.955	1.475
$C : \hat{\beta}$	0.975	0.329	0.462	0.975	1.487
$D : \hat{\beta}$	0.971	0.338	0.441	0.971	1.487
	ARS - IP				
Node	Mean	Sd	5%	50%	95%
$A\ \hat{\mu}_{2012}$	1402.274	0.559	1401.354	1402.274	1403.194
$B\ \hat{\mu}_{2012}$	8335.91	20.26	8302.59	8335.91	8369.25
$C\ \hat{\mu}_{2012}$	9098.42	11.94	9078.80	9098.41	9118.06
$D\ \hat{\mu}_{2012}$	5926.043	2.935	5921.215	5926.043	5930.870
$A : \hat{s}$	1.050	0.017	1.023	1.050	1.079
$B : \hat{s}$	40.648	1.111	38.865	40.624	42.519
$C : \hat{s}$	22.666	0.411	22.001	22.660	23.352
$D : \hat{s}$	5.464	0.086	5.324	5.462	5.607
$A : \hat{\beta}$	0.968	0.000	0.968	0.968	0.969
$B : \hat{\beta}$	0.956	0.002	0.952	0.956	0.960
$C : \hat{\beta}$	0.976	0.001	0.974	0.976	0.978
$D : \hat{\beta}$	0.972	0.000	0.971	0.972	0.978

Table 3.23: ARS: AR(1) NIP/IP prevision

Next model allows for time-varying β parameters. The model is:

$$M^{TVAR1} = \begin{cases} ARS_{x,t} \sim N(\mu_{x,t}, \tau_x), \ t = 1, 2, 3, 4; \ \ x = \{A, B, C, D\} \\ \mu_{x,t} = \beta_{x,t} ARS_{x,t-1} \\ \beta_{x,t} \sim N(\beta_{x,t-1}, \tau_\beta) \\ NIP : \{\beta_{x,t}, ARS_{x,0}\} \sim N(0, .001), \ \{\tau_x, \tau_\beta\} \sim \Gamma(0.001, 0.001) \end{cases}$$

Only non informative priors are used. Again β standard deviations are too low.

	NIP
MD	308
P	10.7
PD	318.7

Table 3.24: TARS: TVAR(1) NIP deviance

	ARS - NIP				
Node	Mean	Sd	5%	50%	95%
$A\ \hat{\mu}_{2012}$	1371.3	1394.8	-766.1	1370.1	3509.8
$B\ \hat{\mu}_{2012}$	7625	8396	-5278	7621	20533
$C\ \hat{\mu}_{2012}$	8543	8978	-5247	8547	22316
$D\ \hat{\mu}_{2012}$	5876	5862	-3113	5866	14894
$A : s_{\hat{ARS}}$	1.767	2.974	0.578	1.201	4.402
$B : s_{\hat{ARS}}$	1.768	3.078	0.578	1.201	4.408
$C : s_{\hat{ARS}}$	1.768	2.874	0.578	1.203	4.421
$D : s_{\hat{ARS}}$	1.753	2.681	0.577	1.199	4.390
$A : \hat{\beta}_1$	0.912	0.879	-0.514	0.929	2.278
$A : \hat{\beta}_2$	0.986	0.002	0.984	0.986	0.988
$A : \hat{\beta}_3$	0.972	0.002	0.970	0.972	0.974
$A : \hat{\beta}_4$	0.945	0.002	0.943	0.945	0.947
$A : \hat{\beta}_5$	0.947	0.963	-0.529	0.946	2.424
$B : \hat{\beta}_1$	0.941	0.879	-0.488	0.959	2.307
$B : \hat{\beta}_2$	1.019	0.000	1.019	1.019	1.019
$B : \hat{\beta}_3$	0.973	0.000	0.972	0.973	0.973
$B : \hat{\beta}_4$	0.874	0.000	0.873	0.874	0.874
$B : \hat{\beta}_5$	0.874	0.963	-0.605	0.874	2.354
$C : \hat{\beta}_1$	0.960	0.880	-0.467	0.978	2.328
$C : \hat{\beta}_2$	1.041	0.000	1.041	1.041	1.041
$C : \hat{\beta}_3$	0.972	0.000	0.971	0.972	0.972
$C : \hat{\beta}_4$	0.917	0.000	0.917	0.917	0.918
$C : \hat{\beta}_5$	0.916	0.963	-0.563	0.917	2.394
$D : \hat{\beta}_1$	0.925	0.877	-0.498	0.942	2.289
$D : \hat{\beta}_2$	1.001	0.000	1.001	1.001	1.002
$D : \hat{\beta}_3$	0.952	0.000	0.952	0.952	0.953
$D : \hat{\beta}_4$	0.961	0.000	0.961	0.961	0.962
$D : \hat{\beta}_5$	0.964	0.961	-0.511	0.962	2.443
$A : \hat{s}_{\beta}$	0.867	0.418	0.457	0.764	1.613
$B : \hat{s}_{\beta}$	0.869	0.41703	0.458	0.766	1.616
$C : \hat{s}_{\beta}$	0.868	0.418	0.458	0.765	1.615
$D : \hat{s}_{\beta}$	0.867	0.416	0.458	0.764	1.612

Table 3.25: Stats TVAR(1) NIP on ARS

One year forecasting: the *TARS* case again

The second estimation method uses the following equations:

$$
M^{Bin} = \begin{cases}
TARS = 4ARS_A + 3ARS_B + 2ARS_C + ARS_D \\
ARS_{x,t} \sim Bin(\theta_x, STD_{x,t}), \ t = 1,2,3,4; \ \ x = \{A,B,C,D\} \\
STD_{x,t} = f(STD_{x,t-1}) \\
NIP : \theta_x \sim B(1,1) \forall x \\
IP : \theta_A \sim B(111.8, 57.9), \ \theta_B \sim B(96.7, 97.22), \\
\theta_C \sim B(193.6, 215.9), \ \theta_D \sim B(1316, 1453)
\end{cases}
$$

ARS is interpreted as depending on two sources: one is quantitative, the total number of students *STD*, and the other qualitative, θ, featuring all possible conditions helping or preventing acquiring the status (e.g. student ability, course difficulty, ecc). Since *STD* is treated as a parameter for *TARS*, the model doesn't make use of the third equation. θ informative prior parameters come from the method of moments on empirical ratios. Both parametrization return the same results, so they are displayed once.

	NIP/IP
MD	833
P	3.99
PD	836.99

Table 3.26: TARS: AR(1) NIP/IP Deviance

n.iter 1M	ARS - NIP/IP				
Node	Mean	Sd	5%	50%	95%
$A : \hat{ARS}_A$	1707.21	27.33	1662.00	1707.00	1752.00
$B : \hat{ARS}_B$	9162.34	75.14	9039.00	9162.00	9286.00
$C : \hat{ARS}_C$	8397.23	73.21	8277.00	8397.00	8518.00
$D : \hat{ARS}_D$	5779.95	60.97	5680.00	5780.00	5880.00
$A\ \hat{\theta}_A$	0.658	0.005	0.650	0.658	0.666
$B\ \hat{\theta}_B$	0.501	0.002	0.498	0.501	0.504
$C\ \hat{\theta}_C$	0.472	0.002	0.469	0.472	0.475
$D\ \hat{\theta}_D$	0.475	0.002	0.472	0.475	0.479

Table 3.27: Stats NIP/IP on ARS Binomial model

Two years forecasting: A1 **2012 prevision**

Now everything is ready for the long term prediction. Every involved quantity is shown in the next table:

Var	Mean	p05	p95	Model	2011	2010
S	3920434	3724412	4116456	$N(1\mu,1\sigma)$	3920434	3920434
MKA	22.35	17.9	28.26	AR(1) - NIP	19.83	17.37
TEA	1478	1454	1500	Bin NIP/IP	1667	1858
TDC	72.98	70.88	75.10	AR(1) IP	80.96	86.41
ARS_A	1402	1401	1403	Bin IP	1448	1494
ARS_B	8335	8302	8369	Bin IP	8721	9844
ARS_C	9098	9078	9118	Bin IP	9322	9960
ARS_D	5926	5921	5930	Bin IP	6097	6223
$A1$	2.96	3.37	2.57		3.27	3.66

Table 3.28: A1 2012 estimation and C.I.

A1 2010 observed value was 3.66, the estimation for 2011 is 3.22 and

for 2012 is 2.96[2.57, 3.37]. This discouraging result comes from an overall decreasing of $A1$ factors and calls for new policies.

3.6 A1 Causal analysis

Causal analysis deals with prediction under intervention. What this task is useful for? It is crucial for policy evaluation. For example an interesting question could be: what is going to be the effect on *A*1 of taking an intervention such that *TARS* is going to increase to, say, 3%? Notice this is different from the question: what is going to be the effect on *A*1 of predicting an *ARS* to increase by 3%. This is because observations are compatible with diagnostic reasoning: *TARS* increased because the number of students did, and likely, new enrolments too. But if new enrolments increase, *KA* will decrease. Whereas actions are not compatible with diagnostic reasoning: *TARS* increased because students earned more CFU, the number of students remained the same, *KA* too. Only the first scenario involves *NE*.

Another example demonstrates that models useful for making predictions are not always acceptable for causal purposes. It can be safely assumed, taking an action on $TTEA_t$ has effect on $TTEA_{t+1}$ because of the identity $TTEA_{t+1} = TTEA_t + IN_{t+1} - OUT_{t+1}$, being $TTEA_t$ on the right hand side means it is a direct causal factor of $TTEA_{t+1}$, and iteratively $TTEA_{t-k(>0)}$ affects $TTEA_{t+1}$ only through $TTEA_t$, that is $TTEA_{t-k}$ is an indirect causal factor of $TTEA_{t+1}$. Due to a linearisation operation [14], that model can be arranged as an *AR*(1), that is, such model can be used both for prediction and for intervention. But, this is not always the case: let us suppose to predict NE_{t+1} on the basis of NE_t. New enrolments at time *t* are not part of new enrolments at time *t* + 1 so, the identity $NE_{t+1} = NE_t + IN_{t+1} - OUT_{t+1}$ is no longer valid, and a *AR*(1) model cannot be interpreted as a causal model [15] but it can as prevision model). This entails, even though NE_t could be a good predictor of NE_{t+1}, it cannot be a causal factor of NE_{t+1}: applying a

---

[14] $TTEA_{t+1} = TTEA_t + IN_{t+1} - OUT_{t+1}$, $IN_{t+1} = k_t TTEA_t + r_t$, $OUT_{t+1} = k'_t TTEA_t + r'_t$ leads to $TTEA_{t+1} = TTEA_t + (k_t TTEA_t + r_t) - (k'_t TTEA_t + r'_t) = \beta_t TTEA_t + \varepsilon_t$

[15] such as $NE_{t+1} = kPOP_{t+1}$ ($k \in [0,1]$, $POP_{t+1} = POP_t + IN_{t+1} - OUT_{t+1}$, where *POP* is the total population in a range of ages (e.g 18-29) and IN_{t+1}, OUT_{t+1} are people entering and leaving that range.

policy on NE_t given no hints about what NE_{t+1} is going to be. Basically the correlation is spurious **??**.

3.6.1 Intervening on NE. Policy evaluation: Is it useful increasing new enrolments?

This section deals with causal effects and counterfactuals in order to evaluate an intervention on the system, namely a new enrolments increasing policy[16]. Before implementing it, next figure shows how the mechanism works:
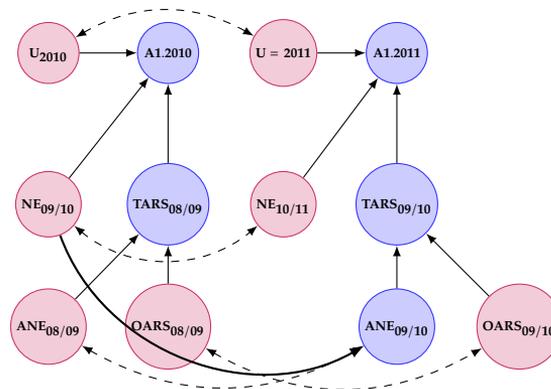
Figure 3.9: $A1$ Two years Causal DAG

The graph is built as follows: $A1$ 2010 depends on $ARS_{08/09}$, $NE_{09/10}$ (through KA) and on other variables jointly labelled U_{2010}. $ARS_{08/09}$ is the sum of active new enrolments[17], $ANE_{08/09}$, and other regular active students, $OARS_{08/09}$. The same structure holds for $A1$ 2011.

Every exogenous variable is connected by a bi-directed arc, with the next year same variable, e.g. $ANE_{08/09} \leftrightarrow ANE_{09/10}$, highlighting the spurious

---

[16]NE depends on university external factors, such as teenage population, fees or family income level.

[17]new enrolments are necessarily regular.

dependence. Note that $ANE_{09/10}$ depends on $NE_{09/10}$[18].

An intervention aiming to increase new enrolments has two opposite effects: one leads to an $A1$ decrease, by decreasing KA (NE is the denominator), the other leads to an $A1$ increase, because increasing NE increases in probability ANE which in turn increases ARS and finally $A1$. But how the following graph shows, these two effects have an impact on two different years.
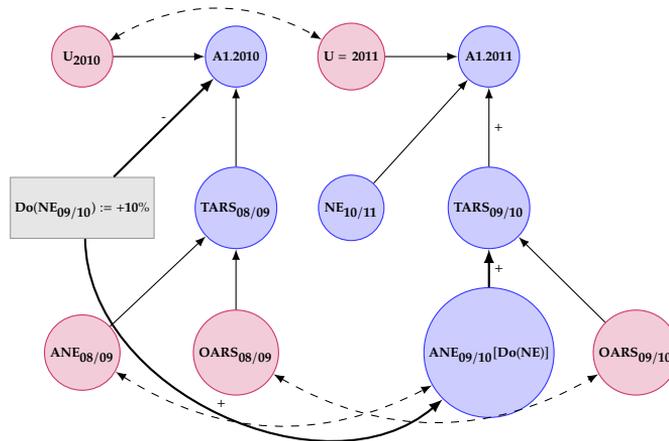
Figure 3.10: Causal effect of $NE_{09/10}$ on $\{A1_{2010}, A1_{2011}\}$

Clearly, it is impossible to evaluate the causal effect of $NE_{09/10}$ by only looking at the difference between $A1_{2010}$ and $A1_{2011}$, because they depend on other time varying quantities.

How do we solve this effect in order to make comparison? Counterfactuals are required. We will rephrase the question: what $A1$ 2010 would be had $NE_{09/10}$ increased by 10% and had $ANE_{08/09}$ equal to $ANE_{09/10}$? In this way, by virtually changing the indicator definition, turning back the effect of NE on ARS to the same year of NE on KA, the model is able to evaluate the

---

[18]probabilistically could be a binomial $ANE_{09/10} \sim BIN(\theta, NE_{09/10})$

double opposite effect jointly and answer the question.
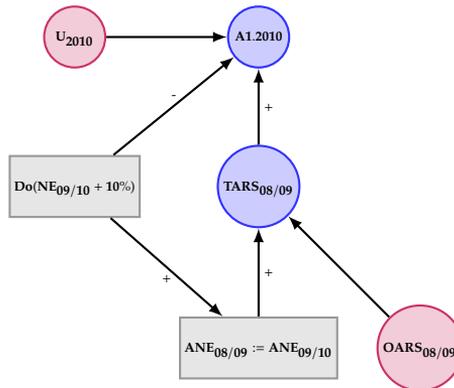
Figure 3.11: Counterfactual effect of $NE_{09/10}$ on $A1_{2010}$

In the light of these arguments, let's perform the computation:

- Observed $A1$ 2010=3.66;

- Direct effect[19] $A1$ 2010$[NE_{09/10} := +10\%] = 3.54$;

- Counterfactual effect $A1$ 2010$[NE_{09/10} := +10\%, ANE_{08/09} := ANE_{09/10}] = 3.69$.

The counterfactual value is higher than the observed one, meaning the increasing policy is advantageous (in the counterfactual world).

3.6.2   Searching weak TARS performances

Let us start describing the student population. The next table shows the number of students, new enrolments and graduates by academic year:

---

[19]Pearl (2001)

A.Y.	STD	NE	GR
2006/07	54577	10189	5763
2007/08	57324	10424	5827
2008/09	57858	9503	6440
2009/10	56390	8519	6973
2010/11	50830	7496	7076(on 28/11/11)

Table 3.29: Students by A.Y.

Students have been increasing from 2006 to 2008, followed by a slight decreasing in 2009 and after that, a strong decline in 2010. The latter could be the effect of the generalized introduction in 2009/2010 of a barrier to entry . New enrolments have been decreasing from 2007 to 2010 by a rate of approximately 10%. Graduates have been increasing every year. Student population is decreasing due to the jointly effect of decreasing *NE* and increasing *GR*. All indicator factors depending on it (e.g. *TARS*, *A*2 *TEO* and *EFF*) have to take into account of this.

From now on, we are going to consider only A.Y. 2009/2010 student population, because it is the last data available. Group *A* is the smallest with 4% students, followed by *D* with 23%, *B* and *C* are quite similar with 35% and 38%.

G	A	B	C	D
%STD	4.3	34.8	37.7	23.3

Table 3.30: Students by group - A.Y. 2009/2010

40% are male, with a large difference for group *B* and *C*.

111

Ge \ Gr	A	B	C	D	T
F	54.0	44.1	72.0	64.9	59.4
M	46.0	56.0	28.0	35.1	40.6
T	4.3	34.8	37.7	23.3	100

Table 3.31: Students by gender for each group - A.Y. 2009/2010

Faculty distribution between groups is more dispersed. Group A has students belonging only to Medicine. Group B is split between 7 Faculties, the most represented being Engineering, Science and Architecture. Group C is split between 5 Faculties, the most represented being Arts and Humanities, Educational Science and Economics. Group C is split between 5 Faculties, the most represented being Law and Arts and Humanities.
It turns out that the group completely depends on the Faculty, but not vice versa.

Fac	Prevalent Group	%STD ∈ PG
Agriculture	B	100
Architecture	B	100
Arts and Humanities	C	71.7
Economics	C	90.2
Educational Science	C	79.1
Engineering	B	100
Law	D	100
Medicine	A	60.1
Pharmacy	B	100
Political Science	D	73.5
Science	B	99.5
Sport Science	C	100

Table 3.32: most represented groups by Faculty, A.Y. 2009/2010

Table above reads 71.7% of Arts and Humanities students belong to group *C*.

Group	Prevalent Fac	%STD∈PF
A	MED	100
B	ENG-SCI-ARC	76.7
C	AH - ES - EC	89.7
D	LAW - AH - ES	86.5

Table 3.33: most represented Faculties by group A.Y. 2009/2010

Table above reads 76.7% group *B* students belong to group ENG-SCI-ARC Faculties.

Let us get started with regular and active students analysis. *ARS* unveils the great mass of out-of-course students as a typical Italian phenomenon.

RS \ G	A	B	C	D	T
No	27.6	43.4	42.9	39.6	41.6
Yes	72.4	56.6	57.1	60.4	58.4
T	4.3	34.8	37.6	23.3	100

Table 3.34: Regular Students by Group

58.4% of students are regulars, irrespective of gender. Percentage increases to 72.4% in group *A*.

AS \ G	A	B	C	D	T
No	20.0	22.7	25.7	25.0	24.2
Yes	80.0	77.3	74.3	75.0	75.28
T	4.3	34.8	37.6	23.3	100

Table 3.35: Active Students by Group

75.8% of students are active, 78.12% female, 80% for group A.

ARS \ S	F	M	T
OC	41.56	41.72	41.63
AR	47.23	42.61	45.37
IR	11.21	15.66	13.00
T	59.4	40.6	100

41.6% of students are active and regular, both male and female, 60.1% for group A.

ARS \ G	A	B	C	D	T
OC	27.6	43.4	42.9	39.6	41.63
RA	60.1	44.4	43.9	46.5	45.37
RI	12.2	12.2	13.2	14.0	13.00
T	4.3	34.8	37.7	23.3	100

Table 3.36: Active and Regular students by group

It can't be performed an intervention on actual out-of-course students, because it is an ultimate status. They are, probably, students living in campus, using university service, paying fees but no more fund increaser. It

can be intervened upon Regular students preventing them to become out-of-course.

But the main point is regular inactive students, 13% of students (22.17% of only regular students) taking no exams for the entire year. It's a transversal phenomena taking place uniformly in each group. Here it is the variable need to be intervened on.

An increasing in the number of active and regular students must have a combination of two sources:

- one quantitative, namely an increasing of the numbers of students (STD);

- the other qualitative, namely their capacity of acquiring more CFUs θ_{ARS}[20].

First source explains ARS increasing as a diagnostic evidence of students increasing[21], hence it cannot be considered an intervention on ARS (at most on STD).

Second source explains ARS increasing as a diagnostic evidence of a higher student capacity in acquiring CFU, θ_{ARS}, and that doesn't have any effect on the number of students, leading to a pure causal effect on $A1$.

---

[20]there is a controversy about wheter student quality can be represented by a quantitative feature as passing more and more exams because it can derive by less strict policy, that is allowing low quality student to get more exams.

[21]and in particular of new enrolments so leading a back-door (Pearl, 2000) effect on KA. Therefore ARS increasing has a causal direct effect on $A1$ and at the same time a spurious effect, through KA denominator, on $A1$.

3.7 Indicator A2

Indicator $A2$ aims to measure student abilities to earn the credits required each year by the Degree Course. This capacity is measured by the ratio (R) between the actual CFU earned (EFF) and the teoric of CFU to be earned (TEO) in one academic year. The latter is typically 60 per year, the former a part of it. Similar to $A1$, the ratio is computed separately for the 4 degree course groups. The credit acquisition different capabilities between groups is taken into account by dividing such ratio by the national median value of each group ($Median(R)$ or MR). This results in a coefficient ranging around 1. Finally, this coefficient multiplies the number of actual credits again, turning the indicator in a number that is a proxy for university dimension[22]. In other words, $A2$ is an indicator based on the number of CFU earned multiplied by a coefficient which depends on the greater or lesser propensity to acquire such credits compared to a theoretical maximum (TEO) and to an external reference value (the national median). Empirically, based on 2010 data, it was found that coefficient had a minimum of about 0.65 and a maximum of 1.50 and, about 50% of the coefficients range in (0.87, 1.12).

3.7.1 A2 Structural equations and induced DAG

The formalization of $A2$, as defined by the Ministy of University, is shown by the following system:

---

[22]The Italian evaluation agency blog defines it as rough and productivist. http://cronaca.anvur.it/2010/09/produttivita-e-produttivismo.html

$$A2 \begin{cases} R_X = 100\frac{EFF_X}{TEO_X}, \quad X = \{A, B, C, D\} \\ R_X^* = \{R_{1,X}, ..., R_{53,X}\} \\ MR_X = Median(\{R_X, R_X^*\}) \\ WCFU_X = \frac{R_X}{MR_X}EFF_X \\ I = WCFU_A + WCFU_B + WCFU_C + WCFU_D \\ S = \sum_{i=1}^{54} I_i = I + \sum_{i=1}^{53} I = I + S^* \\ A2 = 100\frac{I}{S} \end{cases}$$

Notice that the indicator actually doesn't depend directly on TEO because $WCFU_X = \frac{R_X}{MR_X}TEO_X = \frac{1}{MR_X}\frac{EFF_X}{TEO_X}TEO_X = \frac{EFF_X}{MR_X}$. Moreover, TEO and EFF are both function of the number of students (STD): $TEO = 60STD$ and $EFF = \mu STD$, where μ represents average credits earned by students par year. Usually μ is computed a posteriori by the ratio $\frac{EFF}{STD}$, so it turns out the latter equation can not be used for prevision purpose and it will be replaced with $EFF \sim Bin(\theta, TEO)$ where $\theta \in [0, 1]$ is representing (the unobservable) student abilities to earn credits or, in general, every condition may help (or prevent) that acquisition. Notice that R is the estimate of θ[23].

As usual, local and national variables, $\{S^*, I\}$ and $\{R^*, R\}$, could be correlated due to common national factors. These remarks add other equations to the system:

$$\begin{cases} TEO_X = 60STD_X \\ EFF_X \sim Bin(\theta_X, TEO_X) \\ Corr(S^*, I) \neq 0 \\ Corr(R_X^*, R_x) \neq 0 \end{cases}$$

The induced DAG is represented in Figure 3.12:

---

[23]what is the difference between R and θ? θ is perceived as the quality of students, as a "cause" of a high o low credits achievement. This quantity is actually an abstraction, not directly observable, and, maybe, not exactly measurable. R is its palpable expression, a rough measure, the "effect" of a high or low quality. This dichotomy between R and θ allows to explain contradictory cases such as "good students" with "poor results".
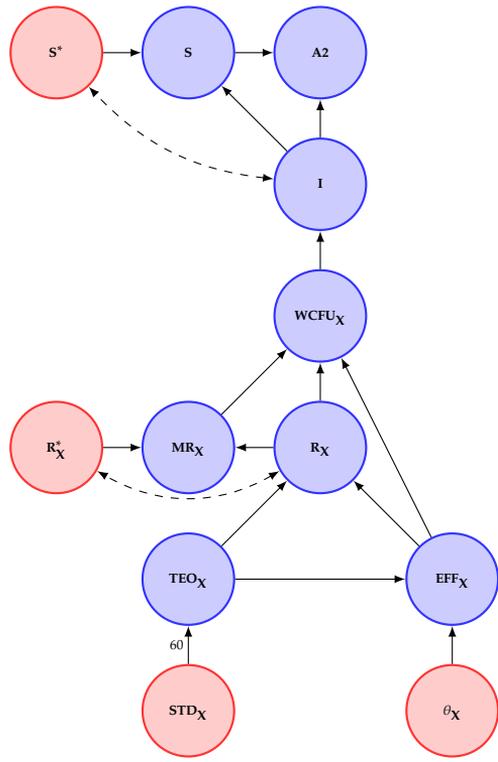
Figure 3.12: *A*2 induced DAG

In order to use adequate statistical techniques, Table 3.37 distinguishes quantities which are already known and which have to be estimated at a fixed time (in bold). National quantities are labelled with the same year as *A2*:

Year	T		T+1		T+2	
Quantity	RT	AT	RT	AT	RT	AT
EFF	2009	01/01/10	2010	01/01/11	2011	**01/01/12**
TEO	2008/09	01/01/09	2009/10	01/01/10	2010/11	01/01/11
MR	2010	12/10	2011	**12/11**	2012	**12/12**
S	2010	12/10	2011	**12/11**	2012	**12/12**
A2	2010	12/10	2011	**12/11**	2012	**12/12**

Table 3.37: A2 Reference Time and Available Time

It turns out that for *A2* 2010 estimate, all quantities are known, for 2011 local quantities, *EFF* and *TEO* are known, while for 2012 all quantities except *TEO* have to be estimated. Let us start with 2010 results.

3.7.2 A2 2010 results

The following quantities provided *A2* 2010 results:

$$A2_{2010} \begin{cases} \text{EFF=\{63767,450706,428157,269522\}} \\ \text{TEO=\{105300,1110160,1231877,776298\}} \\ \text{R=\{60.56, 40.60, 34.76, 34.72\}} \\ \text{MR=\{59.7, 43.5,38.6,37.1\}} \\ \text{I=1122225.29} \\ \text{S=40656680.44} \\ \text{A2=2.76} \end{cases}$$

University of Palermo indicator value, 2.76, is the tenth among all universities. Compared with *A1* value, 3.66, *A2* shows its weakness. It is

119

a quite insufficient result, as it leaps to the eye a poor capacity in earning credits: except for group A, the other groups have a ratio between achieved and teoric credits less than the national median, on average less than 3%.

3.8 A2 retrospective analysis

This section uses counterfactuals in order to know what *A*2 2010 would be, had some of its components be different from observed.
The observed *A*2 2010 value is 2.76.

Q1 What *A*2 would have been had it computed today?

A1 *A*2 would be 2.95. This counterfactual aims to discover delay in data entry by enquiring the same database twice. The first refers to the indicator official computation time, the second to the present time. The difference between these two values is interpreted as the data entry (or errors) delay cost. In 2010 this was 232.56M€, a quite high price.

Q2 What A2 would be had the local ratio, R, at least equal to the national median value, Median(R)?

A2 *A*2 would be equal to 2.98, which means an allocation increasing of 269.28M€. This counterfactual shows efforts have to be made to get a pass-mark, by forcing equality only for those R values which are less the the median value, the others remain intact. Section 3.10.2 provides a policy to reach this goal.

Q3 What *A*2 would be had all teoric CFU acquired by each university?

A3 *A*2 would be equal to 3.20, which means an increasing of 538.56M€. This counterfactual translates into a definition of *A*2 equal to $\frac{TEO}{\sum_{i=1}^{54} TEO_i}$, that is transforming the indicator proportional to the number of enrolled, a merely university dimension indicator.

Q4 What A2 would be had EFF increased by 1%?

A4 *A*2 would be 2.82, this means every increasing of 1% in EFF would lead an increasing of 0.6 in *A*2, that is a gain of 73,44M€.

3.9   A2 previsional analysis

Next task is to make a prediction on *A*2. According to Table 3.37, this task is split in two subtasks: a short term prevision, namely one year forecast, for which only national quantities need to be estimated, and a long term prevision, namely two year forecast, for which both national and local data have to be estimated.

3.9.1   Short term prevision

In order to predict *A*2 2011 only national quantities, S and MR_X are needed, because local quantities, EFF_X and TEO_X are already available, as they refer to C.Y. 2010 and A.Y. 2009/2010.

Table 3.38 shows one year difference for *EFF*, *TEO* and *R*:

	A2 2010	A2 2011	Δ%
EFF_A	63767	59090	-7.33
TEO_A	105300	120838	14.76
R_A	60.56	48.07	-12.49
EFF_B	450706	411626	-8.67
TEO_B	1110160	1129270	1.72
R_B	40.60	36.45	-4.15
EFF_C	428157	418220	-2.32
TEO_C	1231877	1239947	0.66
R_C	34.76	33.73	-1.03
EFF_D	269522	268540	-0.36
TEO_D	776298	760860	-1.99
R_D	34.72	35.29	0.57

Table 3.38: $A2$ {EFF, TEO, R} one year difference

There was a marked worsening of *EFF* for all groups and, except for group D, also *R* decreased. The worst percentage reduction was for group A, albeit it represents only 4.3% of students. Group B got a slight reduction but it represents almost 35% of students so this reduction had a wide negative effect on *A2*. These variations would indicate a reduction for 2011 and, assuming invariance on national quantities, $S = \hat{S}_{2010}$ and $MR_X = M\hat{R}_{X,2010}$, they lead to a value of 2.50, corresponding to a temporary loss of 318,24M€.

Next task is providing one year forecasting for *S* and *Median(R)*.

One year forecasting: the *S* case

Figure 3.13 shows graphically how a simple AR(1) model on S_t is able to hide complex relation between its factors.
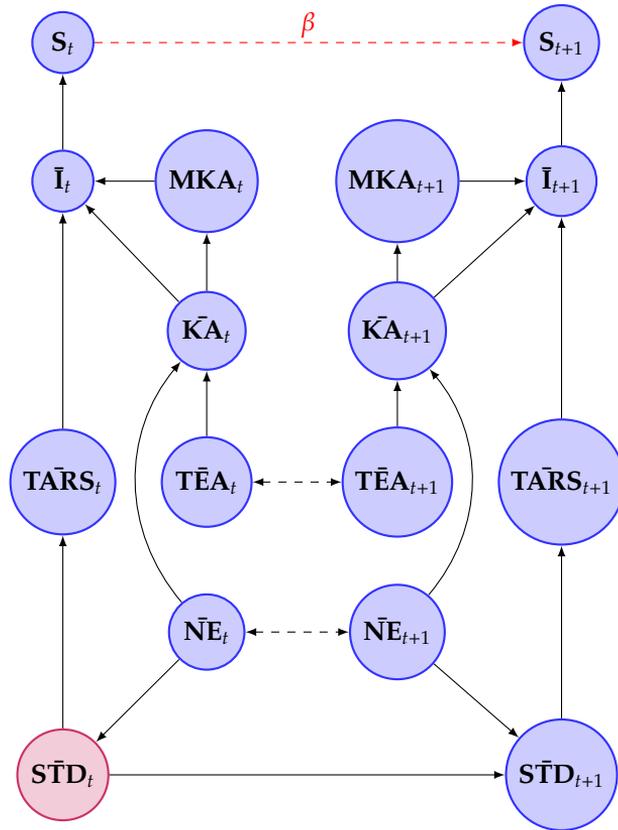
Figure 3.13: The spurious association between S_t and S_{t+1}

However having only one observation available, estimation is performed using the normal approximation.

The model is:

$$S \begin{cases} S_{2011} \sim N(\hat{\mu}, \hat{\sigma}) \\ \hat{\mu} = k_1 S_{2010}^{obs} \\ \hat{\sigma} = k_2 \sigma(I_{2010}^{obs}) \\ \{k_1, k_2\} \in \{0.95, 1, 1.05\} \end{cases}$$

Where S_{2010}^{obs} is the S observed value in 2010, and $\sigma(I_{2010}^{obs})$ is the empirical standard deviation among all 54 I values in 2010. k_1, and k_2 are two coefficients useful for sensitivity analysis.

Table 3.39 shows a simulation with 1 million samples.

n.iter=1Ml	S 2011					
	Min	25%	50%	Mean	75%	Max
$0.95\hat{\mu}, 0.95\hat{\sigma}$	35970434	38193587	38634265	38633906	39081025	40794750
$0.95\hat{\mu}, \hat{\sigma}$	36317998	38155047	38618797	38621280	39086423	41099934
$\hat{\mu}, 0.95\hat{\sigma}$	38265309	40216123	40653516	40657186	41093949	43015965
$\hat{\mu}, \hat{\sigma}$	37689367	40189597	40660970	40658456	41131302	43125385
$1.05\mu, \hat{\sigma}$	40194152	42224726	42698360	42692497	43157456	45412835
$\hat{\mu}, 1.05\hat{\sigma}$	38008765	40180217	40673394	40664743	41150794	43411202
$1.05\hat{\mu}, 1.05\hat{\sigma}$	39607079	42205980	42689397	42687213	43173425	45372949

Table 3.39: Simulation table For S_{A2} 2011

This simulation offers a grid of values inside which S should hopefully be. Summaries helps on deciding whether a value, e.g. 42000000, belongs to $(1.05\mu, \hat{\sigma})$ but not to $(0.95\mu, \hat{\sigma})$.

One year forecasting: the MR case

Figure 3.14 shows the causal DAG the between MR_t and MR_{t+1}
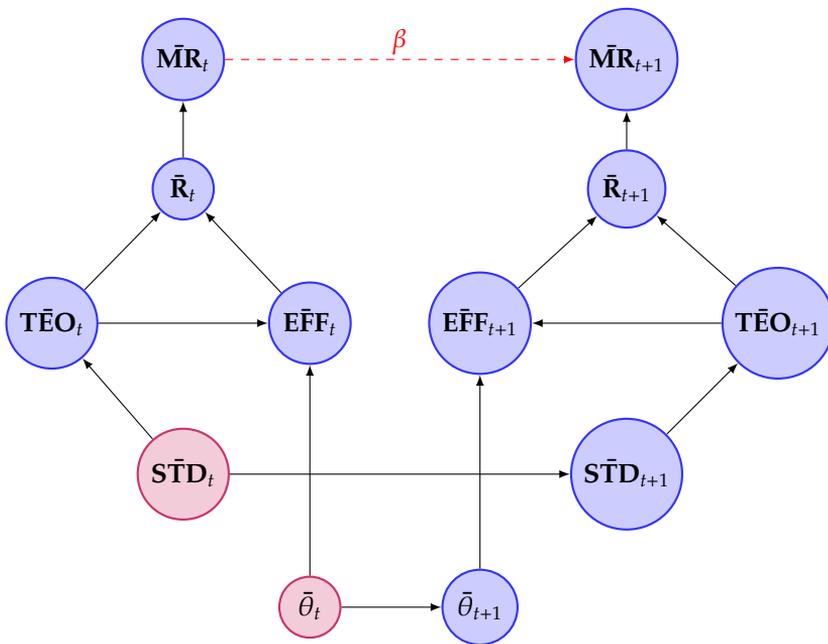
Figure 3.14: The spurious association between $\hat{M}R_t$ and $\hat{M}R_{t+1}$

Also in this case only one observation is available and then the normal approximation is used.

The model is:

$$MR \begin{cases} MR_X \sim N(\hat{\mu}_X, \hat{\sigma}_X) \quad X = \{A, B, C, D\} \\ \hat{\mu}_X = k_1 MR^{obs}_{X,2010} \\ \hat{\sigma}_X = k_2 \sigma(R^{obs}_{X,2010}) \\ \{k_1, k_2\} \in \{0.95, 1, 1.05\} \end{cases}$$

Where $MR^{obs}_{X,2010}$ is the MR_X 2010 observed value, and $\sigma(R^{obs}_{X,2010})$ is the 2010 empirical standard deviation among the 54 values of R.

As for the S case the next table provides a grid of values to choose from as MR estimation points.

Table 3.40 shows simulation results:

n.iter=1Ml	Median(R) Group A					
MR_{2010}	Min	25%	50%	Mean	75%	Max
$0.95\hat{\mu}, 0.95\hat{sd}$	45.84	54.64	56.67	56.67	58.63	67.38
$0.95\hat{\mu}, \hat{sd}$	45.93	54.68	56.67	56.73	58.73	67.99
$\hat{\mu}, 0.95\hat{sd}$	47.51	57.74	59.68	59.68	61.63	69.38
$\hat{\mu}$, sd	48.85	57.55	59.65	59.66	61.72	71.43
$1.05\hat{\mu}, \hat{sd}$	50.16	60.63	62.70	62.68	64.72	72.92
$\hat{\mu}, 1.05\hat{sd}$	48.72	57.58	59.71	59.73	61.83	72.95
$1.05\hat{\mu}, 1.05\hat{sd}$	51.31	60.55	62.72	62.69	64.82	75.28
n.iter=1Ml	Median(R) Group B					
MR_{2010}	Min	25%	50%	Mean	75%	Max
$0.95\hat{\mu}, 0.95\hat{sd}$	33.64	39.93	41.34	41.33	42.79	48.88
$0.95\hat{\mu}, \hat{sd}$	32.95	39.83	41.31	41.30	42.82	49.86
$\hat{\mu}, 0.95\hat{sd}$	34.99	42.05	43.48	43.50	44.93	50.97
$\hat{\mu}, \hat{sd}$	33.99	42.04	43.52	43.51	45.01	51.44
$1.05\hat{\mu}, \hat{sd}$	36.72	44.17	45.68	45.68	47.20	54.43
$\hat{\mu}, 1.05\hat{sd}$	34.34	41.95	43.47	43.50	45.06	51.66
$1.05\hat{\mu}, 1.05\hat{sd}$	36.35	44.10	45.64	45.66	47.21	55.33
n.iter=1Ml	Median(R) Group C					
MR_{2010}	Min	25%	50%	Mean	75%	Max
$0.95\hat{\mu}, 0.95\hat{sd}$	33.64	39.93	41.34	41.33	42.79	48.88
$0.95\hat{\mu}, \hat{sd}$	29.64	35.36	36.67	36.67	37.98	43.93
$\hat{\mu}, 0.95\hat{sd}$	31.13	37.32	38.60	38.60	39.87	45.81
$\hat{\mu}, \hat{sd}$	30.62	37.24	38.56	38.59	39.91	47.40
$1.05\hat{\mu}, \hat{sd}$	32.92	39.23	40.56	40.54	41.85	47.14
$\hat{\mu}, 1.05\hat{sd}$	30.54	37.19	38.60	38.58	39.94	48.18
$1.05\hat{\mu}, 1.05\hat{sd}$	33.25	39.16	40.54	40.55	41.94	48.31
n.iter=1Ml	Median(R) Group D					
MR_{2010}	Min	25%	50%	Mean	75%	Max
$0.95\hat{\mu}, 0.95\hat{sd}$	33.64	39.93	41.34	41.33	42.79	48.88
$0.95\hat{\mu}, \hat{sd}$	27.71	33.98	35.22	35.25	36.54	42.26
$\hat{\mu}, 0.95\hat{sd}$	30.03	35.87	37.08	37.08	38.27	43.04
$\hat{\mu}, \hat{sd}$	31.06	35.84	37.09	37.11	38.41	45.10
$1.05\hat{\mu}, \hat{sd}$	31.52	37.70	38.97	38.96	40.23	46.00
$\hat{\mu}, 1.05\hat{sd}$	28.63	35.77	37.10	37.10	38.41	45.06
$1.05\hat{\mu}, 1.05\hat{sd}$	31.47	37.63	38.97	38.97	40.32	46.31

Table 3.40: Simulation table For *Median*$(R)_{A2}$ 2011

One year forecasting: *A*2 **2011 prevision**

After these two simulation tasks, every quantity necessary to make a prediction on *A*2 2011 is known or estimated. The *S* estimation is chosen by the mean and the couple (Min, Max) of $N(\mu, \sigma)$, namely the forth row of Table 3.39:

$$S = 40658456[37689367, 43125385]$$

The same choice is made for *Median*(*R*) (see Table 3.40):

$$Median(R)_A = 59.66 \ [48.85, 71.43]$$
$$Median(R)_B = 43.51 \ [33.99, 51.44]$$
$$Median(R)_C = 38.59 \ [30.62, 47.40]$$
$$Median(R)_D = 37.11 \ [31.06, 45.10]$$

The known values are:

$$EFF = \{58090, 411626, 418220, 268540\}$$
$$TEO = \{120838, 1129270, 1239947, 760860\}[24]$$

The prevision and the credibility interval are:

$$A2_{2011} = 2.49 \ [1.94, 3.35]$$

In respect with *A*2 2010 there could be a lost of 330.48M€ [-1003.68M€,+722.16M€].

The credibility interval seems to be too large so also the couple (25%, 75%) of $N(\mu, \sigma)$ is chosen:

---

[24]not negligible differences was discovered between local and national registry in order to compute *EFF* and *TEO*. In this section national registry data are used because they are "official", in the prevision section a time series on local registry data will be used instead.

$$S=40658456 \ [37689367,43125385]$$
$$Median(R)_A=59.66 \ [57.55, \ 51.72]$$
$$Median(R)_B=43.51 \ [42.04, \ 55.01]$$
$$Median(R)_C=38.59 \ [37.24, \ 39.91]$$
$$Median(R)_D=37.11 \ [35.84, \ 38.41]$$

The alternative credibility interval is tighter:

$$A2_{2011}=2.49 \ [2.40, \ 2.61]$$

In respect with $A2$ 2010 there could be a lost of 330.48M€ [440,64M€,İ83,6M€].

It has to be noticed as an asymmetric interval for $A2$ comes from symmetric intervals for both $Median(R)$ and S.

3.9.2 Long-term prevision

Long-term prevision refers to predicting A2 for 2012. This aim is achieved estimating both national quantities, S and $Median(R)$, and the local quantity EFF, while TEO is already known (see Table 3.37).
Having only one past observation on $\{S, Median(R)\}$ the same results from the previous section are used for the estimation. Whereas in estimating EFF two different models are used: one is an AR(1) with normal approximation, since it contains only temporal information about EFF. The other model is a binomial model which adds information driven by TEO (as EFF parameter). There are 4 past observations available of EFF and 5 of TEO as next table shows.

Group	A		B		C		D	
C.Y.	EFF	TEO	EFF	TEO	EFF	TEO	EFF	TEO
2007	83812	137700	445258	1121425	467251	1191768	286846	823675
2008	85460	138720	464396	1176342	446935	1300068	284244	824275
2009	80925	140040	485369	1200322	453291	1315793	285205	815055
2010	68755	144538	447691	1177122	435225	1273776	278938	787166
2011	-	155536	-	1096819	-	1066124	-	728777

Table 3.41: *EFF* and *TEO* by Year and group

One year forecasting: the *EFF* Normal model

As well as for modelling $A1$ factors, the prediction is based on temporal variation of the quantity to be estimated. Both non informative and informative priors are used. The model is:

$$
M_{EFF}^{AR1} = \begin{cases}
EFF_{t,g} \sim N(\mu_{t,g}, \tau_g), \ t = 1, 2, 3, 4, \ g = \{A, B, C, D\} \\
\mu_{t,g} = \beta_g EFF_{t-1,g} \\
NIP : \beta_g \sim N(0, .001), \ \tau_g \sim \Gamma(0.001, 0.001), \ EFF_{0,g} \sim N(0, 0.001), \ \forall g \\
IP : \ EFF_{0,1} \sim N(79738, 1.313e - 08), \ EFF_{0,2} \sim N((460679, 2.185e - 09) \\
EFF_{0,3} \sim N(450675, 4.211e - 09), \ EFF_{0,4} \sim N(283808, 6.413e - 08) \\
\beta_{0,1} \sim N(39.70, 1/1.89), \ \beta_{0,2} \sim N(229.4, 1/4.6) \\
\beta_{0,3} \sim N(224.38, 1/3.38), \ \beta_{0,4} \sim N(141.303, 1/0.891) \\
\tau_1 \sim \Gamma(148.4, 0.001861), \ \tau_2 \sim \Gamma(824.3, 0.001789) \\
\tau_3 \sim \Gamma(1521, 0.003374), \ \tau_4 \sim \Gamma(9182, 0.03235)
\end{cases}
$$

Informative priors come from: EFF_0 is the average of 4 year observed values, τ from the method of moments, and β from a linear regression of *EFF* on time.

Results:

	NIP	IP
MD	431	23580
P	10.7	8.1
PD	441.7	23588.1

Table 3.42: $EFF_{2011/12}$ AR(1) NIP/IP deviance

n.iter 1M	EFF - NIP				
Node	Mean	Sd	5%	50%	95%
A: $\hat{\mu}_{2011/12}$	64432	23152	28022	64521	100826
B: $\hat{\mu}_{2011/12}$	445855	142515	221784	447763	668658
C: $\hat{\mu}_{2011/12}$	425268	147864	195582	424612	656994
D: $\hat{\mu}_{2011/12}$	275362	93479	128503	276138	422055
A: $\hat{s}$	45038	18942	25267	40552	78903
B: $\hat{s}$	237815	98872	134371	215189	414788
C: $\hat{s}$	248782	105400	140187	224539	438113
D: $\hat{s}$	152757	66987	86687	137142	268332
A: $\hat{\beta}$	0.937	0.337	0.408	0.938	1.466
B: $\hat{\beta}$	0.996	0.318	0.495	1.000	1.494
C: $\hat{\beta}$	0.977	0.340	0.449	0.976	1.510
D: $\hat{\beta}$	0.987	0.335	0.461	0.990	1.513
	EFF - IP				
Node	Mean	Sd	5%	50%	95%
A: $\hat{\mu}_{2011/12}$	64696	274	64246	64696	65147
B: $\hat{\mu}_{2011/12}$	447673.2	645.1	446611.9	447673.2	448733.3
C: $\hat{\mu}_{2011/12}$	424866.4	206.3	424526.9	424866.5	425205.7
D: $\hat{\mu}_{2011/12}$	276358.13	21.25	276323.15	276358.13	276393.10
A: $\hat{s}$	575.36	23.61	537.96	574.53	615.51
B: $\hat{s}$	1161.76	20.24	1128.96	1161.47	1195.51
C: $\hat{s}$	374.3	4.8	366.5	374.3	382.3
D: $\hat{s}$	37.674	0.197	37.352	37.673	37.998
A: $\hat{\beta}$	0.941	0.004	0.934	0.941	0.948
B: $\hat{\beta}$	1.000	0.001	0.998	1.000	1.002
C: $\hat{\beta}$	0.976	0.000	0.975	0.976	0.977
D: $\hat{\beta}$	0.991	0.000	0.991	0.991	0.991

Table 3.43: $EFF_{2011/12}$: AR(1) NIP/IP estimation

Mean values for $\mu_{2011/2012}$ follow the trend coming out from past observations: decreasing in all areas. β coefficients give information about the strength of such trend. β standard deviations are too narrow.

One year forecasting: the *EFF* binomial model

Another way of formalizing *EFF* is with a binomial model. $EFF_{t,i}$ can be thought as the number of successes on $TEO_{t,i}$ trials, the probability of success being $\theta_{t,i}$, representing the student abilities and/or university facilities in earning credits. i represents the area and t the time. However, for first approximation, the model assumes a constant θ across time.

Both non informative and informative conjugate priors are used. Parameters for Beta distributions are derived with the method of moments. The model is:

$$
M_{EFF}^{Bin} = \begin{cases}
EFF_{i,t} \sim Bin(\theta_i, TEO_{i,t}), \ t = 1,2,3,4 \ \ i = \{A,B,C,D\} \\
NIP : \theta_i \sim Beta(1,1) \ \forall i \\
IP : \theta_A \sim Beta(58.67,44.34) \ \theta_B \sim Beta(1650,2536) \\
\theta_C \sim Beta(242.8,440.2) \ \theta_D \sim Beta(9023,16805)
\end{cases}
$$

The induced DAG is:
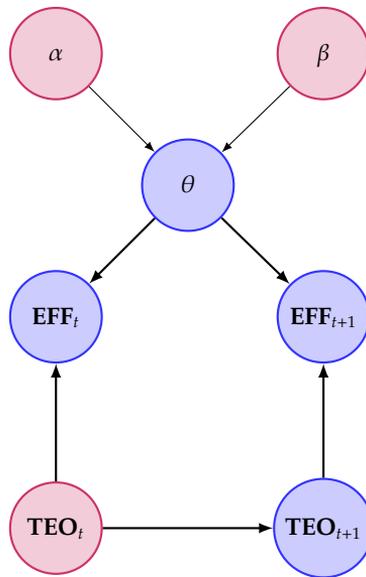
Figure 3.15: The spurious association between EFF_t and EFF_{t+1}

Both models lead to the same estimations so only one is showed.

AR(1)	NIP/IP
MD	18569
P	4
PD	18573

Table 3.44: EFF: fixed effects binomial model. NIP/IP deviance

n.iter	EFF - NIP/IP				
Node	Mean	Sd	5%	50%	95%
$A\ E\hat{F}F_{2011}$	88428.8	220.6	88066.0	88429.0	88792.0
$B\ E\hat{F}F_{2011}$	432306.5	568.3	431372.0	432307.0	433241.0
$C\ E\hat{F}F_{2011}$	378222.6	543.7	377328.0	378223.0	379116.0
$D\ E\hat{F}F_{2011}$	254550.4	450.4	253809.0	254550.0	255292.0
$A : \hat{\theta}$	0.569	0.001	0.567	0.569	0.570
$B : \hat{\theta}$	0.394	0.000	0.394	0.394	0.395
$C : \hat{\theta}$	0.355	0.000	0.354	0.355	0.355
$D : \hat{\theta}$	0.349	0.000	0.349	0.349	0.350

Table 3.45: EFF: fixed effects binomial model. NIP/IP estimation

EFF estimations are quite different from the normal model. Except for group *A* for which the value is higher, all other groups get a lower prevision. θ estimations are the *R* average among the 4 year observations. There are suspicious null standard deviations. An huge value of MD calls for a further investigation.

Figure 3.16 shows *EEF*, *TEO* and *R* time series by group. Two considerations on *R*: there are clear systematic trends, and patterns don't look like any of their two sources, *EFF* and *TEO*, meaning the knowledge of only one quantity among *EFF* and *TEO* is insufficient for prediction on *R*.
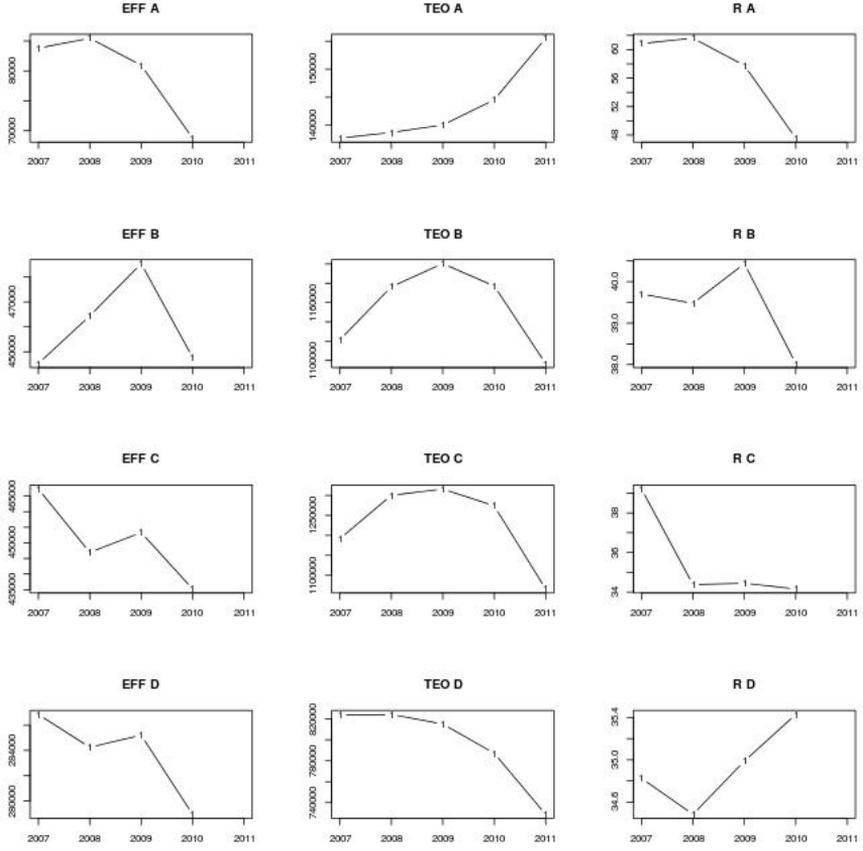
Figure 3.16: Time series on {EFF, TEO, R}

136

To take into account systematic variations on R it is used a binomial model with time-varying parameter θ_x (on logit scale, Congdon (2006) pag 122), for which R_X is an estimation.

The model is:

$$
M_{EFF}^{TVAR} = \begin{cases}
EFF_t \sim Bin(\theta_t, TEO_t), \ t = 1, 2, 3, 4, 5 \\
TEO_t = f(TEO_{t-1}) \\
\text{logit}(\theta_t) = b_t \\
b_t \sim N(\mu_t, \tau) \\
\mu_t = \beta \mu_{t-1} \\
NIP : \{\beta, \mu_1\} \sim N(0, 0.001), \ \tau \sim \Gamma(0.001, 0.001) \\
IP : \beta_1 \sim N(0.374, 1.133), \ \beta_2 \sim N(1.03, 11.48) \\
\beta_3 \sim N(1.079, 8.381), \ \beta_4 \sim N(0.979, 98.581) \\
\mu_{1,1} \sim N(0.417, 6.797), \ \mu_{1,2} \sim N(-0.401, 51.068) \\
\mu_{1,3} \sim N(-0.504, 22.009), \ \mu_{1,4} \sim N(-0.637, 112.420) \\
\tau_1 \sim \Gamma(0.9292, 0.04717), \ \tau_2 \sim \Gamma(0.9687, 0.002862) \\
\tau_3 \sim \Gamma(0.9544, 0.007514), \ \tau_4 \sim \Gamma(0.9898, 0.001185)
\end{cases}
$$

Informative prior parameters come from non informative prior model outcome.

The induced DAG is shown in figure 3.17:
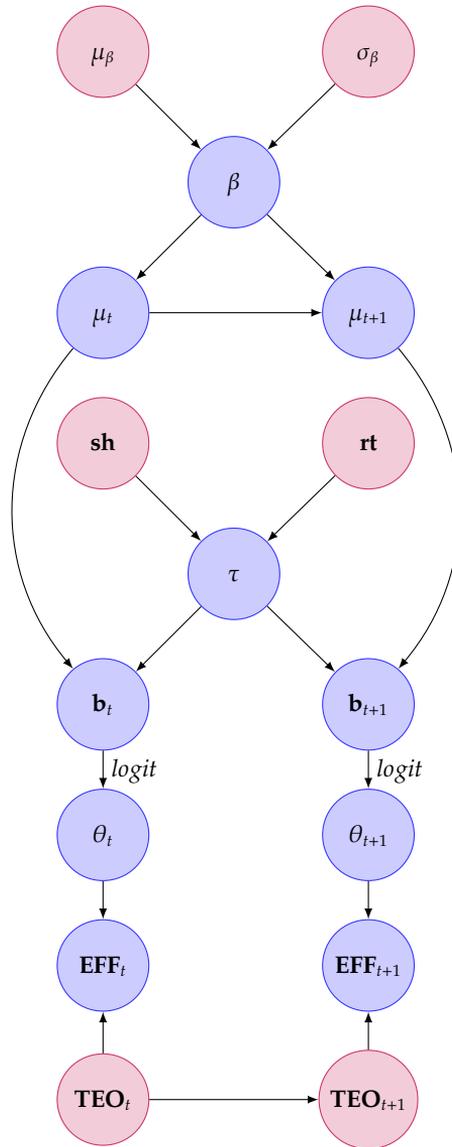
Figure 3.17: *EFF* time-varying binomial model induced DAG

Also for this model both non informative and informative prior lead to the same estimations, so results are shown once.

AR(1)	NIP/IP
MD	236
P	16
PD	252

Table 3.46: EFF: time-varying effects binomial model. NIP/IP deviance

n.iter	EFF - NIP/IP				
Node	Mean	Sd	5%	50%	95%
$A\ \hat{EFF}_{2011}$	81249	17584	54493	81060	109198
$B\ \hat{EFF}_{2011}$	423307	57833	352769	421306	498846
$C\ \hat{EFF}_{2011}$	348081	86355	243324	341004	471634
$D\ \hat{EFF}_{2011}$	259726	26463	232870	258903	286714
$A : \hat{\theta}_{2011}$	0.522	0.113	0.350	0.521	0.702
$B : \hat{\theta}_{2011}$	0.386	0.053	0.322	0.384	0.455
$C : \hat{\theta}_{2011}$	0.326	0.081	0.228	0.320	0.442
$D : \hat{\theta}_{2011}$	0.356	0.036	0.320	0.355	0.393
$A : \hat{\beta}$	0.390	1.233	-1.013	0.527	1.190
$B : \hat{\beta}$	1.023	0.278	0.817	1.038	1.226
$C : \hat{\beta}$	1.071	0.420	0.806	1.101	1.328
$D : \hat{\beta}$	0.974	0.185	0.907	0.983	1.055
$A : \hat{s}$	0.354	0.301	0.129	0.279	0.808
$B : \hat{s}$	0.101	0.131	0.031	0.066	0.280
$C : \hat{s}$	0.167	0.228	0.051	0.108	0.481
$D : \hat{s}$	0.064	0.102	0.020	0.042	0.162

Table 3.47: EFF: time-varying effects binomial model. NIP/IP estimation

139

With respect to the fixed effect, this model provides lower estimation points for *EFF* (except for group *D*). Moreover the θ estimation is relative only to the forecast year (actually it is a real forecast instead of an average within the period of observation as for the fixed effect model), and the MD returns to an acceptable value.

Two year forecasting: A2 2012 prevision

Now prediction can be fulfilled. For *S* and MR_X the same 2011 estimation values are used:

$$S=40658456 \ [37689367,43125385]$$
$$Median(R)_A=59.66 \ [57.55, 61.72]$$
$$Median(R)_B=43.51 \ [42.04, 45.01]$$
$$Median(R)_C=38.59 \ [37.24, 39.91]$$
$$Median(R)_D=37.11 \ [35.84, 38.41]$$

For *EFF* estimation the time-varying binomial model is used:

$$EFF_A=81058 \ [62834, 99763]$$
$$EFF_B=420689 \ [378187, 463038]$$
$$EFF_C=340883 \ [274670, 409420]$$
$$EFF_D=258906 \ [245242, 272680]$$

All these quantities lead to an estimation of:

$$A2_{2012}=2.18 \ [1.74,2.38]$$

with a lost respect to 2010 (2.76) of 709.92M€[1248.48M€,465.12M€] This discouraging result comes jointly from a reduction in *TEO* in the last year (see Table 3.41), which is a *EFF* parameter, and from a general negative trend in *R* (see Figure 3.16).

3.10 A2 Causal analysis

This section answers two capital questions: what variables does it need to intervene upon? and how much time does it need to move up in the ranking? These questions are going to be dealt with separately. From the definition formula it is clear that increasing *EFF* is more beneficial than decreasing *TEO*, because *EFF* is a squared term whereas *TEO* is linear. Moreover *TEO* depends on the number of students (by a coefficient equal to 60), which in turn depends mainly on three features: new enrolments, graduates and drop-outs. These are governed by their own rules and needs their own policy, and none of them is directly related with FFO, e.g. new enrolments depend on family incomes, drop-outs depend on educational offer students relationship, graduates regards the whole student carriers. *EFF* on the other hand is directly related to student characteristics and year by year educational offer sustainability, and can be investigated in the light of them.

The next section is dedicated to the search of weak performances, meaning which are the student characteristics and educational offer features with the lowest value of *R*. This search gives the idea about where it is useful to intervene upon.

The last section offers to decision-makers a three-year plan aimed to positioning the University of Palermo in the median position of the ranking.

3.10.1 Searching weak R performances

Starting from the efficiency on acquiring *CFU*, namely $R = \frac{EFF}{TEO}$, table 3.48 shows by area *R* negative performances both in terms of one year difference and respect to the national median.

141

	R_A	R_B	R_C	R_D
R_{2010}	60.6	40.6	34.8	34.7
R_{2011}	47.6	38.0	34.2	35.4
$R_{2011} - R_{2010}$	-13	-2.6	-0.6	0.7
$Median_{2010}$	59.7	43.5	38.6	37.1
$R_{2011} - Median_{2010}$	-12.1	-5.5	-4.4	-2.3

Table 3.48: R_{A2} and $Median(R)_{A2}$ by group

From now on, tables refer to C.Y. 2011 data. Table 3.49 shows females get always better results then males. This is a quite settled result.

Gender	R_A	R_B	R_C	R_D
F	48.12	40.43	36.84	36.52
M	46.92	36.13	27.26	33.40

Table 3.49: R_{A2} by group and gender

Table 3.50[25] shows lower values in Bachelor rather than in other type of courses.

	R_A	R_B	R_C	R_D
MA	-	51.8	56.3	50.2
MALT	47.6	47.9	-	40.0
MA old	-	51.4	44.7	37.4
BA	-	33.7	31.2	31.1

Table 3.50: R_{A2} by group and type of degree course

---

[25]BA= Bachelor, MA=master, MA old= Master old classification, MALT= long term (4/5/6 years) master

Table 3.51 adds to the table above information on year of enrolment[26]: percentage of acquired CFU increases with years of enrolments, first year is always the worst case (expect for group A), in particular for bachelors.

A	MA	MALT	MA old	BA
1C	-	45.58	-	-
2C	-	35.57	-	-
3C	-	33.98	-	-
>3C	-	54.40	-	-
1FC	-	-	-	-
2FC	-	38.33	-	-
Oth	-	51.19	-	-

B	MA	MALT	MA old	BA
1C	47.95	37.78	37.45	29.43
2C	78.55	50.19	61.26	36.68
3C	-	44.03	-	47.62
>3C	-	60.87	-	25.00
1FC	-	0.00	-	41.39
2FC	-	-	-	31.78
Oth	52.22	44.77	41.45	20.95

C	MA	MALT	MA old	BA
1C	49.87	-	74.22	29.52
2C	84.08	-	57.05	37.61
3C	-	-	-	42.07
>3C	-	-	-	40.34
1FC	-	-	-	37.75
2FC	-	-	-	27.85
Oth	16.25	-	33.33	17.61

D	MA	MALT	MA old	BA
1C	50.41	26.231	-	32.37
2C	51.11	33.942	39.06	45.36
3C	-	40.624	-	46.73
>4C	-	49.024	-	0.00
1FC	-	4.444	-	36.69
2FC	-	25.000	-	26.35
Oth	0.00	55.754	36.62	15.61

Table 3.51: R_{A2} by group, type of degree course and year of enrolment

Next table shows the dramatic plague of students with no credits achieved throughout one year:

Area	A	B	C	D
No CFU	18.68	19.04	23.30	23.59

Table 3.52: Percentage of zero-credit students by area

3.10.2 Intervening on R

From the previous section it emerges a non encouraging situation, R values are in general lower than the national median. Figure 3.18 shows R variability among degree courses.
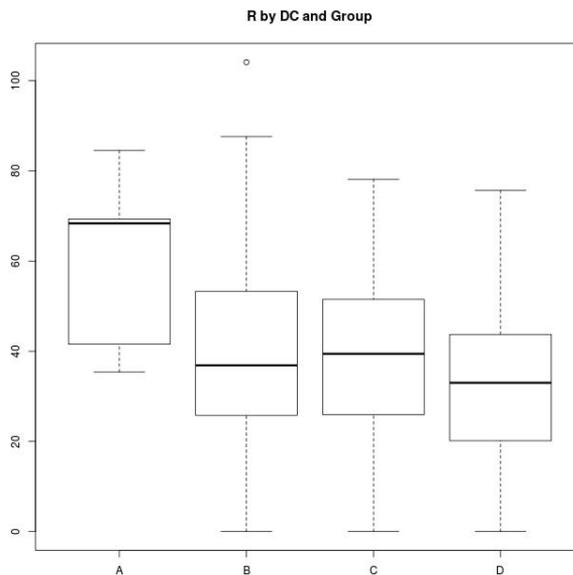
Figure 3.18: Degree Courses R_{A2} by groups

Given this status, one aim could be setting a policy to reach, in an limited number of years, the median value MR of the area. Thus each year, every degree course compares R and MR and according to the difference defines its new target T, namely the new value it must reach:

C1) if $R \geq MR$ then $T \geq R$;

C2) if $0.95MR \leq R < MR$ then $T \geq MR$;

C3) if $0.90MR \leq R < 0.95MR$ then $T \geq 1.05R$;

C4) if $R < 0.90MR$ then $T \geq 1.1R$;

The rules above are interpreted as follows: when the ratio is higher than the median, no effort is required, except not reducing its value. When the ratio is less than the median but not more than 5%, the effort required is to catch up the median. When the ratio is between 90% and 95% then degree courses have to increase it at least by 5%. Finally, when the ratio is smaller than the median by 10%, the ratio must be increased at least by 10%. This procedure, if satisfied, reaches the goal in no more than three years.

The table 3.76 shows for three consecutive years the percentage of degree courses in each class and the value of $A2$[27]:

Year	C1	C2	C3	C4	A2
2010	38.6	12.0	14.6	34.8	2.76
2011	40.3	11.4	8.9	39.4	2.81
2012	53.1	2.9	12.3	31.7	2.90
2013	56.0	3.1	13.7	27.1	2.97

Table 3.53: Percentage of target classes degree courses by area and year

The induced DAG is shown in figure 3.30:

[27]2010 and 2011 got a different number of degree courses; 2011-13 assume it constant. *TEO*, MR and *S* are assumed constant.
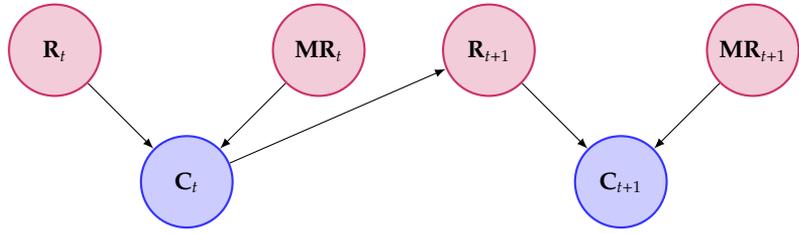
Figure 3.19: Act on EFF policy DAG

Only a word about the link $C_t \rightarrow R_{t+1}$. The class C_t doesn't be interpreted as a (direct) factor of R_{t+1} because R_{t+1} depends on EFF_{t+1} and TEO_{t+1}. The link represents the "intervention requirement ": if $C_t \in \{2, 3, 4\}$ then an intervention either on *EFF* or on *TEO* is needed. Given that the DAG marginalizes on both $\{EFF_{t+1}, TEO_{t+1}\}$, the link becomes direct.

In A.Y. 2010/2011 University of Palermo took (at least) two actions in order to increase *EFF*: 80 tutoring contracts and several remedial courses for the so called "hard to pass teaching", namely where students have problems to pass. Such interventions is called *AOE* (act on *EFF*).

The causal effect of *AOF* on *EFF* can be computed as the difference between the *EFF* predicted value, using data collected in past years when *AOF* did not occur, and the *EFF* 2011 observed value when *AOF* did. Unfortunately that value will be known in late 2012 (due to delay in data entry) and so only the predicted value can be computed (as done). What is important to underline now is the causal assumption that the difference between prediction and observation (under intervention) is interpreted as the causal effect of *AOF* on *FF* without any possible empirical verifiability, that is we are in the field of counterfactual assumptions formalized by the potential outcomes (Rubin, 2005).

147

3.11   Indicator B1

Indicator $B1$ belongs to the set of indicators which deals with scientific research quality. It is defined as the percentage of researchers positively evaluated in national projects (PRIN) during years 2006-2009, weighted by the ratio of success of the scientific area.

The National evaluation agency, Anvur, criticized the double use of the same evaluation: universities already receive fund based on positive evaluation of research projects PRIN, through the financed projects, so why allocating again fund based on the same evaluation?[28]

Formally it is defined as the ratio, (R_i), between the mean of positively evaluated researchers in a 4-year period (PE_i) and the mean of the total number of researcher during the same period, $(TTEA_i)$, for each of the 14 scientific areas, i. This ratio is then divided by the national median, (MR_i), and multiplied again by $TTEA_i$, obtaining the local value per area, I_i. Then final value is obtained by adding up for all areas (I) and finally normalized by dividing for the sum of all 54 university (S).

Replacing PE and $TTEA$ with EFF and TEO, the formula translates into indicator $A2$, that means all remarks already made about $A2$, e.g. university-size depending, and the same analysis structure still holds for $B1$.

3.11.1   B1 Structural equations and induced DAG

The formal definition of $B1$ is depicted by the following structural equations system:

---

[28]http://cronaca.anvur.it/2010/12/alla-ricerca-di-un-premio.html

$$
B1 \begin{cases}
i = 1, ..., 14 \ = area \\
TTEA_i = \frac{TTEA_{05,i} + TTEA_{06,i} + TTEA_{07,i} + TTEA_{08,i}}{4} \\
PE_i = \frac{PE_{05,i} + PE_{06,i} + PE_{07,i} + PE_{08,i}}{4} \\
R_i = 100 \frac{PE_i}{TTEA_i} \\
R_i^* = \{R_{1,i}, ..., R_{53,i}\} \\
MR_i = Median(\{R_i, R_i^*\}) \\
I = \sum_{i=1}^{14} \frac{R_i}{MR_i} TTEA_i = \sum_{i=1}^{14} \frac{PE_i}{MR_i} \\
S = \sum_{k=1}^{54} I_k = I + \sum_{i=k}^{53} I_i = I + S^* \\
B1 = 100 \frac{I}{S}
\end{cases}
$$

Notice that the indicator actually doesn't depend directly on $TTEA$ because $I = \frac{R}{MR} TTEA = \frac{1}{MR} \frac{PE}{TTEA} TTEA = \frac{PE}{MR}$.

As for the previous indicator, PE_i may be modelled as the number of successes in $TTEA_i$ trials, θ_i being the probability of success, that is the set of conditions may facilitate or make difficult to be positively evaluated. $TTEA_{i,t}$ is a function of time [29], but since the statistical models employed either marginalize on it or act conditioned on it, the functional form doesn't need to be further investigated. As usual, (I, S^*) and (R_i, R_i^*) can be correlated due to national common causes.

Such considerations translate in other equations to be added to the system:

$$
\begin{cases}
TTEA_{i,t} = f(TTEA_{i,t-1}) \\
PE_{i,t} \sim Bin(\theta_i, TTEA_{i,t}) \\
Corr(S^*, I) \neq 0 \\
Corr(R_i^*, R_i) \neq 0
\end{cases}
$$

The induced DAG is represented in figure 3.20:

---

[29]through the formula $TTEA_t = TTEA_{t-1} + IN_t - OUT_t$ where IN_t and OUT_t are respectively new enrolled researchers and retirements.
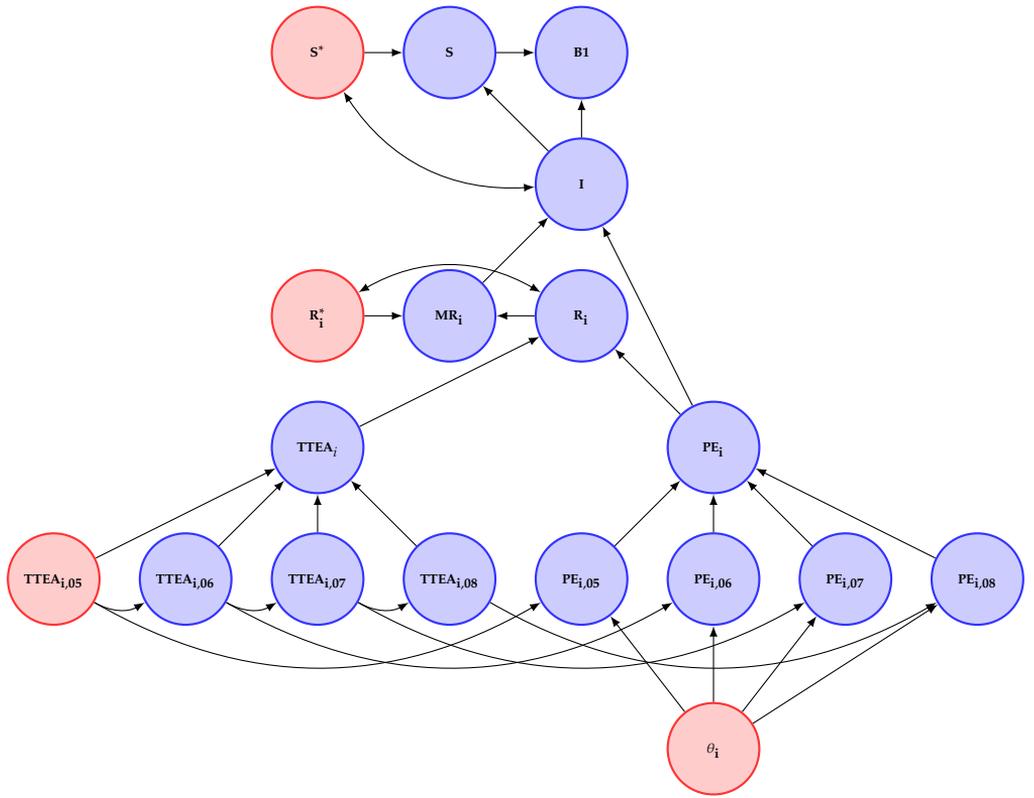
Figure 3.20: $B1$ induced DAG

Table 3.54 shows three-year allocations schema, that is which quantities are known and which have to be estimated (in bold) in order to compute $B1$. National data are labelled with the same year as $B1$:

Year	T		T+1		T+2	
Quantity	RT	AT	RT	AT	RT	AT
TTEA	2005-08	01/01/09	2006-2009	01/01/10	2007-2010	01/01/11
PE	2005-08	4/10	2006-2009	11/11	2007-2010	**-**
MR	2010	12/10	2011	**12/11**	2012	**12/12**
S	2010	12/10	2011	**12/11**	2012	**12/12**
B1	2010	12/10	2011	**12/11**	2012	**12/12**

Table 3.54: *B1* Reference Time & Available Time

As usual, quantities responsible for $B1$ 2010 are already known and ready for a retrospective analysis. For 2011 estimate only national quantities are needed, while for 2012 *TTEA* is the only known quantity. Let us start with 2010 results.

3.11.2   B1 2010 results

In 2010, the University of Palermo achieved the following results:

Area	TTEA(05-08)	PE(05-08)	R	MR	R-MR	I_i
01	83.5	30.25	36.2	43	-6.8	70,89
02	68.25	33.75	49.5	51	-1.5	65,68
03	108	52.75	48.8	55	-6.2	95,18
04	43.5	15.75	36.2	49	-12.8	31,89
05	187	60.75	32.5	49	16.5	123,67
06	378.75	79	20.9	27	-6.1	291,97
07	122.75	64	52.1	50	2.1	128,00
08	198.756	101.25	50.9	51	-0.1	197,62
09	17.5	96.5	54.1	53	1.1	180,84
10	180.25	61	33.8	36	-2.2	167,26
11	168.5	45.5	27.0	37	-10.0	124,19
12	172.25	45.5	26.4	31	-4.6	146,92
13	106.5	35	32.9	33	-0.1	104,74
14	45.5	13	28.6	34	-5.4	38,22
T	2.042	734	35.9	-	-	I=1767.06
						(S=53937.25)

Table 3.55: B1 2010 quantities

$B1$=3.28 is the ninth among all universities. The local rate exceeds the median value only in two areas, while in three areas it is inferior at least of 10%. These poor results compared with the quite high value of the indicator give the intuition that the university size effect operated. In the retrospective section this intuition will be confirmed.

3.12 B1 retrospective analysis

This section highlights possible delays or errors in data entry or other hypothetical (counterfactual) conditions able to explain what actions would

have be performed to increase the indicator.

Q1 What $B1$ would have been had it computed today (namely was there a measurement error)?

A1 $B1$ would be 3.30. The official value is 3.28 meaning there was a little delay or errors in data entry in PE_i values. Actually, this indicator is unlikely to be subject to errors of this kind.

Q2 What $B1$ would be had the local ratio equal at least to the national median?

A2 $B1$ would be 3.78 corresponding to an increasing allocation of 814,97M€.

Q3 what $B1$ would be had all universities achieved 100% PE?

Q4 $B1$ would be 3.57. This counterfactual correspond to a definition of $B1 = 100 \frac{\sum_{i=1}^{14} TTEA_i}{\sum_{i=1}^{14} \sum_{k=1}^{54} TTEA_{i,k}}$, that is depending only on the university dimension.

Q4 what $B1$ would be had R_i not multiplied by $TTEA_i$?

A4 $B1$ would be 1.66. This counterfactual correspond to a definition of $B1 = 100 \frac{\sum_i \frac{R_i}{MR_i}}{\sum_{i,k} \frac{R_{i,k}}{MR_{i,k}}}$, that is erasing the university dimension effect. Notice that the observed value 3.28, is closer to 3.57 (Q3) than to 1.66, confirming the dimension effect.

Q5 what $B1$ would be had PE_i increased by 1%?

A5 $B1$ would be 3.30, meaning every increasing of 1% in PE_i would lead an increasing of 0.2 in $B1$, that is a gain of 332,64M€

3.13   B1 previsional analysis

As for previous indicators, after checking the past, the most urgent task is predicting the future. In the next section national quantities will be forecast one year ahead in order to make prediction on $B1$ 2011.

3.13.1 Short-term prevision

Table 3.56 compares the mean of positive evaluated researchers in two subsequent four-year periods related with $B1$ 2010 and $B1$ 2011 with the aims of evaluating differences in performances. The last column measures the difference between the number of researchers on 2009 and on 2005, because two consecutive indicators depend on it.

Area	PE(05-08)	PE(06-09)	2-1	Δ=(2-1)/1 %	PE09-PE05
01	30.25	26	-4.25	-14.38	-19
02	33.75	33.25	-0.50	-1.48	-1
03	52.75	51	-1.75	-3.32	-9
04	15.75	15.25	-0.50	-3.17	-2
05	60.75	55	-5.75	-9.47	-6
06	79	69.75	-9.25	-11.70	-38
07	64	55.75	-8.25	-12.89	-34
08	101.25	86	-15.25	-15.06	-63
09	96.5	85.75	-10.75	-11.14	-41
10	61	57	-4	-6.56	-16
11	45.5	51.25	5.75	12.64	24
12	45.5	41.75	-3.75	-8.24	-17
13	35	28.25	-6.75	-19.29	-27
14	13	15	2	15.38	7

Table 3.56: B1 2010 national variables - one year difference

Two areas only increased the average of positively evaluated researchers. In the other ten areas, there were substantial decrements, on average by 8%. This would lead, had θ and MR equal to 2010 observed value, to a prevision for $B1$ 2011 of 3.02, which means a temporary loss of 449.07M€. Next table highlights how only little areas (in terms of PE_i) increased their PE_i, while as dimension grew, loss increased too.

154

Quartile	Q1	Q2	Q3	Q4
PE	(13,34.1]	(34.1,49.1]	(49.1,63.2]	(63.2,101]
$E(\Delta|Q)$	40.422	-9.012	-32.472	-50.871

Table 3.57: B1 2010 national variables - one year difference

One year forecasting: the S case

Figure 3.21 shows the casual diagram involving $\S_t$ and $\S_{t+1}$.
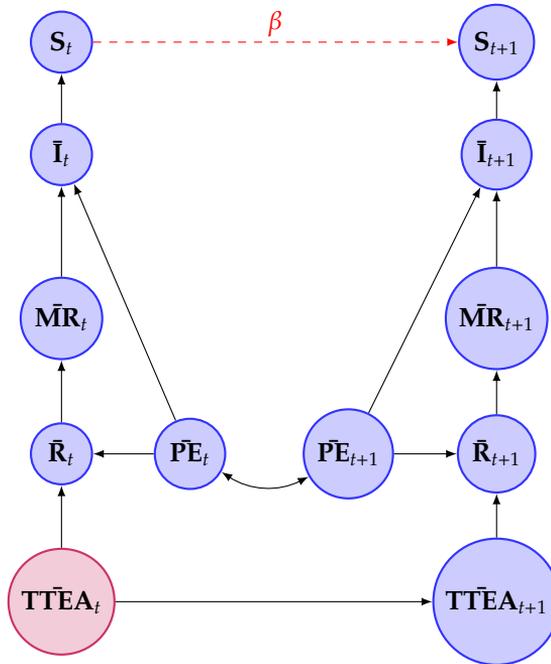
Figure 3.21: The spurious association between S_t and S_{t+1}

However being only one observation available, the normal approxima-
tion is used.

The model is:

$$
S \begin{cases}
\theta_{2011} \sim N(\hat{\mu}, \hat{\sigma}) \\
\hat{\mu} = k_1 S_{2010}^{obs} \\
\hat{\sigma} = k_2 \sigma(I_{2010}^{obs}) \\
\{k_1, k_2\} \in \{0.95, 1, 1.05\}
\end{cases}
$$

Where S_{2010}^{obs} is the 2010 S observed value, and $\sigma(I_{2010}^{obs})$ is the empirical standard deviation among all 54 I values for 2010. Next table shows summaries form 1M samples drawn from such model:

n.iter=1Ml	S 2011					
$\S_{2011}$	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	46583	49076	49560	49566	50059	52671
0.95μ , 1sd	46786	49029	49572	49565	50090	52336
1μ, 0.95sd	49438	51680	52179	52179	52683	55482
1μ, 1sd	49282	51641	52168	52165	52685	55125
1.05μ, 1sd	52074	54272	54793	54794	55315	57846
1μ, 1.05sd	49335	51632	52180	52169	52705	55272
1.05μ, 1.05sd	51178	54227	54788	54782	55332	57789

Table 3.58: Simulation table for $\S_{B1}$ 2011

One year forecasting: the MR case

The model is the same as for S:

$$
MR_i \begin{cases}
MR_{i,2011} \sim N(\hat{\mu}_i, \hat{\sigma}_i) \quad 1 = \{1, ..., 14\} \\
\hat{\mu}_i = k_1 MR_{i,2010}^{obs} \\
\hat{\sigma}_i = k_2 \sigma(MR_{i,2010}^{obs})/2 \\
\{k_1, k_2\} \in \{0.95, 1, 1.05\}
\end{cases}
$$

Where $MR_{2010}^{i,obs}$ is the MR_i observed value in 2010, and $\sigma(MR_{2010}^{i,obs})/2$ is the empirical standard deviation among all 54 MR_i values for 2010 divided by

2 (ad hoc imputation).

The causal diagram is:
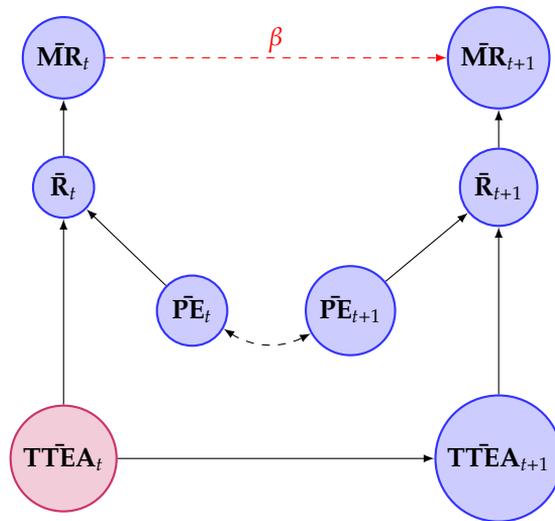
Figure 3.22: The spurious association between $\hat{MR}_t$ and $\hat{MR}_{t+1}$

It follows the simulation table for all 14 areas:

n.iter=1Ml	*MR* 2011 AREA 1					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	22.48	37.46	40.77	40.78	44.12	65.14
0.95μ , 1sd	20.66	37.47	40.91	40.90	44.45	59.07
1μ, 0.95sd	25.97	39.79	43.04	43.06	46.27	61.34
1μ, 1sd	25.13	39.46	42.97	42.99	46.43	64.51
1.05μ, 1sd	25.05	41.72	45.20	45.19	48.74	65.62
1μ, 1.05sd	23.06	39.36	43.05	43.06	46.77	62.39
1.05μ, 1.05sd	23.29	41.54	45.19	45.15	48.83	65.65
n.iter=1Ml	*MR* 2011 AREA 2					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	30.83	45.13	48.51	48.45	51.75	67.50
0.95μ , 1sd	29.97	44.97	48.47	48.42	51.85	67.11
1μ, 0.95sd	31.54	47.73	51.05	51.05	54.42	68.04
1μ, 1sd	34.79	47.57	51.06	51.09	54.65	71.66
1.05μ, 1sd	31.58	50.11	53.56	53.53	57.01	72.77
1μ, 1.05sd	30.95	47.38 5	0.96	51.00	54.67	71.00
1.05μ, 1.05sd	33.04	49.79	53.55	53.48	57.19	73.31
n.iter=1Ml	*MR* 2011 AREA 3					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	35.29	48.72	52.30	52.21	55.63	70.80
0.95μ , 1sd	32.76	48.77	52.16	52.15	55.58	71.47
1μ, 0.95sd	35.06	51.65	55.04	55.03	58.43	72.60
1μ, 1sd	37.50	51.59	55.03	55.00	58.52	73.78
1.05μ, 1sd	39.47	54.25	57.73	57.73	61.20	75.26
1μ, 1.05sd	32.78	51.34	55.00	54.99	58.61	77.91
1.05μ, 1.05sd	35.44	54.16	57.67	57.74	61.44	79.75
n.iter=1Ml	*MR* 2011 AREA 4					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	28.56	43.27	46.57	46.54	49.84	65.12
0.95μ , 1sd	26.90	42.88	46.45	46.42	49.99	66.61
1μ, 0.95sd	29.48	45.78	49.06	49.05	52.35	69.08
1μ, 1sd	30.14	45.45	48.97	48.93	52.40	68.66
1.05μ, 1sd	28.74	48.09	51.56	51.49	54.95	73.64
1μ, 1.05sd	25.94	45.22	48.85	48.98	52.70	71.73
1.05μ, 1.05sd	30.89	47.81	51.52	51.54	55.24	71.54
n.iter=1Ml	*MR* 2011 AREA 5					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	28.52	43.28	46.63	46.62	49.92	64.74
0.95μ , 1sd	27.00	43.13	46.52	46.57	50.07	66.20
1μ, 0.95sd	29.73	45.68	49.09	49.04	52.34	67.16
1μ, 1sd	31.91	45.59	48.99	49.04	52.46	68.73
1.05μ, 1sd	29.21	47.99	51.50	51.51	55.00	70.47
1μ, 1.05sd	25.91	45.32	49.03	49.01	52.66	69.45
1.05μ, 1.05sd	30.71	47.80	51.44	51.46	55.04	73.05
n.iter=1Ml	*MR* 2011 AREA 6					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	6.753	22.343	25.597	25.641	28.961	42.956
0.95μ , 1sd	6.095	22.122	25.577	25.598	29.057	45.509
1μ, 0.95sd	7.763	23.700	26.957	26.990	30.302	45.330
1μ, 1sd	7.919	23.568	27.064	27.057	30.511	45.024
1.05μ, 1sd	9.546	24.864	28.331	28.334	31.880	48.671
1μ, 1.05sd	7.003	23.233	26.938	26.940	30.610	46.461
1.05μ, 1.05sd	6.453	24.707	28.431	28.381	32.010	50.098

Table 3.59: Simulation table for *Median(R)*$_{B1}$ 2011 AREA 1-6

n.iter=1Ml	MR 2011 AREA 7					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	29.15	44.16	47.46	47.46	50.82	68.85
0.95μ, $1sd$	25.80	43.97	47.42	47.46	50.98	66.93
1μ, $0.95sd$	28.22	46.71	50.03	50.01	53.35	68.43
1μ, $1sd$	29.55	46.44	49.98	49.95	53.50	71.50
1.05μ, $1sd$	30.78	49.03	52.59	52.54	56.01	72.45
1μ, $1.05sd$	31.28	46.25	50.00	49.98	53.58	74.59
1.05μ, $1.05sd$	30.25	48.79	52.46	52.50	56.11	75.11
n.iter=1Ml	MR 2011 AREA 8					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	28.65	45.19	48.57	48.52	51.87	70.84
0.95μ, $1sd$	29.46	45.04	48.54	48.49	51.97	68.30
1μ, $0.95sd$	31.78	47.61	50.95	50.98	54.30	68.48
1μ, $1sd$	32.32	47.57	51.08	51.06	54.53	70.82
1.05μ, $1sd$	32.80	50.11	53.60	53.56	57.03	73.16
1μ, $1.05sd$	30.22	47.30	51.00	51.01	54.72	73.07
1.05μ, $1.05sd$	32.17	49.93	53.50	53.56	57.17	74.25
n.iter=1Ml	MR 2011 AREA 9					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	30.61	47.10	50.29	50.35	53.67	66.92
0.95μ, $1sd$	32.29	46.78	50.26	50.31	53.87	68.50
1μ, $0.95sd$	34.64	49.59	52.87	52.95	56.33	72.68
1μ, $1sd$	32.82	49.62	52.99	53.04	56.56	72.72
1.05μ, $1sd$	34.20	52.16	55.68	55.68	59.17	75.23
1μ, $1.05sd$	33.17	49.25	52.98	52.95	56.57	73.11
1.05μ, $1.05sd$	33.81	52.05	55.77	55.75	59.40	73.77
n.iter=1Ml	MR 2011 AREA 10					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	13.65	30.86	34.19	34.24	37.62	53.49
0.95μ, $1sd$	14.93	30.61	34.17	34.17	37.68	55.92
1μ, $0.95sd$	15.31	32.64	35.96	35.94	39.31	57.28
1μ, $1sd$	17.11	32.39	35.98	35.93	39.55	53.65
1.05μ, $1sd$	14.35	34.32	37.84	37.79	41.35	56.06
1μ, $1.05sd$	14.13	32.47	36.08	36.07	39.69	57.10
1.05μ, $1.05sd$	19.15	34.23	37.87	37.86	41.50	58.24
n.iter=1Ml	MR 2011 AREA 11					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	15.29	31.86	35.15	35.18	38.49	54.34
0.95μ, $1sd$	16.34	31.68	35.13	35.15	38.70	56.91
1μ, $0.95sd$	18.95	33.67	37.06	37.00	40.33	57.62
1μ, $1sd$	19.81	33.51	37.07	37.05	40.54	57.12
1.05μ, $1sd$	20.24	35.36	38.92	38.85	42.34	57.11
1μ, $1.05sd$	14.60	33.17	36.96	36.95	40.61	60.44
1.05μ, $1.05sd$	17.10	35.14	38.87	38.79	42.48	59.94
n.iter=1Ml	MR 2011 AREA 12					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, $0.95sd$	11.00	26.15	29.51	29.50	32.89	48.20
0.95μ, $1sd$	8.31	25.96	29.45	29.46	32.95	47.91
1μ, $0.95sd$	11.59	27.74	31.10	31.07	34.43	48.60
1μ, $1sd$	12.90	27.49	30.95	30.98	34.43	50.35
1.05μ, $1sd$	14.07	28.95	32.37	32.43	35.87	52.02
1μ, $1.05sd$	11.21	27.45	31.15	31.14	34.84	53.35
1.05μ, $1.05sd$	9.959	28.873	32.550	32.523	36.177	52.293

Table 3.60: Simulation table for $Median(R)_{B1}$ 2011 AREA 7-12

n.iter=1Ml	MR 2011 AREA 13					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	11.16	28.05	31.34	31.36	34.70	49.74
0.95μ, 1sd	7.013	27.848	31.352	31.365	34.904	51.753
1μ, 0.95sd	12.59	29.69	32.93	32.95	36.20	50.95
1μ, 1sd	13.41	29.45	33.05	33.03	36.61	49.86
1.05μ, 1sd	13.44	31.08	34.61	34.63	38.12	54.28
1μ, 1.05sd	12.66	29.43	33.06	33.05	36.63	52.27
1.05μ, 1.05sd	15.45	31.06	34.74	34.69	38.35	54.64
n.iter=1Ml	MR 2011 AREA 14					
θ_{2011}	Min	25%	50%	Mean	75%	Max
0.95μ, 0.95sd	13.17	29.02	32.22	32.27	35.54	49.79
0.95μ, 1sd	11.82	28.85	32.32	32.34	35.80	52.16
1μ, 0.95sd	15.46	30.71	33.96	34.01	37.28	53.44
1μ, 1sd	12.92	30.45	33.90	33.95	37.47	52.97
1.05μ, 1sd	15.97	32.12	35.66	35.63	39.11	57.71
1μ, 1.05sd	14.23	30.22	33.97	33.97	37.73	55.44
1.05μ, 1.05sd	15.06	31.95	35.79	35.74	39.51	55.69

Table 3.61: Simulation table for $Median(R)_{B1}$ 2011 AREA 13-14

One year forecasting: B1 2011 prevision

Now every quantity needed to make a prediction on B1 2011 is known. $\S_{2011}$ is estimated with the mean and the couple (Min, Max) of $N(\mu, \sigma)$, namely the forth row of Table 3.60:

$$S=52165 \ (49282, 55125]$$

Due to the large credential intervals for MR_{2011}, the estimate is performed with the couple (25pc, 75pc) of $N(\mu, \sigma)$ (see Tables 3.59–3.61):

$$MR_{mean}=\{42.99, 51.09, 55.00, 48.93, 49.04, 27.05, 49.95, 51.06, 53.04, 35.93,$$
$$37.05, 30.98, 33.03, 33.95\}$$
$$MR_{25pc}=\{39.46, 47.57, 51.59, 46.45, 45.59, 23.56, 46.44, 47.57, 49.62, 32.39,$$
$$33.51, 27.49, 29.45, 30.45\}$$
$$MR_{75pc}=\{46.43, 54.65, 58.52, 52.40, 52.46, 30.51, 53.50, 54.53, 56.56, 39.55,$$
$$40.54, 34.43, 36.61, 37.47\}$$

PE is already known (see Table 3.56):

PE={26,33.25,51,15.25,55,69.75,55.75,86,85.75,57,51.25,41.75,28.25,15}

The final estimate is:

$$B1_{2011}=3.02\ [2.63,3.49]$$

Compared to *B*1 2010 there is an expected loss of 449.07M€(-1081,08M€, +349272M€].

3.13.2 Long-term prevision

Long term prevision deals essentially with PE_i prediction, both because *TTEA* is known and because it affects *B*1 only through *PE*.

One year forecasting: the PE case

Figures 3.23-3.26 compare $TTEA_i$, PE_i and R_i in a 5-year time series, displaying that R_i patterns are similar to the PE_i ones. This result can be explained by the low variability of $TTEA_i$ compared to PE_i as shown in the next table:

area	TTEA	PE	R
AREA 01	2.302	11.234	12.821
AREA 02	2.408	5.167	8.190
AREA 03	2.588	16.697	14.165
AREA 04	1.095	5.857	13.992
AREA 05	5.32	10.92	5.47
AREA 06	6.107	30.158	7.636
AREA 07	3.421	13.191	10.799
AREA 08	4.494	16.547	7.497
AREA 09	4.324	11.345	6.081
AREA 10	6.107	13.554	7.342
AREA 11	4.147	16.903	9.576
AREA 12	7.436	16.634	8.767
AREA 13	3.564	13.387	11.559
AREA 14	1.517	8.264	17.191

Table 3.62: TTEA, PE and R standard deviations
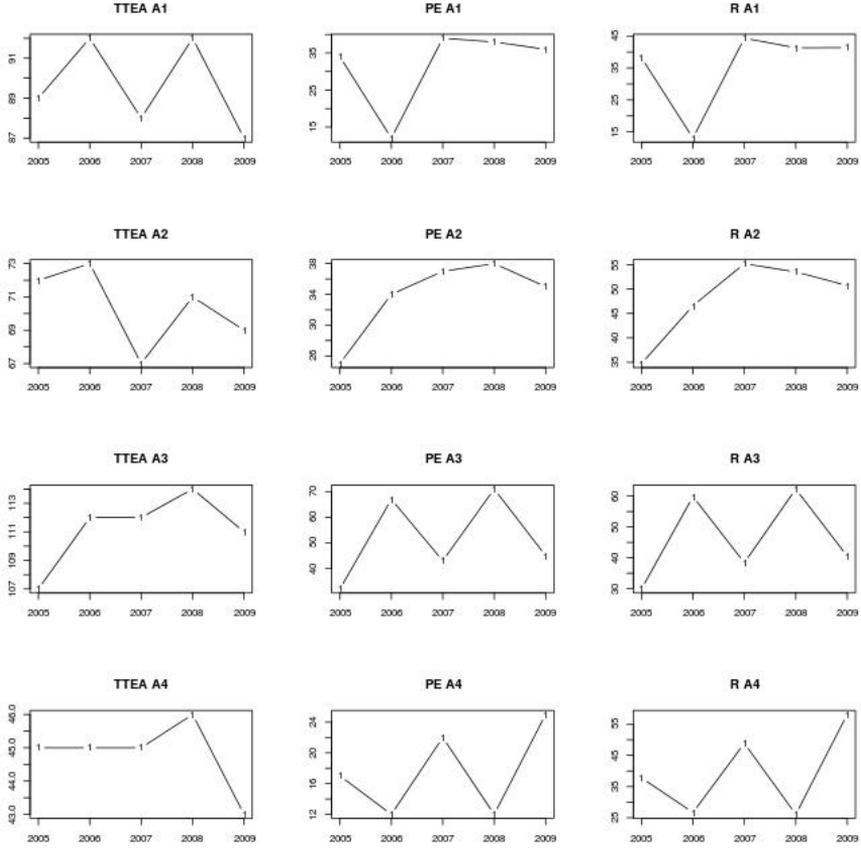
Figure 3.23: PE_i vs TEA_i vs R_i time series - Area 1-4

163
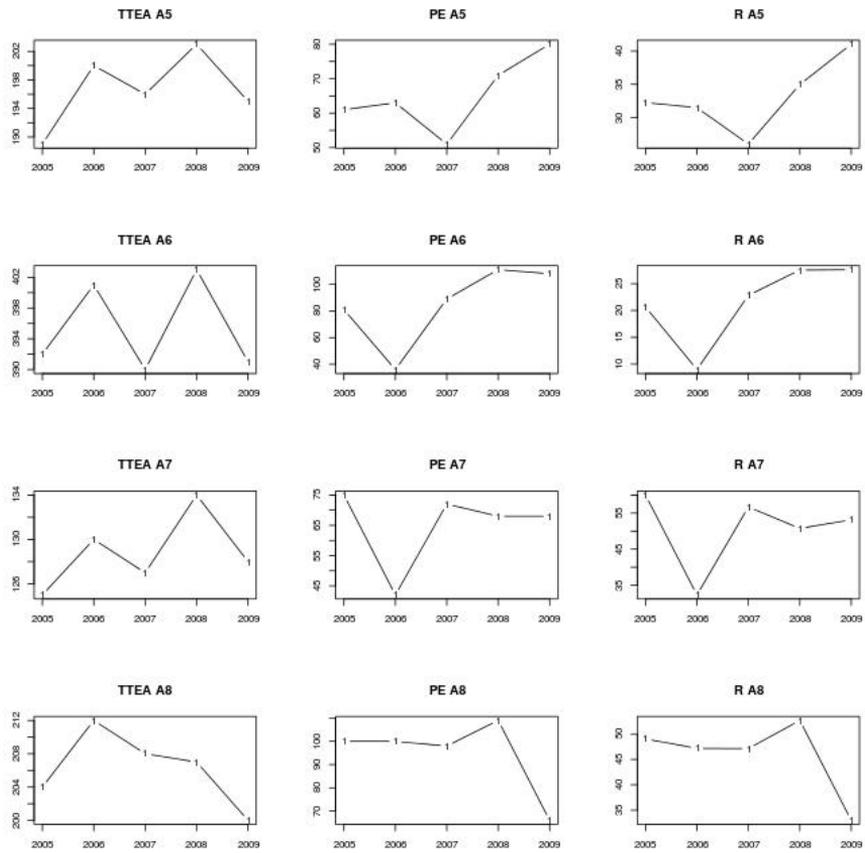
Figure 3.24: PE_i vs TEA_i vs R_i time series - Area 5-8

164
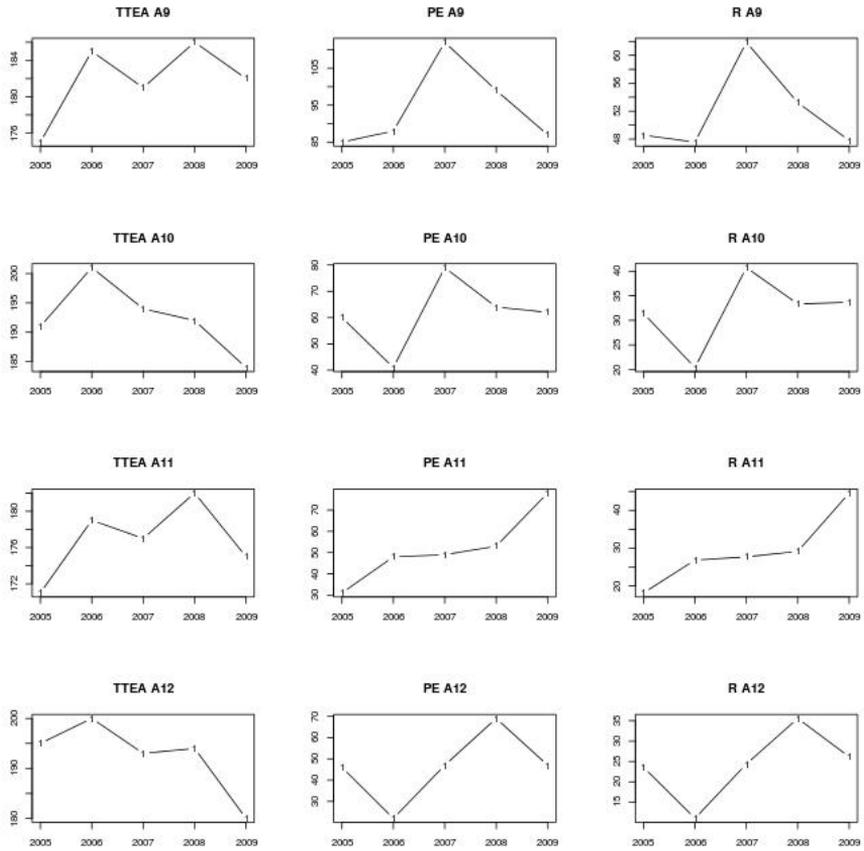
Figure 3.25: PE_i vs TEA_i vs R_i time series - Area 9-12
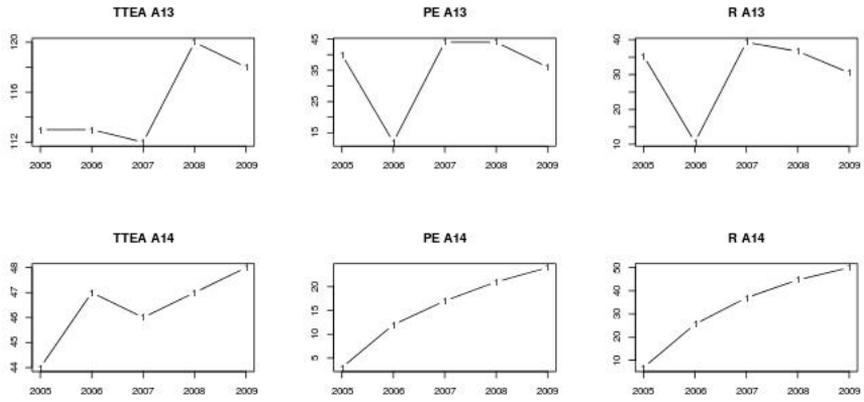
Figure 3.26: PE_i vs TEA_i vs R_i time series - Area 13-14

166

In order to make a prediction, a binomial model with fixed effect on θ is applied on PE_i.

The model is:

$$M_{PE}^{Bin} = \begin{cases} PE_{i,t} \sim Bin(\theta_i, TTEA_{i,t}), \ i = 1, ..., 14 \ t = 1, 2, 3 \\ NIP : \theta_i \sim Beta(1, 1) \ \forall i \\ IP : \\ \theta_1 \sim Beta(3.416, 7.667), \theta_2 \sim Beta(13.258, 16.227) \\ \theta_3 \sim Beta(3.266, 4.468), \theta_4 \sim Beta(10.526, 19.668) \\ \theta_5 \sim Beta(15.014, 37.587), \theta_6 \sim Beta(4.986, 22.411) \\ \theta_7 \sim Beta(7.277, 8.421), \theta_8 \sim Beta(6.172, 8.217) \\ \theta_9 \sim Beta(6.713, 7.544), \theta_{10} \sim Beta(11.426, 26.717) \\ \theta_{11} \sim Beta(24.072, 66.305), \theta_{12} \sim Beta(5.041, 17.743) \\ \theta_{13} \sim Beta(2.709, 7.476), \theta_{14} \sim Beta(2.697, 7.305) \end{cases}$$

Beta parameters are estimated according to the method of moments. The induced DAG is:
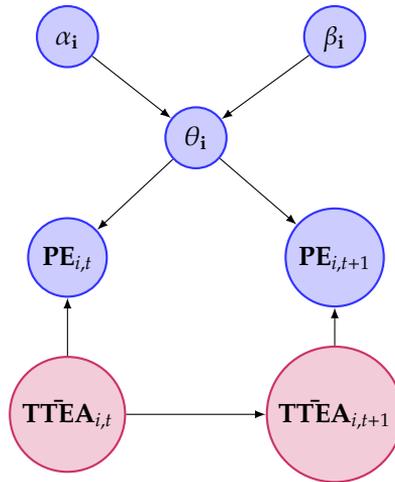
Figure 3.27: Spurious association between PE_t and PE_{t+1}

167

Informative parametrization leads to lower PE_i and θ_i estimates than the non informative one. Standard deviations of estimates are, instead, equal.

AR(1)	NIP	IP
MD	736	891
P	14	13.5
PD	740	904.5

Table 3.63: PE: Binomial model with fixed on θ NIP/IP Deviance

n.iter	PE_i - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{PE}_{A1,2011}$	29.521	4.736	22.000	29.000	37.000
$\hat{PE}_{A2,2011}$	32.167	4.457	25.000	32.000	39.000
$\hat{PE}_{A3,2011}$	49.218	5.604	40.000	49.000	58.000
$\hat{PE}_{A4,2011}$	17.338	3.541	12.000	17.000	23.000
$\hat{PE}_{A5,2011}$	63.049	7.081	52.000	63.000	75.000
$\hat{PE}_{A6,2011}$	78.793	8.525	65.000	79.000	93.000
$\hat{PE}_{A7,2011}$	61.585	6.006	52.000	62.000	71.000
$\hat{PE}_{A8,2011}$	83.982	7.296	72.000	84.000	96.000
$\hat{PE}_{A9,2011}$	91.198	7.247	79.000	91.000	103.000
$\hat{PE}_{A10,2011}$	56.699	6.778	46.000	57.000	68.000
$\hat{PE}_{A11,2011}$	48.716	6.387	38.000	49.000	59.000
$\hat{PE}_{A12,2011}$	40.909	6.046	31.000	41.000	51.000
$\hat{PE}_{A13,2011}$	34.60	5.34	26.00	35.00	44.00
$\hat{PE}_{A14,2011}$	15.671	3.528	10.000	16.000	22.000
$\hat{s}_{A1}$	0.356	0.023	0.319	0.355	0.393
$\hat{s}_{A2}$	0.480	0.026	0.437	0.480	0.524
$\hat{s}_{A3}$	0.464	0.021	0.429	0.464	0.499
$\hat{s}_{A4}$	0.394	0.033	0.341	0.394	0.448
$\hat{s}_{A5}$	0.332	0.015	0.308	0.332	0.357
$\hat{s}_{A6}$	0.215	0.009	0.200	0.21	5 0.231
$\hat{s}_{A7}$	0.505	0.020	0.472	0.505	0.537
$\hat{s}_{A8}$	0.459	0.016	0.433	0.459	0.484
$\hat{s}_{A9}$	0.518	0.017	0.491	0.518	0.545
$\hat{s}_{A10}$	0.318	0.015	0.294	0.318	0.343
$\hat{s}_{A11}$	0.293	0.015	0.269	0.293	0.319
$\hat{s}_{A12}$	0.241	0.014	0.218	0.240	0.264
$\hat{s}_{A13}$	0.306	0.019	0.275	0.306	0.338
$\hat{s}_{A14}$	0.333	0.031	0.283	0.333	0.385

Table 3.64: PE: binomial model with fixed effects on θ NIP estimation

			PE_i - IP		
Node	Mean	Sd	5%	50%	95%
$\hat{\hat{\beta}}_{A1,2011}$	25.573	4.558	18.000	25.000	33.000
$\hat{\hat{\beta}}_{A2,2011}$	30.093	4.414	23.000	30.000	37.000
$\hat{\hat{\beta}}_{A3,2011}$	44.983	5.527	36.000	45.000	54.000
$\hat{\hat{\beta}}_{A4,2011}$	15.32	3.41	10.00	15.00	21.00
$\hat{\hat{\beta}}_{A5,2011}$	54.317	6.751	43.000	54.000	66.000
$\hat{\hat{\beta}}_{A6,2011}$	66.683	8.025	54.000	67.000	80.000
$\hat{\hat{\beta}}_{A7,2011}$	56.432	5.986	47.000	56.000	66.000
$\hat{\hat{\beta}}_{A8,2011}$	78.827	7.262	67.000	79.000	91.000
$\hat{\hat{\beta}}_{A9,2011}$	82.868	7.213	71.000	83.000	95.000
$\hat{\hat{\beta}}_{A10,2011}$	53.296	6.634	43.000	53.000	64.000
$\hat{\hat{\beta}}_{A11,2011}$	44.29	6.15	34.00	44.00	55.00
$\hat{\hat{\beta}}_{A12,2011}$	37.645	5.844	28.000	38.000	47.000
$\hat{\hat{\beta}}_{A13,2011}$	30.027	5.135	22.000	30.000	39.000
$\hat{\hat{\beta}}_{A14,2011}$	12.746	3.324	7.000	13.000	18.000
$\hat{s}_{A1}$	0.308	0.022	0.273	0.308	0.344
$\hat{s}_{A2}$	0.449	0.025	0.407	0.449	0.491
$\hat{s}_{A3}$	0.424	0.021	0.390	0.424	0.459
$\hat{s}_{A4}$	0.348	0.030	0.300	0.348	0.398
$\hat{s}_{A5}$	0.286	0.014	0.263	0.286	0.309
$\hat{s}_{A6}$	0.182	0.009	0.168	0.182	0.196
$\hat{s}_{A7}$	0.463	0.019	0.431	0.463	0.495
$\hat{s}_{A8}$	0.431	0.015	0.406	0.431	0.456
$\hat{s}_{A9}$	0.471	0.016	0.444	0.471	0.498
$\hat{s}_{A10}$	0.299	0.014	0.276	0.299	0.323
$\hat{s}_{A11}$	0.267	0.014	0.244	0.267	0.290
$\hat{s}_{A12}$	0.221	0.013	0.200	0.221	0.243
$\hat{s}_{A13}$	0.266	0.018	0.236	0.265	0.296
$\hat{s}_{A14}$	0.271	0.028	0.226	0.271	0.319

Table 3.65: PE: binomial model with fixed effects on θ IP estimation

Allowing time-varying θ_i leads to the new model:

$$M_{PE}^{TVAR} = \begin{cases} PE_{i,t} \sim Bin(\theta_{i,t}, TTEA_{i,t}), \; i = 1,...14 \; t = 1,2,3 \\ TTEA_{i,t} = PETTEA_{i,t-1} + \varepsilon_{TTEA,t} \\ logit(\theta_{i,t}) = b_{i,t} \\ b_{i,t} \sim N(\mu_{i,t}, \tau_i) \\ \mu_{i,t} = PE_i \mu_{i,t-1} \\ NIP : \{PE, \mu_1\} \sim N(0, 1.0e-6), \; \tau \sim \Gamma(0.001, 0.001) \\ IP : \\ \mu_{1,1} \sim N(-0.71640, 1/0.924^2 * 5), \mu_{1,2} \sim N(-0.29434, 1/0.4880^2 * 5) \\ \mu_{1,3} \sim N(-0.09038, 1/0.9494^2 * 5), \mu_{1,4} \sim N(-0.56008, 1/0.4942^2 * 5) \\ \mu_{1,5} \sim N(-0.5204, 1/0.3423^2 * 5), \mu_{1,6} \sim N(-1.54504, 1/0.7068^2 * 5) \\ \mu_{1,7} \sim N(0.3064, 1/0.6019^2 * 5), \mu_{1,8} \sim N(0.4714, 1/0.6011^2 * 5) \\ \mu_{1,9} \sim N(0.4468, 1/0.6347^2 * 5), \mu_{1,10} \sim N(-0.848284, 1/0.4719^2 * 5) \\ \mu_{1,11} \sim N-1.4991, 1/0.15966^2 * 5), \mu_{1,12} \sim N(-1.49289, 1/0.6784^2 * 5) \\ \mu_{1,13} \sim N(-0.7432, 1/1.0176^2 * 5), \mu_{1,14} \sim N(-2.1786, 1/0.9134^2 * 5) \\ PE_1 \sim N(-0.06296, 1/0.2786^2 * 5), PE_2 \sim N(0.02836, 1/0.1471^2 * 5) \\ PE_3 \sim N(-0.08767, 1/0.2863^2 * 5), PE_4 \sim N(-0.02802, 1/0.1490^2 * 5) \\ PE_5 \sim N(-0.1401, 1/0.1032^2 * 5), PE_6 \sim N(-0.01422, 1/0.2131^2 * 5) \\ PE_7 \sim N(-0.1546, 1/0.1815^2 * 5), PE_8 \sim N(-0.2668, 1/0.1812^2 * 5) \\ PE_9 \sim N(-0.1945, 1/0.1914^2 * 5), PE_{10} \sim N(-0.008475, 1/0.1423^2 * 5) \\ PE_{11} \sim N(0.1569, 1/0.04814^2 * 5), PE_{12} \sim N(0.05472, 1/0.2045^2 * 5) \\ PE_{13} \sim N(-0.1391, 1/0.3068^2 * 5), PE_{14} \sim N(0.3417, 1/0.2754^2 * 5) \\ \tau_1 \sim \Gamma(0.41127, 0.8333), \tau_2 \sim \Gamma(0.11414, 0.8333) \\ \tau_3 \sim \Gamma(0.44013, 0.8333), \tau_4 \sim \Gamma(0.11701, 0.8333) \\ \tau_5 \sim \Gamma(0.08953, 0.8333), \tau_6 \sim \Gamma(0.23686, 0.8333) \\ \tau_7 \sim \Gamma(0.21301, 0.8333), \tau_8 \sim \Gamma(0.29461, 0.8333) \\ \tau_9 \sim \Gamma(0.25644, 0.8333), \tau_{10} \sim \Gamma(0.10556, 0.8333) \\ \tau_{11} \sim \Gamma(0.05483, 0.8333), \tau_{12} \sim \Gamma(0.22310, 0.8333) \\ \tau_{13} \sim \Gamma(0.52391, 0.8333), \tau_{14} \sim \Gamma(0.59772, 0.8333) \end{cases}$$

Informative prior parameters come from non informative priors. The induced DAG is shown on figure 3.28:
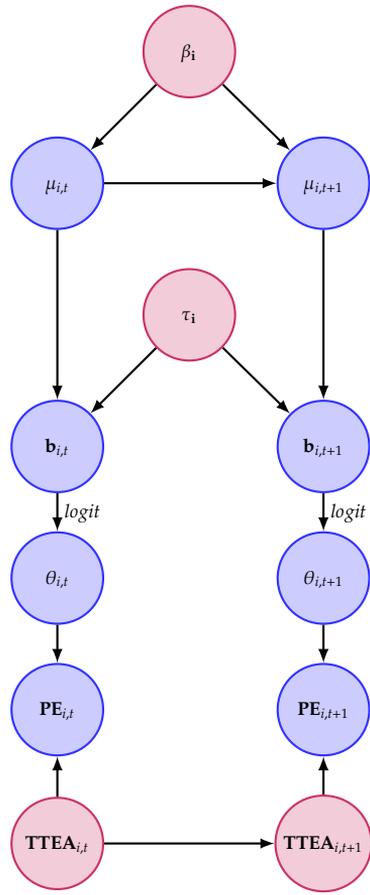
Figure 3.28: Induced DAG from *PE* DLM model

AR(1)	NIP	IP
MD	431	424
P	64.9	66.9
PD	495.9	490.9

Table 3.66: PE: Binomial model with time-varying on θ NIP/IP Deviance

n.iter	PE - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{PE}_{A1,2011}$	32.72	22.42	2.00	30.00	75.00
$\hat{PE}_{A2,2011}$	33.885	8.568	21.000	34.000	48.000
$\hat{PE}_{A3,2011}$	75.12	19.59	36.00	79.00	101.00
$\hat{PE}_{A4,2011}$	16.569	7.552	5.000	16.000	30.000
$\hat{PE}_{A5,2011}$	42.49	25.94	12.00	38.00	91.00
$\hat{PE}_{A6,2011}$	93.71	81.39	7.00	70.00	278.0
$\hat{PE}_{A7,2011}$	59.30	21.74	22.00	59.00	96.00
$\hat{PE}_{A8,2011}$	147.23	61.94	0.00	182.00	183.00
$\hat{PE}_{A9,2011}$	96.28	49.52	6.00	95.00	175.0
$\hat{PE}_{A10,2011}$	61.32	32.86	15.00	57.00	125.00
$\hat{PE}_{A11,2011}$	58.74	11.62	42.00	58.00	76.00
$\hat{PE}_{A12,2011}$	53.69	37.43	6.00	46.00	133.00
$\hat{PE}_{A13,2011}$	39.08	33.09	0.00	30.00	105.00
$\hat{PE}_{A14,2011}$	21.707	8.831	7.000	22.000	37.000
$\hat{\theta}_{A1}$	0.325	0.164	0.106	0.354	0.550
$\hat{\theta}_{A2}$	0.467	0.086	0.318	0.479	0.586
$\hat{\theta}_{A3}$	0.483	0.196	0.209	0.437	0.813
$\hat{\theta}_{A4}$	0.361	0.095	0.223	0.355	0.518
$\hat{\theta}_{A5}$	0.276	0.080	0.147	0.288	0.372
$\hat{\theta}_{A6}$	0.195	0.115	0.076	0.199	0.318
$\hat{\theta}_{A7}$	0.472	0.127	0.287	0.495	0.646
$\hat{\theta}_{A8}$	0.501	0.229	0.166	0.494	1.000
$\hat{\theta}_{A9}$	0.483	0.162	0.222	0.499	0.699
$\hat{\theta}_{A10}$	0.308	0.100	0.183	0.303	0.450
$\hat{\theta}_{A11}$	0.282	0.061	0.181	0.282	0.378
$\hat{\theta}_{A12}$	0.238	0.122	0.097	0.223	0.412
$\hat{\theta}_{A13}$	0.281	0.170	0.072	0.319	0.496
$\hat{\theta}_{A14}$	0.311	0.158	0.053	0.314	0.552

Table 3.67: binomial model with time varying effects on θ NIP estimation

n.iter	PE - NIP				
Node	Mean	Sd	5%	50%	95%
$\hat{\hat{\beta}}_{A1,2011}$	0.540	0.964	-1.107	0.739	1.718
$\hat{\hat{\beta}}_{A2,2011}$	-0.012	0.914	-1.452	0.062	1.162
$\hat{\hat{\beta}}_{A3,2011}$	-1.041	0.728	-1.855	-1.151	0.404
$\hat{\hat{\beta}}_{A4,2011}$	0.794	0.694	-0.546	0.930	1.481
$\hat{\hat{\beta}}_{A5,2011}$	1.125	0.407	0.700	1.149	1.489
$\hat{\hat{\beta}}_{A6,2011}$	0.885	0.488	-0.025	0.963	1.348
$\hat{\hat{\beta}}_{A7,2011}$	-0.053	1.081	-1.394	-0.199	1.610
$\hat{\hat{\beta}}_{A8,2011}$	-3.387	5.894	-11.801	-3.871	6.106
$\hat{\hat{\beta}}_{A9,2011}$	-0.332	3.142	-4.559	-0.186	3.132
$\hat{\hat{\beta}}_{A10,2011}$	0.844	0.603	-0.296	0.942	1.410
$\hat{\hat{\beta}}_{A11,2011}$	0.840	0.107	0.732	0.848	0.951
$\hat{\hat{\beta}}_{A12,2011}$	0.803	0.521	-0.250	0.900	1.303
$\hat{\hat{\beta}}_{A13,2011}$	0.666	1.044	-1.035	0.876	1.830
$\hat{\hat{\beta}}_{A14,2011}$	0.435	0.402	-0.217	0.486	0.892
$\hat{s}_{A1}$	1.175	0.711	0.485	1.008	2.391
$\hat{s}_{A2}$	0.271	0.235	0.041	0.216	0.686
$\hat{s}_{A3}$	0.422	0.386	0.062	0.324	1.090
$\hat{s}_{A4}$	0.429	0.415	0.045	0.321	1.160
$\hat{s}_{A5}$	0.360	0.338	0.069	0.273	0.949
$\hat{s}_{A6}$	0.992	0.719	0.392	0.783	2.270
$\hat{s}_{A7}$	0.599	0.348	0.259	0.519	1.194
$\hat{s}_{A8}$	0.210	0.284	0.030	0.110	0.752
$\hat{s}_{A9}$	0.611	0.397	0.177	0.536	1.281
$\hat{s}_{A10}$	0.618	0.450	0.216	0.496	1.407
$\hat{s}_{A11}$	0.132	0.158	0.028	0.088	0.365
$\hat{s}_{A12}$	0.923	0.653	0.349	0.738	2.087
$\hat{s}_{A13}$	1.354	0.833	0.553	1.151	2.804
$\hat{s}_{A14}$	0.709	0.566	0.124	0.582	1.680

Table 3.68: binomial model with time varying effects on θ NIP estimation

n.iter	PE - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{PE}_{A1,2011}$	41.55	21.08	7.00	42.00	76.00
$\hat{PE}_{A2,2011}$	33.53	12.44	13.00	34.00	54.00
$\hat{PE}_{A3,2011}$	53.07	22.56	16.00	53.00	91.00
$\hat{PE}_{A4,2011}$	21.998	9.864	6.000	22.000	38.000
$\hat{PE}_{A5,2011}$	94.95	46.55	19.00	95.00	171.00
$\hat{PE}_{A6,2011}$	183.2	108.8	15.0	183.0	351.1
$\hat{PE}_{A7,2011}$	60.98	23.07	22.00	61.00	99.00
$\hat{PE}_{A8,2011}$	91.42	38.35	28.00	91.00	155.00
$\hat{PE}_{A9,2011}$	88.01	35.07	29.00	88.00	147.00
$\hat{PE}_{A10,2011}$	88.91	41.97	20.00	89.00	158.00
$\hat{PE}_{A11,2011}$	83.00	37.12	21.00	83.00	145.00
$\hat{PE}_{A12,2011}$	85.15	46.71	10.00	85.00	160.00
$\hat{PE}_{A13,2011}$	56.62	31.03	7.00	57.00	106.00
$\hat{PE}_{A14,2011}$	23.178	9.498	7.000	23.000	39.000
$\hat{\theta}_{A1}$	0.345	0.173	0.114	0.369	0.655
$\hat{\theta}_{A2}$	0.461	0.116	0.289	0.464	0.638
$\hat{\theta}_{A3}$	0.438	0.173	0.191	0.409	0.682
$\hat{\theta}_{A4}$	0.385	0.136	0.207	0.364	0.633
$\hat{\theta}_{A5}$	0.324	0.141	0.162	0.309	0.647
$\hat{\theta}_{A6}$	0.236	0.182	0.080	0.211	0.709
$\hat{\theta}_{A7}$	0.472	0.134	0.278	0.499	0.657
$\hat{\theta}_{A8}$	0.443	0.144	0.174	0.477	0.618
$\hat{\theta}_{A9}$	0.477	0.140	0.226	0.498	0.663
$\hat{\theta}_{A10}$	0.336	0.139	0.184	0.314	0.637
$\hat{\theta}_{A11}$	0.309	0.135	0.160	0.287	0.627
$\hat{\theta}_{A12}$	0.270	0.171	0.097	0.232	0.680
$\hat{\theta}_{A13}$	0.309	0.184	0.084	0.336	0.680
$\hat{\theta}_{A14}$	0.317	0.169	0.056	0.309	0.603

Table 3.69: binomial model with time varying effects on θ IP estimation

n.iter	PE - IP				
Node	Mean	Sd	5%	50%	95%
$\hat{\hat{\beta}}_{A1,2011}$	-0.049	0.125	-0.255	-0.049	0.157
$\hat{\hat{\beta}}_{A2,2011}$	0.029	0.066	-0.079	0.029	0.137
$\hat{\hat{\beta}}_{A3,2011}$	-0.091	0.128	-0.301	-0.091	0.121
$\hat{\hat{\beta}}_{A4,2011}$	-0.025	0.067	-0.135	-0.025	0.084
$\hat{\hat{\beta}}_{A5,2011}$	-0.139	0.046	-0.215	-0.139	-0.063
$\hat{\hat{\beta}}_{A6,2011}$	-0.001	0.096	-0.159	-0.001	0.156
$\hat{\hat{\beta}}_{A7,2011}$	-0.158	0.081	-0.291	-0.157	-0.025
$\hat{\hat{\beta}}_{A8,2011}$	-0.266	0.081	-0.399	-0.266	-0.133
$\hat{\hat{\beta}}_{A9,2011}$	-0.195	0.085	-0.335	-0.195	-0.055
$\hat{\hat{\beta}}_{A10,2011}$	-0.004	0.064	-0.109	-0.004	0.101
$\hat{\hat{\beta}}_{A11,2011}$	0.158	0.022	0.122	0.158	0.194
$\hat{\hat{\beta}}_{A12,2011}$	0.069	0.092	-0.082	0.069	0.221
$\hat{\hat{\beta}}_{A13,2011}$	-0.124	0.138	-0.351	-0.124	0.102
$\hat{\hat{\beta}}_{A14,2011}$	0.372	0.119	0.175	0.373	0.565
$\hat{s}_{A1}$	1.347	0.522	0.774	1.233	2.293
$\hat{s}_{A2}$	0.800	0.332	0.450	0.725	1.396
$\hat{s}_{A3}$	1.009	0.387	0.582	0.924	1.720
$\hat{s}_{A4}$	1.052	0.456	0.573	0.949	1.863
$\hat{s}_{A5}$	1.289	0.532	0.730	1.166	2.253
$\hat{s}_{A6}$	1.867	0.745	1.076	1.699	3.202
$\hat{s}_{A7}$	0.850	0.340	0.485	0.773	1.467
$\hat{s}_{A8}$	0.999	0.393	0.575	0.911	1.709
$\hat{s}_{A9}$	0.926	0.368	0.530	0.842	1.595
$\hat{s}_{A10}$	1.211	0.498	0.684	1.096	2.120
$\hat{s}_{A11}$	1.109	0.466	0.623	1.000	1.947
$\hat{s}_{A12}$	1.587	0.635	0.908	1.445	2.735
$\hat{s}_{A13}$	1.568	0.586	0.916	1.442	2.631
$\hat{s}_{A14}$	0.886	0.341	0.502	0.812	1.515

Table 3.70: binomial model with time varying effects on θ IP estimation

PE_i estimates are quite different fro the fixed model.

Two year forecasting: B1 2012 prevision

Now every quantity needed to make an estimation on B1 2012 is known.
For S and MR_i estimations the same values of 2011 are used:
S=52165 [49282, 55125]
MR_{mean}={42.99, 51.09, 55.00, 48.93, 49.04, 27.05, 49.95, 51.06, 53.04, 35.93, 37.05, 30.98, 33.03, 33.95}
MR_{25pc}={39.46, 47.57, 51.59, 46.45, 45.59, 23.56, 46.44, 47.57, 49.62, 32.39, 33.51, 27.49, 29.45, 30.45}
MR_{75pc}={46.43, 54.65, 58.52, 52.40, 52.46, 30.51, 53.50, 54.53, 56.56, 39.55, 40.54, 34.43, 36.61, 37.47}
For PE_{2010} estimation the time-varying binomial model is used:
PE_{mean}={47,33,60,20,74,122,59,115,88,70,69,60,47,23}
PE_{5pc}={8,25,46,6,47,31,41,22,65,29,47,15,7,16}
PE_{95pc}={70,42,74,34,103,256,76,177,112,119,91,123,97,30}
These PE_{2010} must be averaged out with years 2007-2009 before using for
$B1$ computation.

$$B1_{2012}=3.6 \ [2.91, \ 4.15]$$

Compared to $B1$ 2010, 3.28, there could be a gain of 532.22M€[-615.38,1446.98].
This countertrend result could be explained by the high variability of PE_{2010}
estimates, responsible for the generalized increases (PE(10)-PE(06) is almost
always positive) and, in particular, for abnormal value in area 6, as shown
in Table 3.71.

Area	PE_{05-08}	PE_{06-09}	PE_{07-10}	$\Delta_{2-1}\%$	$\Delta_{3-2}\%$	PE(09)-PE(05)	PE(10)-PE(06)
01	30.25	26	33.37	-14.38	22.96	-19	21.4
02	33.75	33.25	33.13	-1.48	-0.36	-1	-0.9
03	52.75	51	47.50	-3.32	-5.26	-9	-19.5
04	15.75	15.25	17.76	-3.17	14.13	-2	5.8
05	60.75	55	63.03	-9.47	12.74	-6	0.0
06	79	69.75	106.60	-11.70	34.57	-38	70.6
07	64	55.75	60.51	-12.89	7.87	-34	18.5
08	101.25	86	83.87	-15.06	2.54	-63	-16.1
09	96.5	85.75	85.77	-11.14	0.02	-41	-2.2
10	61	57	69.04	-6.56	17.44	-16	28.0
11	45.5	51.25	59.95	12.64	14.51	24	12.0
12	45.5	41.75	57.43	-8.24	27.30	-17	35.4
13	35	28.25	39.39	-19.29	28.28	-27	27.4
14	13	15	17.78	15.38	15.64	7	5.8

Table 3.71: 3 year PE values

3.14   B1 Causal analysis

Miming statistics performed on $A2$, this section aims to answer two questions: what variable does it need to intervene upon and how much time does it take to move up to the median position in the ranking?

As it has been already mentioned, the indicator does not depend directly on $TTEA$, but only indirectly because $TTEA$ is a parameter in the function modelling $PE \sim Bin(\theta, TTEA)$. The crucial variable to intervene on is R, namely the ratio between positively evaluated researcher and the total number of researcher (by area), that is on a quantity representing the efficiency of the system. Notice that being TEA fixed, an intervention on R is equivalent to an intervention of PE because of the relation $PE = R * TTEA$.

Let us getting started searching weak R performances and after making the plane.

3.14.1 searching weak R performances

Table 3.72 shows several statistics on R: the difference, absolute and as a fraction of R, between R and the median MR on 2010. After R 2011 is compared with R 2010 and with MR_{2010} showing both temporal dynamics and benchmark comparison. Unfortunately all 12 areas are under the median value.

Area	R_{2010}	MR_{2010}	$1-2$	$\frac{1-2}{1}\%$	R_{2011}	$5-1$	$\frac{5-1}{1}\%$	5-2	$\frac{5-2}{5}\%$
01	36.2	43	-6.8	-18.8	29.0	-7.2	-19.9	-14	-48.3
02	49.5	51	-1.5	-3.0	47.5	-2.0	-4.0	-3.5	-7.1
03	48.8	55	-6.2	-12.7	45.4	-3.4	-7.0	-9.6	-19.7
04	36.2	49	-12.8	-35.4	34.1	-2.1	-5.8	-14.9	-43.7
05	32.5	49	-16.5	-50.8	27.7	-4.8	-14.8	-21.3	-65.5
06	20.9	27	-6.1	-29.2	17.6	-3.3	-15.8	-3.3	-15.8
07	52.1	50	2.1	4.0	43.0	-9.1	-17.5	-7	-14
08	50.9	51	-0.1	-0.2	41.6	-9.3	-18.3	-9.4	-22.6
09	54.1	53	1.1	+2.0	46.7	-7.4	-13.7	-6.3	-13.5
10	33.8	36	-2.2	-6.5	29.6	-4.2	-12.4	-6.4	-21.6
11	27.0	37	-10	-37.0	28.8	1.8	6.3	-8.2	-28.5
12	26.4	31	-4.6	-17.4	21.8	-4.6	-17.4	-9.2	-42.2
13	32.9	33	-0.1	-0.3	24.4	-8.5	-25.8	-8.6	-35.2
14	28.6	34	-5.4	-18.9	31.9	3.3	11.5	-2.1	-6.6

Table 3.72: R_{2010}, R_{2011}, MR_{2010}, differences

Next statistics shows the percentage of researchers, by area, with no positive evaluations in the 2006-2009 period. Results are discouraging.

area	% no PE
AREA 01	35.7
AREA 02	20.8
AREA 03	14.6
AREA 04	18.2
AREA 05	34.0
AREA 06	52.6
AREA 07	11.6
AREA 08	16.2
AREA 09	6.5
AREA 10	36.2
AREA 11	32.3
AREA 12	35.5
AREA 13	38.9
AREA 14	26.1

Table 3.73: % researchers with no *PE* in 2006-09 per area

Table 3.74 performs a gender analysis. Last columns shows female prevalence.

Area	R_F	R_M	$|R_F - R_M|$	$100 * \frac{|R_F - R_M|}{R}$	R	% F
AREA 01	26.95	30.73	3.78	13.05	28.97	46.5
AREA 02	36.00	51.71	15.71	33.07	47.50	26.8
AREA 03	42.78	46.47	3.69	8.12	45.43	40.1
AREA 04	36.36	33.76	2.60	7.63	34.08	12.3
AREA 05	26.72	28.61	1.89	6.82	27.71	47.6
AREA 06	25.07	15.26	9.81	55.74	17.60	23.9
AREA 07	39.29	43.49	4.20	9.77	42.97	21.6
AREA 08	44.26	40.31	3.95	9.50	41.60	29.5
AREA 09	54.24	46.07	8.17	17.48	46.73	8.0
AREA 10	29.31	29.64	0.33	1.12	29.57	60.2
AREA 11	26.53	30.54	4.01	13.95	28.75	48.1
AREA 12	22.27	21.55	0.72	3.31	21.77	31.0
AREA 13	28.99	22.46	6.53	26.75	24.41	29.8
AREA 14	38.71	27.78	10.93	34.25	31.91	33.0

Table 3.74: % researchers with no *PE* in 2006-09 per area

3.14.2 Intervening on R

From the previous section it emerges a non encouraging situation, *R* values are in general lower than the national median. Figure 3.18 shows *R* variability among departments.
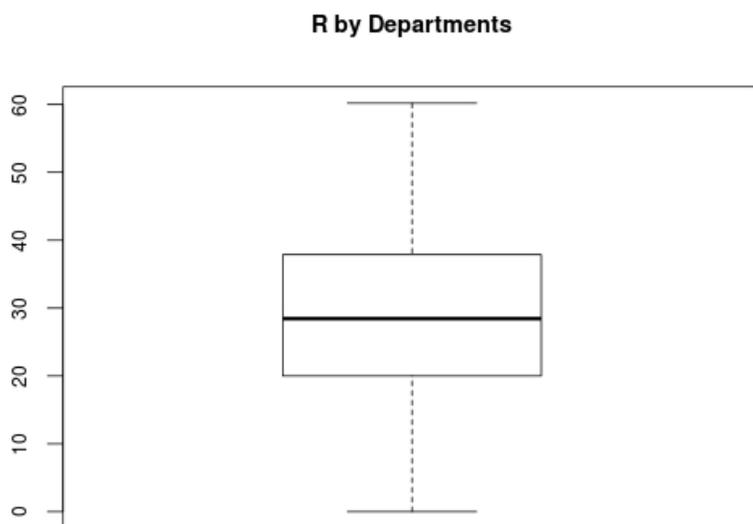
Figure 3.29: R_{B1} by Departments

No department goes over 60% of researcher positively evaluated.
Finally, these are the number of positive evaluated researchers by area needed to get the equality:

Area	1	2	3	4	5	6	7	8	9	10	11	12	13	14
PE	6	2	7	6	31	24	0	1	0	4	17	8	1	3

Table 3.75: *PE* needed to get $R \geq MR$

Given this status, one aim could be setting a policy to reach, in an limited

183

number of years, the median value MR of the area. Thus each year, every degree course compares R and MR and according to the difference defines its new target T, namely the new value it must reach:

C1) if $R \geq MR$ then $T \geq R$;

C2) if $0.95MR \leq R < MR$ then $T \geq MR$;

C3) if $0.90MR \leq R < 0.95MR$ then $T \geq 1.05R$;

C4) if $R < 0.90MR$ then $T \geq 1.1R$;

The rules above are interpreted as follows: when the ratio is higher than the median, no effort is required, except not reducing its value. When the ratio is less than the median but not more than 5%, the effort required is to catch up the median. When the ratio is between 90% and 95% then degree courses have to increase it at least by 5%. Finally, when the ratio is smaller than the median by 90%, the ratio must be increased by at least by 10%. This procedure, if satisfied, reaches the goal in no more than three years. The table 3.76 shows for three consecutive years the percentage of degree courses in each class and the value of $B1$[30]:

Year	C1	C2	C3	C4	A2
2011	0	2	9	3	3.02
2012	2	3	8	1	3.34
2013	5	7	2	0	3.63
2013	12	1	1	0	3.84
2014	13	1	0	0	3.96

Table 3.76: Percentage of target classes degree courses by area and year

The counterfactual value (What $B1$ 2011 would be had the local ratio at least equal to the national median?) is 3.97. Differently from $A2$ case, where

---

[30]2010 and 2011 got a different number of researchers; 2011-13 assume it constant. MR_i and S are assumed constant.

some units (Degree Courses in that case) showed values of R in the C1 class in the first year, allowing some other units to remain in other classes, the $B1$ case showed all 14 areas under the median value, so to reach it all 14 areas needed to get the C1 class.

The induced DAG is shown in the next figure:
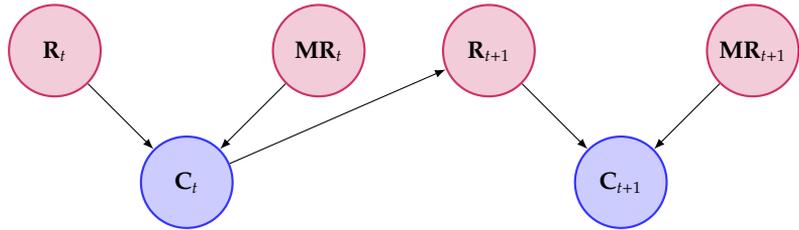
Figure 3.30: Act on EFF policy DAG

Chapter 4

Conclusion and Future work

This thesis turned around the meaning of the word "monitoring" FFO indicators. This word has been assigned 4 separate and interconnected meanings: describe, retrospectively assess, predict and act.

To describe means using elementary statistical tools such as tables and graphs to describe the indicator factors in order to take a picture of the state of the art, past and present, and compare local data with national median values.

Retrospective assessment means using counterfactuals in the simple features of what/if question in order to discover errors or weakness and prepare correctives for the future.

To make a prediction means to use Bayesian forecasting techniques in order to give both point and interval estimates up to two years ahead to indicator factors. One year forecasts are supposed to be more reliable estimates because local variable were already known, so prediction regarded only two national variables. Actually one or few observations available were a poor evidence to establish the degree of stability of those variables. The two year forecasting was a more difficult task because both local and national variable were forecast increasing global uncertainty.

Unlike physics or natural sciences, whose purpose is the discovery of laws, either deterministic or stochastic, external to human beings and in some sense eternal, the regularities observed in the social sciences are the result of human actions, internal to the point of view of the scientist as human behaviour and mutable, so the art to predicting future values deserves a different paradigm: laws determining a certain outcome can change rapidly due to modification in behaviours and often are results of several conflicting behaviours of which observed values are only averages.

Much more attention has to be paid to the counterfactual interpretation of a prediction. A prediction can be summarised as: given what has been observed, e.g. a linear trend, and given some assumptions, e.g. linearity will still hold in the future, a prevision for Y is y. Now suppose to take an action A aiming to increase the value of Y, and to observe $Y_A = 1.1y$. Normally the causal effect of A to Y is computed as $\frac{Y_A}{Y} = 1.1$, but this effect relies on the unobserved (counterfactual) variable $Y = Y_{\bar{A}}$, namely the value of Y had action A not taken, which coincides with the pure prevision of Y.

Turning back to technicalities, particularly difficult was proven the task of choosing parameters of informative prior distributions, often guided by intuition, repeated trials or with tricks as to use the empirical data. But this difficult was overcome because priors are the best way to make models flexible (even to external interventions) by putting into them extra knowledge.

Another difficulty encountered has been to think constantly about the nature of relationships between variables in terms not only associative but mainly causal, that is constantly asking "who is the cause and who is the effect". This explained the several times repeated phrase "spurious association between X and Y" stressing that beyond statistical significance, the relation was the result of a spurious correlation due to (at least) a third unobserved variable interpreted as a common cause. But it also helped greatly in clarifying the role played by variables in relationships, such as causal or mediator or confounder. This drives directly to the task of acting,

namely thinking about what changes and what remains invariant when a decision is taken, an intervention is performed and a shock into the (pre-intervention) relationships is imposed.

Despite several problems arose, good solutions were taken, but only time will tell how good they were.

If writing this thesis brought me light about which levers need to be pulled in order to build statistical and causal models for monitoring the University of Palermo Ordinary financing Fund, even more dubts and new ideas occurred. Several future works are in mind and in process, including:

- implementing such monitoring models in the university ICT system through client/server applications. Statistical models are useful but often are matters for initiates, nowadays software engineering paradigm may help in separate what can be called the "interface", and what can be called the "engine". The latter is represented by the statistical models, usually running in the background, completely transparent to the final user. The former, instead, is usually represented by the output of the statistical models, such as point or interval estimate, provided in a user-friendly web page;

- so far only variables prescribed by the national set of indicators were used. The job was: act on factors and see the results on the indicator. Now a second level analysis is needed: to build regressive models on the factors, as $EFF = f(X, Y, Z, \varepsilon)$ or $PE = g(T, K, F, u)$. This may help to make more precise previsions on the indicators and to discover inter-indicator common cause variables;

- generalize the do() operator, see pag. 49;

- it is a matter of the higher importance revisiting both parameters or functional forms of variables and parent/child relationships on the basis of new facts, such as new national laws or local decree or an

unexpected increase in the rate of retirement due to new national retirement policy, or a decrease in new enrolments due to worse economical perspective. All this can be done improving elicitation techniques on the prior parameters on the basis of meetings with experts or documentary researches.

Exams never end.

Appendix A

Historical pathway of the word Cause

A.0.3 Causality in the Antiquity

The etimology of the ancient greek word $\alpha\iota\tau\iota\alpha$ or $\alpha\iota\tau\epsilon\omega$ can be translate in "to accuse"or "to ask", showing the primitive meaning was related with "volontary action which whom the autor is responsable "(Guzzo and Barone, 1957).[1].

The first greek philosopher that spoke about cause was the atomist Leucippo, giving a definition of cause as necessary purely physics connection between empirical facts, as mechanicistic causes. This definition circumscribes cause to physics, only from Newton onwards cause will include forces between remote objects.

To Plato "everything that arises must be born from a cause, it is impossible that anything is born without a cause"(Timeo 28A). He distinguished between first and second (cooperating) casues. Relationships were considered material.

---

[1]Nowdays that term is used in law to indicate "contract ", or "inheritance "(causa mortis)

The first attempt to systematize the cause is due to Aristotle and peripatetic schoolmen, throught the famouses four causes (Bunge (1959) pag 32):

1. the material cause (to be essence of something), which provides the passive receptacle on which the remaining casuses act;

2. the formal cause (when the movement begins), which contributes the essence, idea, or quality of the thing concerned;

3. the efficient cause (material and substrate), that is the external compulsion that bodies had to obey;

4. the final cause (that for which it acts, the Good), the goal to which everything strove and which everything served.

The first two were causes of being, the last two of becoming. For Aristotle to go back to formal cause is the most important task because it exists a link between cause and demonstration: knowing a fact means showing a proof of it starting from initial principles (causes).
Stoics and Epicureans link causal determinism with liberty of moral life. For Stoics causation is equate to fate, causality is perceived as purely mechanical and physics.
Skeptics did criticize causation.
For Neoplatonists nothing is uncaused, they focused on the cause as product of its own act, *causa sui*.

A.0.4   Causality in The Middle Ages

Patristic and Scholasticism cleared all the material aspects from the cause beacuse material was perceived as expression of Evil (Origen).
For Saint Augustine God is *causa sui*, material objects have been caused by the *rationes aeterne* of God.
With Scholasticism cause reached more and more adjectives: appropriate,

being, ending, immanent, transient, di per se, per accidents, occasional and so on.

It insisted on the substantial forms and consequently on the immanet cause (opposed to transitive).

Saint Thomas did inculturate sacred theology in the Aristotelian philosophy, for him causation is realization of potentiality, God as first cause. Contrary to Neoplatonisms he developped a demonstrative method going from effects to causes.

Ockam is critical: the Scholasticism axiom according to which the sentence "the more efficient the more universal the cause is" is rejected by the nominalistic theory that predicated universal is not real and there needs an empirical knowledge (*notitia intuitiva*) connecting causes and effects.

A.0.5 Causality in The Modern Age

When Modern Age became, it riduced the problem of efficient causality on the naturalistc plane, formal and final causes were banished to theological field, beacuse of impossibility to submit them to experiment. Instead the efficient cause dominate because it seemed to have the necessary requirements to the current paradigma: it is related to the idea of changing, providing something from something else, the possibility to explain it mathematically, and to controll it by an experiment. The definitive turning point was due to the milenstone works of Galileo and in particular to the central role of the experiment as synthesis between "sensible experiences" ans "necessary demonstrations", in which temporal aspects and necessary causation are joined and results are answers of the Nature to human mind activity. The Galileo's definition of cause prescibed cause is something necessary and sufficient in order that somethig else be, if cause is then effect will be, if cause is notthen effect never will (Galilei, 1623). It is a methodological definition: let us add a cause to have an effect and let us remove a

cause to remove the effect.

For Descartes causation is an eternal truth placed in the thought.

For Spinosa every thing is produced by substance and, as finite thing, by a finite modification of his attributes, then it is transient. Leibniz riduced causation to principle of sufficient reason (monadology) not demanding proof or experiment and oriented to moral better (and not for logic necessity).

Bacon and Hobbes drawn attention on inductive searching of causes.

For Locke the ideas of cause and effect come from regularity of a sequence of events perceived through senses or mediation (Locke, 1847).

Hume refused the a priority and analyticity of cause and focused on his empirical charateristic and on the constancy of sequence. Hume's definition of cause is (Hume, 1967): "We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed".

This definition encompasses a relation between two objects, ordering time is supposed (the cause must precede th effect)[2], similarity means generalization, not only a sequence of two "particular" events but classes of events, and the last point describes the necessary condition of causation: removing the cause entails removing the effect. The formers defines the regular theory of causation, the last defines the counterfactual theory of causation. Kant interpreted cause as synthetic a priori judgment (Kant, 1929).

Mill resumed empiricism tradition connecting closely cause and induction, the latter is applicable assuming the former and the former is acceptable on the base on the induction by simple enumeration. Descartes claimed that sensations, such as taste or temperature, are caused by the shape and size of tiny pieces of matter.

---

[2]the axioms of Peano explicitely speak about "successor"

A.0.6 Causality in The Contemporary Age

The XIX century was dominated by classical physics and it defined causal process as knowning its actual state and all the forces that affect it, it is possible to infer its future state. The famous Laplace's Demon: "We may regard the present state of the universe as the effect of its past and the cause of its future. An intellect which at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, if this intellect were also vast enough to submit these data to analysis, it would embrace in a single formula the movements of the greatest bodies of the universe and those of the tiniest atom; for such an intellect nothing would be uncertain and the future just like the past would be present before its eyes"(de Laplace, 1829).

That priciple was extended to other filed of knowledge as Darwinism and logic Positivism: natural laws are laws of real.

Mach proposed a clearing of the philosophycal concept of cause by substitution with the mathematical object of function.

On the same line Russell: scientific law is dominated by the idea of differential equation.

The theory of relativity and the quantum mechanism dictated a real revolution both of classic mechanism and concept of cause. In particular the uncertainty principle of Heisenberg expressed the inapplicability of the determinist criterion of causality to microscopic phenomena.

The contemporary age started with a sort of rejection of the cause, which can be spotted as scientists against causation:

B. Russell: "The law of causality, I believe, like much that passes muster along philosopher, is a relic of a bygobe age, surviving like the monarchy only because it is erroneously supposed to do no har". (Russell, 1912).

For Pearson causality is an obsolete metapysical idea and it has to be replaced with the most general and specid tool as correlation (Pearson, 1892).

Galton (1889):"there was a category borader than causation, namely corre-

lation, of which causation was only the limit".

For Granger the main conditions for defining causality are temporal priority (cause occurs before effect) and that cause need to contain information about the effect that is not contained in any aother event occurring not later than itself (Granger, 1980).

Bunge criticises Hume's definiton of causality and the claim to be the only category of determination (Bunge, 1959).

In the "A probabilistic Theory of Causality", Suppes (1970) transferred Hume's definition in a pure probabilistic language. Other contemporaries dealt with causation, (Dawid, 2000), (Rubin, 2005), (Spirtes *et al.*, 2001), (Cartwright, 2007), I like to end with my favorite, Dr. Judea Pearl who with visionary and pioneer works on structural equations, graphical models and countefactuals Pearl (2009) gave a new semantics, formalism and lifeblood to the word cause.

Finally I would like to underline that the pathway of the word cause followed the pattern as our culture, and society in general, passed from pole of ontological-immanent-metaphysical to the pole of transient-empirical categories.

Bibliography

Balke, A. and Pearl, J. (1994). Probabilistic evaluation of counterfactual queries. In *Proceedings of the National Conference on Artificial Intelligence*, pages 230–230. JOHN WILEY & SONS LTD.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. of the Royal Soc. of London*, **53**, 370–418.

Bernardo, J. M. and Smith, A. (2000). *Bayesian Theory (Wiley Series in Probability and Statistics)*. Wiley, 1 edition.

Bollen, K. (1989). *Structural equations with latent variables*, volume 8. Wiley New York.

Box, G. and Tiao, G. (1992). *Bayesian inference in statistical analysis*. Wiley Online Library.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control (Revised Edition)*. Holden-Day, revised edition.

Brockwell, P. J. and richard A. Davis (2002). *Introduction to Time Series and Forecasting*. Springer, New York.

Bunge, M. (1959). *Causality*. Harvard University Press.

Campagnoli, P., Petrone, S., Petris, G., Petris, G., Petrone, S., and Campagnoli, P. (2009). Dynamic linear models dynamic linear models with r. In

Dynamic Linear Models with R, Use R, chapter 2, pages 31–84. Springer New York, New York, NY.

Cartwright, N. (2007). *Hunting causes and using them: approaches in philosophy and economics*. Cambridge Univ Pr.

Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons Inc.

Dawid, A. (2000). Causal inference without counterfactuals. *Journal of the American Statistical Association*, pages 407–424.

de Finetti, B. (1989). Probabilism. *Erkenntnis*, **31**(2), 169–223.

de Finetti, B. (1990). *Theory of probability. Vol. 1-2*. John Wiley & Sons Ltd., Chichester.

de Laplace, P. S. (1829). *Essai philosophique sur les probabilités*. H. Remy.

Galavotti, M. C. (2005). *A Philosophical Introduction to Probability (Center for the Study of Language and Information - Lecture Notes)*. Center for the Study of Language and Inf, 1 edition.

Galilei, G. (1623). Il saggiatore (roma, 1623). *Italian translation by L. Sosio (Milano, Feltrinelli, 1979)*.

Galton, S. (1889). *Natural inheritance*, volume 42. Macmillan and co.

Good, I. J. (1965). *The Estimation of Probabilities*. MIT Press, Cambridge, MA.

Granger, C. (1980). Testing for causality:: A personal viewpoint. *Journal of Economic Dynamics and Control*, **2**, 329–352.

Guzzo, A. and Barone, F. (1957). Causa. In *Enciclopedia Filosofica*, volume I, pages 958–975. Sansoni, Firenze.

Harrison, P. J. and West, M. (1991). Dynamic linear model diagnostics. *Biometrika*, **78**, 797–808.

Hume, D. (1967). A treatise of human nature (1740). *A more pertinent reference, in this author's opinion, is provided in*, **4**, 35.

Hume, D. (1976). Of the idea of necessary connection. *The nature of causation, edited by Myles Brand. Urbana: University of Illinois Press. Reprint of An inquiry concerning human understanding, section vii.*

Johnson, N., Kotz, S., and Balakrishnan, N. (1994). Continuous univariate distributions, vol. 1.

Johnson, N., Kemp, A., and Kotz, S. (2005). *Univariate discrete distributions*, volume 444. Wiley-Interscience.

Kant, I. (1929). Critique of pure reason. *Norman Kemp Smith (New York: St. Martinâs Press, 1965)*, **323**.

Kolmogorov, A. (1992). *Selected Works of A.N. Kolmogorov: vol. 2 Probability Theory and Mathematical Statistics*. Springer, 1 edition.

Lee, S.-Y. (2007). *Structural equatio modeling: a bayesian approach*. John Wiley & Sons, Chichester.

Lewis, D. (1973). *Counterfactuals*. Blackwell Publishers, Oxford.

Lindley, D. (2010). *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference*. CUP Archive.

Locke, J. (1847). *An essay concerning human understanding*. Troutman & Hayes.

O'Hagan, A. (2006). *Uncertain judgements : eliciting experts' probabilities*. Wiley.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Dystems: Networks of Plausible Inference*. Morgan Kaufmann, 1 edition.

Pearl, J. (1995). Casual diagrams for empirical research. *Bometrika*.

Pearl, J. (1996). Structural and probabilistic causality. In D. R. Shanks, K. J. Holyoak, and D. L. Medin, editors, *The psychology of learning and motivation: Causal learning*, volume 34. Academic Press, San Diego.

Pearl, J. (2000). Comment: Graphical models, causality, and intervention. *Journal of the American Statistical Association*, **95**, 428–431.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*, pages 411–420.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, New York, NY, USA.

Pearl, J. (2012). The causal mediation formula – a guide to the assessment of pathways and mechanisms. *Prevention Science*.

Pearl, J. and Bareinboim, E. (2011). Transportability of causal and statistical relations: A formal approach. In M. Spiliopoulou, H. Wang, D. J. Cook, J. Pei, W. W. 0010, O. R. Zaane, and X. Wu, editors, *ICDM Workshops*, pages 540–547. IEEE.

Pearson, K. (1892). *The grammar of science*, volume 20. W. Scott.

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, **9**(3), 523–539.

Plummer, M. (2011). *rjags: Bayesian graphical models using MCMC*. R package version 3-5.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raiffa, H. and Schlaifer, R. (1968). *Applied statistical decision theory*.

Roberts, C. (1994). The bayesian choice: A decision-theoretic motivation".

Robins, J. (1987). Addendum toa new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Comput Math Appl*, **14**(9-12), 923–945.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, **66**(5), 688.

Rubin, D. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *JASA*.

Russell, B. (1912). On the notion of cause. In *Proceedings of the Aristotelian society*, volume 13, pages 1–26. JSTOR.

Salmon and Wesley, C. (1998). Probabilistic causality. *Causality and Explanation*, pages 208–233.

Shpitser, I. and Pearl, J. (2007). What counterfactuals can be tested. In R. Parr and L. C. van der Gaag, editors, *UAI*, pages 352–359. AUAI Press.

Shpitser, I. and Pearl, J. (2008). Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research*, **9**, 1941–1979.

Shumway, R. H. and Stoffer, D. S. (2006). *Time series analysis and its applications: with R examples*. Springer, New York.

Spirtes, P., Glymour, C., and Scheines, R. (2001). *Causation, Prediction, and Search*. The MIT Press, Cambridge, MA, USA, second edition.

Strotz, R. and Wold, H. (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, pages 417–427.

Suppes, P. (1970). *A probabilistic theory of causality*. North Holland, Amsterdam.

Tian, J. and Pearl, J. (2002). A general identification condition for causal effects. In R. Dechter and R. S. Sutton, editors, *AAAI/IAAI*, pages 567–573. AAAI Press / The MIT Press.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Series in Statistics. Springer, second edition.

Wright, S. (1921). Correlation and causation.