Dottorato in Scienze Economiche e Statistiche
Dipartimento di Scienze Economiche, Aziendali e Statistiche
SECS-S/01 - Statistica

LONGITUDINAL MEDIATION ANALYSIS WITH STRUCTURAL AND MULTILEVEL MODELS: ASSOCIATIONAL AND CAUSAL PERSPECTIVES

IL DOTTORE
CHIARA DI MARIA

IL COORDINATORE
ANDREA CONSIGLIO

IL TUTOR
GIANFRANCO LOVISON

IL CO-TUTOR
ANTONINO ABBRUZZO

CICLO XXXIV
ANNO CONSEGUIMENTO TITOLO 2022

Longitudinal mediation analysis with structural and multilevel models: associational and causal perspectives

Chiara Di Maria
PhD in Economics and Statistics
Department of Economics, Business and Statistics
University of Palermo
2022

Abstract

In many fields of science, it is crucial to monitor and understand the evolution of phenomena over time. Mediation analysis is widely used for understanding how the effect of an exposure on a response propagates, and longitudinal data are acknowledged to be the most suited to answer mediational questions, since they take into account the dynamic nature of most phenomena. Indeed, the effect of a treatment/exposure on an outcome of interest may take more or less time to unfold and, as claimed by many authors, ignoring the temporal aspect may lead to severely biased estimates and to misleading conclusions about the evolution and the 'mechanism' regulating phenomena under investigation.

Over time, several models and approaches have been proposed to address longitudinal mediation and a kind of dichotomy seems to have emerged. Social and behavioural scientists generally privilege models including latent variables, such as structural equation models or multilevel models, while epidemiologists have traditionally been interested in causal questions, and latent variables are hardly ever included in their models. This thesis aims to reconcile some of these approaches, showing how they can be combined and the reasons why this is advantageous.

The thesis starts with a literature review of models and approaches proposed over years to address discrete-time longitudinal mediation analysis. The focus is on both associational and causal approaches: for each we discuss the main features and highlight strengths and limitations. In the final discussion, we remark differences

between the two frameworks and point out some issues which will be addressed in the following chapters.

Next, we propose a formal unification of structural and multilevel mediation models for longitudinal settings. Such a unified framework overcomes some of the limitations of traditional multilevel mediation models, for example the impossibility to include effects from subject-level variables to cluster-level variables (the so called 'bottom-up effects'), or the fact that multilevel mediation models are not simultaneous. We give also a detailed discussion of all multilevel mediational designs which can be addressed from a structural equation perspective.

In the following, we address two popular associational models for longitudinal mediation analysis, mixed-effect and latent growth models, from a causal perspective by means of the separable effects approach. Assumptions for identifiability of the effects and estimation g-formulas are derived. In addition, a simulation study to evaluate the impact of model misspecification on the estimates of separable mediational effects is carried out. This part concludes with a discussion of possible extensions to more general settings, including baseline and time-varying covariates and time-varying treatments.

The last chapter presents an application of the separable effects approach proposed in the thesis to data from the COVCO-Basel study, carried out in Switzerland between July 2020 and August 2021. The analysis focuses on the estimation of direct and indirect effects of income on depression in a phase of acute spreading of Covid-19. The mediators of interest are worries concerning different aspects of life: economy, health, social and cultural life. The findings show that having a higher monthly income has always a beneficial direct effect on depression, while only some kinds of worries play a mediating role.

There is nothing permanent
except change.

–Heraclitus

Omnia aliena sunt,
tempus tantum nostrum est.
All things are external,
only time is ours.

–Seneca

Acknowledgments

This thesis is the result of an individual effort, but it has benefited from a great deal of support and assistance from several people.

I would first like to thank my supervisor, Professor Gianfranco Lovison, for his invaluable guidance over these years. He has been a mentor to me and, because of his deep and broad knowledge of statistics and his 'humanistic' approach towards it, also an inexhaustible source of inspiration. I cannot remember a single occasion when, at the end of a conversation of ours, I did not feel that I had learned something. His insightful comments and great confidence in my abilities played a key role in writing this dissertation.

My deepest gratitude goes also to my co-supervisor, Dr. Antonino Abbruzzo, who was my first statistics teacher and definitely the one who pushed me to follow this path. I want to thank him for all the hours he spent talking with me and his useful suggestions. His advice about this thesis and, more generally, about the different aspects of academia has been really helpful for my training as a researcher.

I wish to thank Professor Vanessa Didelez, who gave me the opportunity to join her research group at the Leibniz Institute for Prevention Research and Epidemiology in Bremen. She has been extremely kind to me since the first day, and her extensive knowledge of causal inference had a great impact on my education. My stay in Bremen, although short, was crucial for the development of a part of this dissertation.

Many thanks also to Professor Nicole Probst-Hensch and Dr. Ayoung Jeong from the Swiss Tropical and Public Health Institute in Basel for allowing me to use data from the COVCO-Basel study – the analysis of which is the object of the last chapter of the thesis – and for their useful feedback and suggestions.

I cannot fail to mention and thank all other PhD fellows from the DSEAS at the University of Palermo. I am grateful for the stimulating conversations, the sense of community and, of course, for all the funny and carefree moments we shared (especially during lunches!).

I also wish to thank my family for their unwavering support and love. They have always given me the motivation and encouragement to pursue this project, and I could not be luckier for being part of such a wonderful family.

Finally, all my gratitude goes to Rodi, an unexpected gift along the way, for whom no thanks would ever be enough. He has given me the strength to hold on even in the most difficult times, and has constantly encouraged me to believe in myself. We shared every moment of this PhD journey since the very beginning, and now that it is coming to an end, my greatest hope is that it provided us with the awareness and the tools to face a much more challenging journey called life. Hand by hand, always together.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In many fields of science, it is crucial to monitor and understand the evolution of phenomena over time. Epidemiologists may be interested in the time at which a specific disease manifests, in psychology patients are often followed over weeks or months to keep track of their mental health, economists collect longitudinal data to evaluate the impact of policies over time. In all these situations, the goal, more or less explicitly stated, is to gain insight into the mechanism regulating the phenomenon under investigation.

Mediation analysis is a statistical technique widely used for understanding how the effect of an exposure on a response propagates, in particular, it tries to disentangle the *direct* effects and the *indirect* effects, i.e. the effects through intermediate variables called *mediators*. As a (very popular) example, consider the risk of developing lung cancer. It is well known that this risk is affected by the presence of genetic variants on chromosome 15q25.1. However, genetic variants do not act only directly, but they contribute to increase also the propensity to smoke, which in turn increases the probability of lung cancer (VanderWeele et al. 2012a).

Most phenomena, like the example just provided, are dynamic in nature and their mechanisms may take more or less time to unravel. As a consequence, the effects of a treatment or exposure on a certain response may not be immediately observable, but may manifest over time. Longitudinal mediation analysis aims to assess mediational effects in longitudinal settings and how they change over time.

1.1 Motivation and aims

The temporal aspect has paramount importance in mediation analysis, for the simple reason that the exposure or treatment has to precede the mediator, and both have to be antecedent to the response. The need of time ordering is motivated by the fact

that, in most real-world situations, the effect of the treatment on the outcome is not instantaneous. Think of a randomised trial where subjects receive a placebo or a drug to contrast insulin resistance, a syndrome which is often the prelude of diabetes. The drug may affect insulin levels directly, or a proportion of the effect may be mediated by the production of adiponectin, a protein hormone involved in different metabolic processes which is known to regulate insulin production. Researchers do not expect to observe a sudden change in subjects' insulin or adiponectin, indeed these kinds of trials must be carried out for months if not years before being able to detect and measure sizeable effects.

In such contexts, researchers collect repeated measures of the variables of interest to follow their evolution over time, quantify the efficacy of the treatment and how it changes over the follow-up period. Longitudinal data are crucial for the investigation of mediational mechanisms and, in recent years, the issue of longitudinal mediation has witnessed a remarkable development. Over the years, several approaches have been proposed to address longitudinal mediation. Each approach has its own unique features, strengths and limitations and their huge variety may generate confusion.

As will become clear in Chapter 2, one of the main differences between these various approaches is how latent variables are tackled. In social and behavioural sciences, it is quite common to address longitudinal data through a broad class of models called structural equation models: they include several model specifications characterising the trajectories of time-varying variables via latent factors. In contrast, in epidemiology these models are not so frequently employed, and latent variables are regarded as a threat for the possibility to estimate treatment effects from observed data. This difference does not concern only the field of application of mediation analysis, but it is deeper and involves the researchers' conceptual framework. Indeed, social sciences have a long associational tradition, while epidemiologists are often interested in the causes behind phenomena under investigation and therefore they try to make causal inference.

The present dissertation addresses longitudinal mediation analysis from both perspectives, associational and causal, focusing on the issue of latent variables. We revisit established methodologies, or develop new ones, providing the following theoretical contributions:

- we provide a comprehensive overview of associational and causal approaches to longitudinal mediation, highlighting their differences and possible gaps which need to be filled in the literature;
- we show the analogies between the two most commonly used associational models for longitudinal data, i.e. multilevel and structural equation models, showing

how they can be unified, and the advantages of such a unification;

- we endow some of these longitudinal associational models with a causal interpretation within a separable effects framework.

These points will be developed throughout the thesis. The rest of the chapter is devoted to providing the tools to understand what comes next and to present the structure of the thesis.

1.2 Preliminary concepts

In this section we introduce the key concepts underlying mediation analysis, which will be extensively used throughout the thesis. We start with the regression-based approach to mediation and continue discussing the causal framework for mediation based on counterfactuals.

1.2.1 Basics about structural equation models

Structural equation models (SEM) are a wide class of models which allows researchers to model relationships among either observed or latent variables. In this section we briefly review some basic concepts about SEMs which will be recurrent in the following.

SEMs are usually expressed through the LISREL notation (Jöreskog and Sörbom 2001), which decomposes each model into two parts, the measurement model and the structural model. The former specifies the relationship between observed variables and latent variables

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\mu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \mathbf{x} &= \boldsymbol{\mu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \tag{1.1}$$

where $\mathbf{y}$ and $\mathbf{x}$ are p- and q-dimensional vectors of endogenous and exogenous observed variables, respectively, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are r- and s-dimensional vectors of endogenous and exogenous latent variables, respectively, $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$ are p- and q-dimensional vectors of intercept/means, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are $p \times r$ and $q \times s$ coefficient matrices, and $\boldsymbol{\varepsilon}, \boldsymbol{\delta}$ are error vectors. The latter model specifies the relationship between latent variables

$$\boldsymbol{\eta} = \boldsymbol{\nu} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1.2}$$

where $\boldsymbol{\nu}$ is an r-dimensional vector of means, $\mathbf{B}$ is a $r \times r$ matrix of coefficients showing the dependencies among endogenous latent variables, $\boldsymbol{\Gamma}$ is a $r \times s$ matrix of coefficients showing how endogenous latent variables depend on the exogenous ones,

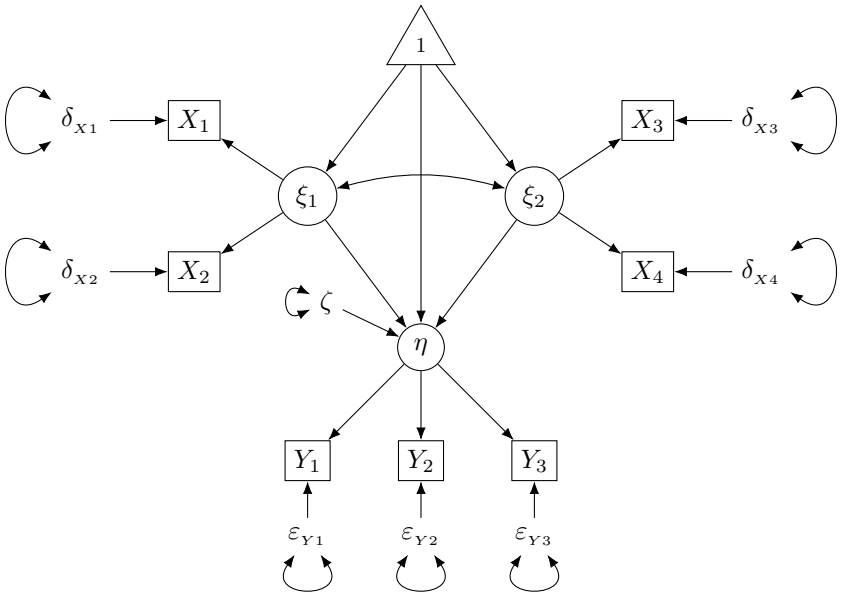

Figure 1.1: Example of SEM.

and $\boldsymbol{\zeta}$ is a vector of errors. It is assumed that $\boldsymbol{\varepsilon}$ is uncorrelated with $\boldsymbol{\eta}$, $\boldsymbol{\zeta}$ and $\boldsymbol{\delta}$ are uncorrelated with $\boldsymbol{\xi}$, and finally that $\boldsymbol{\zeta}, \boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are mutually uncorrelated.

SEMs are primarily focused on the covariance structure characterising model variables: estimation procedures aim to find the parameters which minimise the discrepancy between the observed and the model-implied covariance matrix. In some cases however, the mean structure can be of interest as well.

A useful way to visualise SEMs, especially those including many variables, is by representing them through path diagrams. Introduced by Wright (1921), path diagrams are graphs showing the relationship of dependence and the covariance among variables in the system. They are characterised by five elements:

- Squares, indicating observed variables
- Circles, indicating latent variables
- Triangles, denoting constants (generally intercepts)
- Directional arrows, denoting the effect of a variable on another one
- Bidirectional arrows, denoting covariances or variances of error terms.

Consider Figure 1.1. It represents a model with $p = 3$ endogenous and $q = 4$ exogenous observed variables, and $r = 1$ endogenous and $s = 2$ exogenous latent

variables. The measurement model is

$$\underbrace{\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix}}_{\boldsymbol{\Lambda}_y} \eta + \underbrace{\begin{pmatrix} \varepsilon_{Y1} \\ \varepsilon_{Y2} \\ \varepsilon_{Y3} \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} \alpha_1 & 0 \\ \alpha_2 & 0 \\ 0 & \beta_1 \\ 0 & \beta_2 \end{pmatrix}}_{\boldsymbol{\Lambda}_x} \underbrace{\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}}_{\boldsymbol{\xi}} + \underbrace{\begin{pmatrix} \delta_{X1} \\ \delta_{X2} \\ \delta_{X3} \\ \delta_{X4} \end{pmatrix}}_{\boldsymbol{\delta}}$$

where $\boldsymbol{\eta} \equiv \eta$ is a scalar (i.e. a 1×1 vector), $\boldsymbol{\mu}_y$ and $\boldsymbol{\mu}_x$ are null vectors and the coefficients linking the latent variables to their indicators are not reported in the figure to avoid clutter. The structural model is made up by only one equation, since there is only one endogenous latent variable

$$\eta = \nu + \begin{pmatrix} \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} + \zeta.$$

When all the variables in the system are observed, i.e. $\boldsymbol{\xi} = \mathbf{x}$, $\boldsymbol{\eta} = \mathbf{y}$, structural equation models take the name of *path analytic models* (Wright 1934) or *simultaneous* equation models. Several mediational models in the literature are path-analytic and several scholars have proposed methods to estimate the indirect effects, see for example Alwin and Hauser (1975), Greene (1977), Finney (1972) and Fox (1980). Bollen (1987) proposed a generalisation of their methods, which can be used in a wide variety of SEMs, including both observed and latent variables. We will not provide the details here, what is important to know is that indirect effects can be obtained by appropriately combining the coefficient matrices in (1.1)-(1.2). An extensive discussion can be found in Bollen (1987) and (Bollen 1989, Chapter 8).

1.2.2 Regression-based mediation analysis

Baron and Kenny (1986) paved the way for the use of mediation analysis by researchers from different fields. Although they did not address some crucial concepts now widely recognised in mediational settings (e.g. post-treatment confounders, time ordering) and some of their claims have proved to be incorrect (Hayes 2009, Zhao et al. 2010), their role in spreading the use of mediation methods is unquestionable, as witnessed by the thousands of citations of their paper.

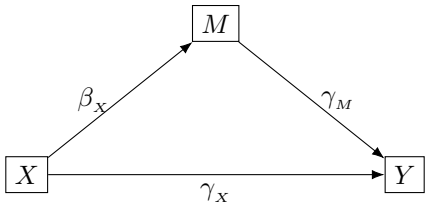

Figure 1.2: The basic mediation model.

In this section, we start from their regression-based approach, which has been extensively addressed and deepened in MacKinnon (2008) and Hayes (2018). The basic mediation model includes an explanatory variable X, an intermediate variable M and a response Y: X may affect Y directly and also indirectly through M, as shown in Figure 1.2. The X variable is called *treatment* in experimental studies, where researchers have a certain degree of control over it (for example, X can be the type of medicine assigned to participants in a randomized control trial, a new drug versus a placebo), or *exposure* in observational studies, where X is not manipulated (for example, it can be the number of cigarettes smoked in a day). If the context is not specified, we will generally refer to X as the *exposure*.

A first important observation to make is that, even in single-time settings, mediation analysis entails a time ordering, since the exposure has to precede the mediator, which in turn has to precede the response. As already mentioned, the quantities of interest are the direct and indirect effects. In the path-analytic approach, which is not inherently causal, effects are obtained as combinations of coefficients lying on paths of interest. The mediator and the response are assumed to be Normally distributed, and all the relationships to be linear. Thus, the model equations are

$$M = \beta_0 + \beta_X X + \varepsilon_M \tag{1.3}$$

$$Y = \gamma_0 + \gamma_X X + \gamma_M M + \varepsilon_Y, \tag{1.4}$$

where the first equation represents the mediator model and the last one the response model including both the exposure and the mediator. The error terms are assumed to be independent. For example, in the simple setting shown in Figure 1.2, the direct effect corresponds to the arrow $X \to Y$, and it is measured by γ_X, while the indirect effect is made up of the arrows $X \to M$ and $M \to Y$, so that it can be obtained as the product $\beta_X \gamma_M$.

From an SEM perspective, Equations (1.3)-(1.4) can be rewritten as follows

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{B}\mathbf{y} + \boldsymbol{\varepsilon} \tag{1.5}$$

$$\begin{pmatrix} X \\ M \\ Y \end{pmatrix} = \begin{pmatrix} \mu_X \\ \beta_0 \\ \gamma_0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \beta_X & 0 & 0 \\ \gamma_X & \gamma_M & 0 \end{pmatrix} \begin{pmatrix} X \\ M \\ Y \end{pmatrix} + \begin{pmatrix} \varepsilon_X \\ \varepsilon_M \\ \varepsilon_Y \end{pmatrix}, \tag{1.6}$$

and the indirect effect can be obtained through the approach described in Bollen (1989, 1987), not reported here for the sake of brevity. This approach entails multiplying the $\mathbf{B}$ matrix by itself,

$$\mathbf{B}^2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_X \gamma_M & 0 & 0 \end{pmatrix}.$$

The only non-null entry is the indirect effect of X on Y.

The effects defined above do not have a causal meaning, unless specific assumptions are satisfied. In the next section, we introduce the counterfactual framework to address causality and show how the definition of causal mediational effects differ from the associational one just described.

1.2.3 Counterfactuals

The concept of causation has been more or less explicitly banned from the mainstream statistical literature for all the 19th and the great majority of the 20th century, since the predominant paradigm was that only associational relationships could be inferred from observed data. Starting from the '70s, a formal causal theory arose within the potential outcome framework (Rubin 1974, 1978). Since then, a great stream of literature has begun defining causal effects through potential outcomes, also called counterfactuals, thanks to the fundamental contributions of many scholars, including James Robins (Robins 1986) and Judea Pearl (Pearl 2009, 2010), although approaches to causality not involving counterfactuals have been proposed as well (Dawid 2000).

Counterfactuals are statements referring to events or situations contrary to those that occurred, such as "If the patient had taken the drug, he/she would be alive" (but he/she did not receive the drug). A researcher may be interested in understanding if the patient's death has been caused by not assuming the drug or not. If one denotes by X the binary variable "Taking the drug" and by Y the binary response indicating patient's survival, one may conceive the causal effect of X on Y as a contrast between the outcome that would have been observed if the patient had taken the drug and that observed if he/she had not, leaving all the rest unchanged. If the two *potential outcomes* (the expression used by Holland, 1986, referring to Rubin's work) differ, assuming we control for all variables confounding the X-Y relationship, then X has

a causal effect on Y. It is possible to translate these concepts in mathematical terms by writing the individual total causal effect for subject i, on the difference scale, as $Y_i(X = 1) - Y_i(X = 0)$, where $X = 1$ means the patient took the drug and $X = 0$ he/she did not, and $Y_i(X = x)$ is the value the response for subject i would have had if X had been set to $x \in \{0, 1\}$.

In fact, it is possible to observe just one counterfactual outcome, that is, the one corresponding to the treatment taken by the patient, while the other one is an artificial construct. In general, subject-specific causal effects are not targets of inference. Population effects are much more useful and widely employed: counterfactuals are defined for a population of interest, then $Y(x)$ becomes a random variable with its own distribution $P(Y(x))$ if Y is discrete, $f(Y(x))$ if it is continuous. The average causal effect (ACE) can be defined comparing the counterfactual distributions of Y under two different values of X, x and x^*, for example as the expected difference

$$ACE = \mathbb{E}[Y(x)] - \mathbb{E}[Y(x^*)] = \mathbb{E}[Y(x) - Y(x^*)]$$

or the causal odds ratio if Y is binary

$$ACE_{OR} = \frac{P(Y(x) = 1)/P(Y(x) = 0)}{P(Y(x^*) = 1)/P(Y(x^*) = 0)}.$$

It is worth noting that both the expressions above entail a comparison of the same population under two different scenarios. For example, the ACE can be read as the difference between the outcome if *all* the subjects in the population of interest had been exposed to x and the outcome if *all* the subjects had been exposed to x^*. This is generally different from the associational difference

$$\mathbb{E}[Y \mid X = x] - \mathbb{E}[Y \mid X = x^*]$$

where a group of subjects in the population has actually been exposed to x and the other group to x^* (see Figure 1.3). The causal and the associational quantities coincide under the exchangeability assumption

$$Y(x) \perp\!\!\!\perp X \qquad \forall\, x \tag{1.7}$$

which says that each counterfactual is independent of X. Let us make this statement clearer with an example. Suppose that Y is an health outcome, say blood pressure, and X is binary, expressing the assumption of a new experimental drug $X = 1$ versus a placebo $X = 0$. A researcher is interested in evaluating the causal effect of the drug on blood pressure. To do so, he/she selects a group of patients suffering from

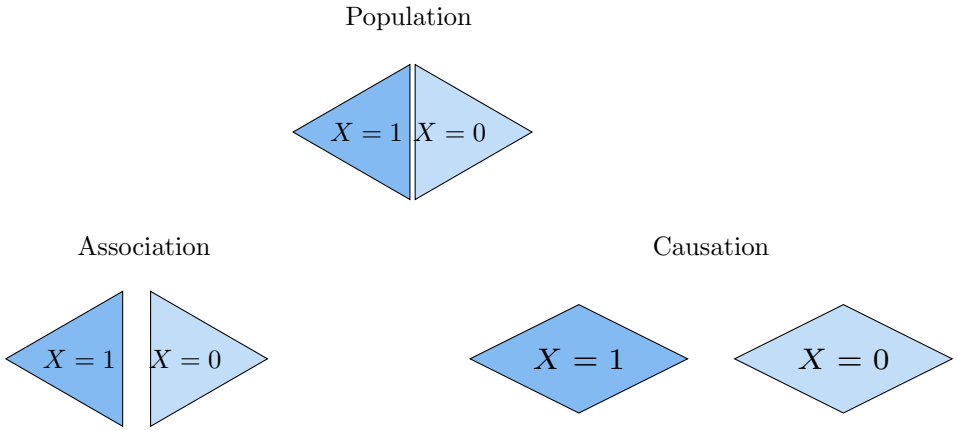

Figure 1.3: Association versus causation.

high pressure levels and randomly assigns a group to the treatment condition and the other group to the control condition. Since the assignment to the groups was random, the two group of patients are equivalent or, in other words, the potential outcome 'observed' for the treated subjects would have been equal to that of subjects in the control group, if they have been assigned to the treatment group and vice versa: the two groups are exchangeable (Hernán and Robins 2020).

The hypothetical experiment we have just described is an example of randomized control trial (RCT). RCTs represent a golden standard in the causal inference literature, since they ensure that the relationship between X and Y is not confounded, i.e. affected by other, possibly unmeasured, covariates. If exchangeability holds, then causation coincides with association and

$$\mathbb{E}[Y(x)] = \mathbb{E}[Y(x) \,|\, X = x] = \mathbb{E}[Y \,|\, X = x]$$

the last equality holding by consistency: $Y(x) = Y$ if $X = x$, i.e. a potential outcome coincides with the observed outcome if the value of X is that actually observed. When a counterfactual quantity can be expressed as a function of observed variables it is said to be *identifiable*. Identifiability is a crucial issue in causal inference.

In observational studies, exhangeability is very unlikely to hold true due to the presence of confounders of the exposure-outcome relationship. Denoting by C these confounders, it is however possible to achieve exchangeability within the strata of these covariates:

$$Y(x) \perp\!\!\!\perp X \,|\, C. \tag{1.8}$$

Assumption (1.8) is called *conditional exchangeability*, and allows to obtain, for ex-

ample, the conditional counterfactual expectation as

$$\mathbb{E}[Y(x) \mid C] = \mathbb{E}[Y(x) \mid X = x, C] = \mathbb{E}[Y \mid X = x, C].$$

It is worth remarking that identification concerns the entire counterfactual distribution $f(Y(x))$, but generally researchers focus on specific aspects of this distribution, like the expectation or the odds ratios.

See Rubin (2004), Pearl (2009), Morgan and Winship (2015) for more details on counterfactuals and causal inference. In the mediation setting, $M(x)$ denotes the value of the mediator if X is set to x, and $Y(x, m)$ the value of the outcome if X is set to x and M is set to m. It is also possible to consider nested counterfactuals like $Y(x,\, M(x^*))$, i.e., the outcome value if the treatment were set to x and the mediator to the value it would assume under $X = x^*$.

1.2.4 DAGs and causal models

The relationships among variables can be visualised through graphs having specific features. These graphs should represent the model structure and the dependencies among the variables of interest hypothesised by the researcher, and this phase comes before the analysis of the actual data. In other words, a graph reflects the researcher's ideas, conjectures and questions on the phenomenon under investigation, generally based on previous knowledge. Representing observed and unobserved confounders in the graph allows the researcher to get an idea about which effects are of interest and if they are identifiable, even before the analysis of observed data. In mediational settings, the relationships among variables are directed, then we will use directed graphs.

A directed acyclic graph (DAG) $\mathcal{G}$ is a couple $\{\boldsymbol{V}, \boldsymbol{E}\}$, where $\boldsymbol{V}$ is the set of vertices or nodes and $\boldsymbol{E}$ is a set of directed edges linking the nodes. DAGs are characterised by the absence of loops, i.e. sets of subsequent edges starting and ending with the same node. A set of successive edges is called *path* and, if the direction given by arrows is respected, the path is said to be *directed*. For two nodes $V_i, V_j \in \boldsymbol{V}$, if there exists a directed path $V_i \to \cdots \to V_j$ from V_i to V_j, V_i is called an ancestor of V_j and V_j a descendant of V_i. In particular, if $V_i \to V_j$, V_i is called *parent* and V_j *child*. The set of parents, ancestors and descendants of a node V_k are denoted by $\text{pa}(V_k)$, $\text{an}(V_k)$ and $\text{de}(V_k)$, respectively[1]. For example, in the DAG represented in Figure 1.4, $V_1 \to V_3 \leftarrow V_5$ is a path and $V_1 \to V_2 \to V_5$ is a directed path. V_1 is a parent of V_2, V_3 and V_4, and an ancestor of V_5, and V_4 is child of V_3 [2]. Given a subset of nodes

[1] Sometimes we will write pa_k, an_k and de_k to simplify the notation

[2] At this stage, the labeling of nodes is arbitrary and it does not imply any temporal ordering.

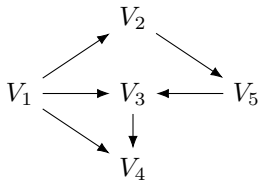

Figure 1.4: Example of a DAG.

$\boldsymbol{A} \subset \boldsymbol{V}$, the induced subgraph $\mathcal{G}_{\boldsymbol{A}}$ is the graph having vertex set $\boldsymbol{A}$ and edge set consisting of links in $\boldsymbol{E}$ with both endpoints in $\boldsymbol{A}$. For example, if $\mathcal{G}$ is the graph in Figure 1.4 and $\boldsymbol{A} = \{V_1, V_3, V_5\}$, the induced subgraph $\mathcal{G}_{\boldsymbol{A}}$ is $V_1 \to V_3 \leftarrow V_5$.

We identify vertices with corresponding random variables, each variable V_k having support $\mathcal{V}_k$, and the edges may assume different meanings on the ground of the assumptions a researcher is willing to make. The random vector $\mathbf{v} = (V_1, V_2, \ldots, V_K)$ factorises according to DAG $\mathcal{G}$ if the joint distribution satisfies

$$P(v_1, \ldots, v_K) = \prod_k P(v_k \,|\, \mathrm{pa}(v_k)). \tag{1.9}$$

In other words, the joint distribution can be factorised as the product of the conditional distribution of each node given only its parents.

Expression (1.9) can be read also in terms of conditional independence, since it implies

$$V_k \perp\!\!\!\perp \{\mathrm{nd}(V_k) \setminus \mathrm{pa}(V_k)\} \,|\, \mathrm{pa}(V_k)$$

where $\mathrm{nd}(V_k)$ are the non-descendants of V_k. It follows that two graphs can encode the same set of conditional independencies, as those shown in Figure 1.5, for which $V_1 \perp\!\!\!\perp V_3 \,|\, V_2$. When two different graphs represent the same set of conditional independencies, they are said *Markov equivalent*.

Figure 1.5: Markov equivalent DAGs.

Probabilities distributions which can be represented by a DAG, i.e. which admit factorisation (1.9) relative to a certain DAG, can be characterised through a graphical criterion called *d-separation*.

Definition 1.2.1 (d-separation criterion)**.** In a DAG $\mathcal{G}$, a path π is said to be d-separated or blocked by a set of nodes $\boldsymbol{Z}$ if and only if

1. π contains a chain $V_1 \to V_2 \to V_3$ or a fork $V_1 \leftarrow V_2 \to V_3$ such that the central node V_2 is in $\boldsymbol{Z}$ or

2. π contains a collider $V_1 \to V_2 \leftarrow V_3$ such that neither V_2 nor any of its descendants are in $\boldsymbol{Z}$.

Two sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ are said to be d-separated by $\boldsymbol{Z}$ if $\boldsymbol{Z}$ blocks any paths from a node in $\boldsymbol{X}$ to a node in $\boldsymbol{Y}$.

In Figure 1.4, for example, V_1 is d-separated from V_5 by V_2, but they are not d-separated by $\{V_3, V_4\}$.

It is worth remarking that d-separation has also a probabilistic implication, since if two sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-separated by $\boldsymbol{Z}$ in a DAG $\mathcal{G}$, then $\boldsymbol{X}$ is independent of $\boldsymbol{Y}$ conditional on $\boldsymbol{Z}$ in every distribution represented by $\mathcal{G}$. Vice versa, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are not d-separated by $\boldsymbol{Z}$ in $\mathcal{G}$, there exists at least one distribution represented by $\mathcal{G}$ in which $\boldsymbol{X}$ and $\boldsymbol{Y}$ are dependent conditional on $\boldsymbol{Z}$.

In light of what said so far, it should be clear that DAGs do not encode any causal relationship per se, but only conditional independencies. They can be endowed with a causal meaning if some assumptions are satisfied.

As discussed in the previous section, the concept of causality entails the idea of manipulating or intervening on some variables in the system of interest, to see how the distribution of the outcome changes. The distribution of the outcome under an intervention on the exposure can be expressed through the counterfactual language as $P(Y(x))$, and this contrasts with traditional associational settings based on conditioning $P(Y \,|X = x)$. This leads to the following definition of causal DAG

Definition 1.2.2 (Causal DAG)**.** Let $\mathcal{G}(\boldsymbol{V}, \boldsymbol{E})$ be a DAG and $\boldsymbol{X} \subset \boldsymbol{V}$ a subset of variables on which an intervention is performed (at least hypothetically) and which can be set to any value $\boldsymbol{x}$ in the domain of $\boldsymbol{X}$. $\mathcal{G}$ is causal for $\boldsymbol{V}$ if the joint distribution $P(\cdot)$ satisfies the following:

(i) P factorises according to $\mathcal{G}$;

(ii) $P(v_k(\boldsymbol{x})) = 1$ for any $V_k \in \boldsymbol{X}$;

(iii) $P(v_k(\boldsymbol{x}) \,|\, \mathrm{pa}_k) = P(v_k \,|\, \mathrm{pa}_k)$ for any $V_k \notin \boldsymbol{X}$.

Properties (ii)-(iii) tell us that the interventional distribution $P(\boldsymbol{v}(\boldsymbol{x}))$ obtained via an intervention setting $\boldsymbol{X}$ to $\boldsymbol{x}$ is given by

$$P(\boldsymbol{v}(\boldsymbol{x})) = \prod_{\{k|V_k \notin \boldsymbol{X}\}} P(v_k \,|\, \mathrm{pa}_k), \tag{1.10}$$

which is known as *truncated factorisation* (Pearl 2009) or *g-formula* (Robins 1986).

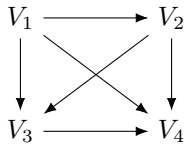

Figure 1.6: Example of back-door adjustment.

In causal DAGs, an arrow between two nodes is interpreted as a causal effect of the source node on the target. However, in complex systems involving many variables and many paths among them, it can be difficult to determine if a causal effect of interest is identified, due to the presence of confounders. As discussed in Section 1.2.3, in observational studies some degree of confounding is unavoidable, and identification can be achieved if conditional exchangeability (1.8) holds. But which are the covariates on which one should condition to achieve conditional exchangeability? Inspecting causal DAGs can be helpful for finding out such confounders and understanding which causal effects can be identified and which are unidentifiable, by means of two graphical criterion known as the *back-door* and the *front-door* criteria.

A back-door path from V_i to V_j is a path which does not have an edge emanating from V_i, so it has the form $V_i \leftarrow \dots V_j$. In Figure 1.4, the path $V_5 \leftarrow V_2 \leftarrow V_1$ is a back-door path from V_5 to V_1.

Definition 1.2.3 (Back-door criterion (Pearl 2009))**.** A set of variables $\boldsymbol{Z}$ satisfies the back-door criterion relative to a pair of variables (V_i, V_j) in a DAG $\mathcal{G}$ if:

(i) no node in $\boldsymbol{Z}$ is a descendant of V_i;

(ii) $\boldsymbol{Z}$ blocks all back-door paths from V_i to V_j.

Analogously, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two disjoint sets of nodes in $\mathcal{G}$, then $\boldsymbol{Z}$ satisfies the back-door criterion relative to $(\boldsymbol{X}, \boldsymbol{Y})$ if it satisfies the back-door criterion relative to any pair (V_i, V_j) such that $V_i \in \boldsymbol{X}$ and $V_j \in \boldsymbol{Y}$.

Theorem 1.2.1 (Back-door adjustment)**.** *If a set of variables $\boldsymbol{Z}$ satisfies the back-door criterion relative to $(\boldsymbol{X}, \boldsymbol{Y})$, then the causal effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is identifiable and is given by*

$$P(Y(x) = y) = \sum_z P(y \,|\, x, z) P(z). \tag{1.11}$$

In Figure 1.4, the causal effect of V_3 on V_4 can be identified conditioning on V_1, since it blocks the back-door path $V_3 \leftarrow V_1 \rightarrow V_4$. As another example, consider Figure 1.6: in order to identify the causal effect of V_3 on V_4, conditioning on V_1 is not sufficient, since it does not block the back-door path $V_3 \leftarrow V_2 \rightarrow V_4$. Then, the

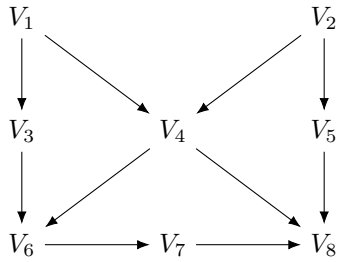

Figure 1.7: Illustration of the front-door criterion.

adjusting set is $\boldsymbol{Z} = \{V_1, V_2\}$, and the interventional density of V_4 is

$$P(V_4(v_3) = v_4) = \sum_{v_1,v_2} P(v_4 \,|\, v_3, v_1, v_2)P(v_1, v_2) = \sum_{v_1,v_2} P(v_4 \,|\, v_3, v_1, v_2)P(v_2 \,|\, v_1)P(v_1).$$

Definition 1.2.4 (Front-door criterion)**.** A set of variables $\boldsymbol{Z}$ is said to satisfy the front-door criterion relative to an ordered pair of variables (V_i, V_j) if:

(*i*) $\boldsymbol{Z}$ intercepts all directed paths from V_i to V_j;

(*ii*) There are no unblocked back-door paths from V_i to $\boldsymbol{Z}$;

(*iii*) All back-door paths from $\boldsymbol{Z}$ to V_j are blocked by V_i.

Theorem 1.2.2 (Front-door adjustment)**.** *If $\boldsymbol{Z}$ satisfies the front-door criterion relative to (X, Y) and $P(x, z) > 0$, then the causal effect of X on Y is identified and given by*

$$P(Y(x)) = \sum_z P(z \,|\, x) \sum_{x'} P(y \,|\, x', z)P(x'). \tag{1.12}$$

In Figure 1.7 , the effect of V_6 on V_8 is identified conditioning on $Z = V_7$, since it intercepts the only directed path between V_6 and V_8; all the back-door paths between V_6 and V_7 are blocked since they pass through the collider V_4 which is not in the conditioning set; all the back-door paths between V_7 and V_8 are blocked by V_6. The causal effect of V_6 on V_8 is then

$$P(V_8(v_6)) = \sum_{v_7} P(v_7|v_6) \sum_{v_6'} P(v_8 \,|\, v_6', v_7)P(v_6').$$

The back-door and the front-door criteria are only sufficient, since there are cases in which none of them is satisfied, still identification can be achieved. We will come back to this point at the end of this subsection. See Pearl (2009, 2014), Didelez (2019*a*), Shpitser (2012, 2019) for a broader discussion on DAGs and identifiability of causal effects.

DAGs can be associated to causal models, which can be defined as sets of factual and counterfactual distributions subject to some restrictions. Such constraints allow us to identify some counterfactual distributions from observed data (Shpitser 2019). There exist several causal models (see Robins and Richardson 2011 for a review), but in this thesis we will focus on the two best known: the Finest Fully Randomized Causally Interpretable Structured Tree Graph (FFRCISTG) (Robins 1986, 2003) and the non-parametric structural equation model with independent errors (NPSEM-IE or simply NPSEM) proposed by Pearl (1995).

Definition 1.2.5 (FFRCISTG)**.** Let $\mathcal{G}(\boldsymbol{V})$ be a DAG with node set $\boldsymbol{V} = \{V_1, \ldots, V_K\}$, where V_i is a non-descendant of V_j if $i < j$. In addition, for any $k = 1, \ldots, K$, let $\bar{v}_k = (v_1, \ldots, v_k)$. A FFRCISTG associated with $\mathcal{G}$ is defined by the following assumptions:

(*i*) All counterfactuals $V_k(\bar{v}_{k-1})$ exist for any value assumed by $\bar{v}_{k-1}$;

(*ii*) $V_k(\bar{v}_{k-1})$ depends on $\bar{v}_{k-1}$ only through the values of V_k's parents in $\mathcal{G}$, i.e. $V_k(\bar{v}_{k-1}) \equiv V_k(\text{pa}_k)$;

(*iii*) Both the observed V_k and the counterfactuals $V_k(w)$, for any $W \subset V$, are obtained recursively from $V_k(\bar{v}_{k-1})$, (e.g. $V_3 = V_3(V_1, V_2(V_1))$);

(*iv*) For each k, any possible values of future counterfactuals, starting from $k+1$, is independent from V_k, given its history, i.e.

$$\{V_{k+1}(\bar{v}_k), \ldots, V_K(\bar{v}_{K-1})\} \perp\!\!\!\perp V_k(\bar{v}_{k-1}) \,|\, \overline{V}_{k-1} = \bar{v}_{k-1} \quad \forall\ k, \bar{v}_{K-1} \in \overline{\mathcal{V}}_{K-1}. \tag{1.13}$$

Definition 1.2.6 (NPSEM)**.** A NPSEM associated with a DAG $\mathcal{G}(\boldsymbol{V})$ assumes the existence of unknown deterministic functions f_k and mutually independent random disturbances ε_k such that, for each $k = 1, \ldots, K$, $V_k = f_k(\text{pa}_k, \varepsilon_k)$. In addition, observed and counterfactual variables can be obtained via recursive substitution as in (*iii*) of Definition 1.2.5.

For example, in the basic mediation model $X := f_{_X}(\varepsilon_{_X})$, $M := f_{_M}(X, \varepsilon_{_M})$ and $Y := f_{_Y}(X, M, \varepsilon_{_Y})$.

A NPSEM implies the following independence

$$\{V_{k+1}(\bar{v}_k), \ldots, V_K(\bar{v}_{K-1})\} \perp\!\!\!\perp V_k(\bar{v}^{**}_{k-1}) \,|\, \overline{V}_{k-1} = \bar{v}^{*}_{k-1} \quad \forall\ k, \bar{v}_{K-1} \in \overline{\mathcal{V}}_{K-1} \text{ and } \bar{v}^{*}_{k-1}, \bar{v}^{**}_{k-1} \in \overline{\mathcal{V}}_{k-1} \tag{1.14}$$

and it can be proved that

Lemma 1.2.3. *A NPSEM can be characterised by (i)-(iv) in Definition 1.2.5, with Equation (1.13) replaced by (1.14).*

It follows that NPSEMs are special cases of FFRCISTGs. Understanding the difference between (1.13) and (1.14) is crucial. In a FFRCISTG, the value of $\bar{v}_{k-1}$ in the counterfactual $V_k(\bar{v}_{k-1})$, in the conditioning and in the set of subsequent counterfactuals $\{V_{k+1}(\bar{v}_k), \ldots, V_K(\bar{v}_{K-1})\}$ is the same, while in NPSEM the values it can assume are allowed to differ. As highlighted by Shpitser and Tchetgen Tchetgen (2016),

> The FFRCISTG model always imposes restrictions on a set of variables under a single set of interventions (a "single world"), while the NPSEM-IE may also impose restrictions on variables across multiple conflicting sets of interventions simultaneously.

While the independencies encoded by a FFRCISTG are, at least in principle, testable, those encoded by NPSEMs in general are not. The 'single-world' nature of FFRCSTGs and the 'multiple-world' nature of NPSEMs impacts the kinds of effects which it is possible to identify within their framework, as will become clear in the next section.

We conclude this section with an important result concerning the identification of total effects in NPSEMs provided by Shpitser et al. (2010). They propose a generalisation of the back-door criterion, called *adjustment criterion*, which, in contrast to the back-door and front-door criteria, is complete.

Definition 1.2.7 (Proper causal path)**.** Let $\mathcal{G}(\boldsymbol{V})$ be a DAG and $\boldsymbol{X}, \boldsymbol{Y} \subset \boldsymbol{V}$. A directed path from a node X in $\boldsymbol{X}$ to a node in $\boldsymbol{Y}$ is called proper causal with respect to $\boldsymbol{X}$ if it intersects $\boldsymbol{X}$ only at X.

Definition 1.2.8 (Adjustment criterion)**.** The adjustment criterion holds for C with respect to $(\boldsymbol{X}, \boldsymbol{Y})$ if C blocks all paths from $\boldsymbol{X}$ to $\boldsymbol{Y}$ which are not proper causal with respect to $\boldsymbol{X}$, and if C is not a descendant of any node on a proper causal path from $\boldsymbol{X}$ to $\boldsymbol{Y}$ (except possibly nodes in $\boldsymbol{X}$ themselves) in the graph where all arcs pointing to $\boldsymbol{X}$ are cut.

Theorem 1.2.4. *In any causal DAG $\mathcal{G}$, C satisfies the adjustment criterion for $(\mathbf{X}, \boldsymbol{Y})$ if and only if in every NPSEM inducing $\mathcal{G}$, $P(Y(x)) = \sum_c P(Y \,|\, x, c)P(C = c)$.*

Theorem 1.2.5. *In any causal DAG $\mathcal{G}$, C satisfies the adjustment criterion for $(\mathbf{X}, \boldsymbol{Y})$ if and only if in every NPSEM inducing $\mathcal{G}$, $Y(x) \perp\!\!\!\perp X \,|\, C$.*

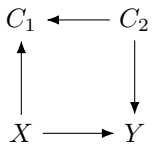

Figure 1.8: Causal DAG where the total effect of X on Y is identified through the adjustment criterion.

These results are very important, since they imply that in NPSEMs the adjustment criterion characterises both covariate adjustment and conditional exchangeability, which are indeed equivalent. As an example, consider Figure 1.8. In this DAG, the back-door criterion does not hold, but the counterfactual distribution $P(Y(x))$ is identified conditioning on $\{C_1, C_2\}$ and is given by

$$P(Y(x)) = \sum_{c_1,c_2} P(Y \mid x, c_1, c_2)P(c_1, c_2).$$

1.2.5 Types of mediational effects

With the background introduced in the previous sections, we are now ready to introduce the different types of causal mediational effects which have been proposed over time. Consider a simple mediation model as that shown in Figure 1.2. We denote by $M(x)$ the value of the mediator if the exposure have been set to x, and by $Y(x, m)$ the value of the outcome if the exposure has been set to x and the mediator to m. For graphical consistency, in the next DAGs we will use squares and circles as in path diagrams.

Natural mediational effects

Consider two values of the exposure, x and x^*, where, for example, if X is binary, $x = 1$ denotes the treatment and $x^* = 0$ denotes no treatment or treatment at a baseline level. Robins and Greenland (1992) introduce four types of mediational effects, which they call pure and (somehow confusingly) total direct and indirect effects, and discuss the conditions for their identifiability. On the difference scale, these effects are defined as follows:

$$PDE = \mathbb{E}[Y(x, M(x^*)) - Y(x^*, M(x^*))] \tag{1.15}$$

$$TDE = \mathbb{E}[Y(x, M(x)) - Y(x^*, M(x))] \tag{1.16}$$

$$PIE = \mathbb{E}[Y(x^*, M(x)) - Y(x^*, M(x^*))] \tag{1.17}$$

$$TIE = \mathbb{E}[Y(x, M(x)) - Y(x, M(x^*))]. \tag{1.18}$$

The direct effects compare the values taken by the outcome when the mediator is fixed to the natural value it would assume either under $X = x^*$ (PDE) or under $X = x$ (TDE), while the exposure changes from x^* to x: this captures the effect along the direct path between X and Y. In contrast, the pure and total indirect effects compare the values of the outcome when the exposure is fixed either to x or x^*, while the mediator changes from $M(x^*)$ to $M(x)$: this captures the effect of X on Y conveyed by M.

Using composition $Y(x) = Y(x, M(x))$, i.e. the counterfactual $Y(x)$ obtained setting X to x does not differ from that obtained setting X to x and M to the value it would have taken if X were set to x, it is easy to prove that the total effect of X on Y can be decomposed into the sum of a pure and a total mediational effect:

$$\begin{aligned} TE =& \mathbb{E}[Y(x) - Y(x^*)] = \\ & \underbrace{\mathbb{E}[Y(x, M(x^*))] - \mathbb{E}[Y(x^*, M(x^*))]}_{PDE} + \underbrace{\mathbb{E}[Y(x, M(x))] - \mathbb{E}[Y(x, M(x^*))]}_{TIE} = \\ & \underbrace{\mathbb{E}[Y(x, M(x))] - \mathbb{E}[Y(x^*, M(x))]}_{PIE} + \underbrace{\mathbb{E}[Y(x^*, M(x))] - \mathbb{E}[Y(x^*, M(x^*))]}_{TDE}. \end{aligned}$$

Pearl (2001) changed the nomenclature, referring to these effects as natural. The PDE and the TIE are more widely used than the TDE and the PIE, so we will often use the term natural effects to denote the former ones, denoting them by NDE and NIE.

Equations (1.15)-(1.18) are all characterised by the presence of cross-world counterfactuals, $Y(x, M(x^*))$ and $Y(x^*, M(x))$, which set X to different values and can never be observed. Consider, for example, the first term of PDE, $Y(x, M(x^*))$. As noted by Robins and Greenland (1992), to observe a potential outcome of the form $Y(x, M(x^*))$, a researcher should assign a subject to treatment x^*, measure $M(x^*)$, then come back to the state of the world before the intervention, assign the same subject treatment level x and measure the value of Y under this treatment and the value of the mediator under the other treatment x^*, measured previously. This kind of experimental manipulation is clearly unfeasible. For this reason, identifiability of natural effects requires a cross-world independence assumption which has long been debated in the literature. Sufficient assumptions for identifying natural effects are listed below, and are discussed subsequently.

First, we assume composition and consistency $Y(x, m) = Y$ if $X = x$, $M = m$, i.e. if the observed exposure and mediator take values x and m, respectively, the counterfactual $Y(x, m)$ equals the observed Y. Second, the following ignorability (or exchangeability) assumptions on unobserved confounders suffice for identification:

(a) $Y(x,m) \perp\!\!\!\perp X$, i.e. no unmeasured confounders for the exposure-outcome relationship;

(b) $Y(x,m) \perp\!\!\!\perp M \,|\, X$, i.e. no unmeasured confounders for the mediator-outcome relationship;

(c) $M(x) \perp\!\!\!\perp X$, i.e. no unmeasured confounders for the exposure-mediator relationship;

(d) $Y(x,m) \perp\!\!\!\perp M(x^*)$.

These assumptions have long been subject of debate among scholars. Indeed, although assumptions (a) and (c) can be satisfied at least in an experimental setting where the treatment is randomised, (b) and (d) are much more complex to deal with. Natural effects rely on a manipulation not only of the treatment, but of the mediator as well, which is considered a second intervening variable. Acting on the mediator is not always possible, and that is the reason why assumption (b), which requires randomisation of the mediator, is generally not satisfied in a single-experiment design, i.e. a setting where only X is randomised.

Imai et al. (2013) claim that in a single-experiment design assumptions (a) and (c) are not sufficient to make natural effects identifiable, and for such effects it is only possible to compute sharp bounds, see also Sjölander (2009) and Robins and Richardson (2011). For point identification, assumption (d) is crucial. In contrast to the other assumptions, it is a cross-world independence assumption that involves counterfactuals corresponding to two different interventions, never observable together. For this reason, as already said, this assumption is untestable and, clearly, it can hold only in a NPSEM. It has traditionally been interpreted as no confounders (measured and unmeasured) of the mediator-outcome relationship affected by the exposure, although Andrews and Didelez (2021) show that this assumption can be violated even in the absence of such a confounder.

Assumption (d) is even stronger than (b), since it does not require a simple intervention on the mediator, but an intervention that, for each individual in the sample, sets the mediator to the value it would assume under treatment x^*, while the outcome assumes the value it would have under another treatment x. Imai et al. (2013) discuss several designs alternative to the single-experiment one where it is possible to manipulate the mediator in order to mimic nested cross-world counterfactuals of the form $Y(x, M(x^*))$. Although these designs allow for identification of mediational effects without assumption (d), they nevertheless require other untestable assumptions whose plausibility is not always trivial to conceive, as the authors themselves and the discussants of their article acknowledge.

See VanderWeele and Vansteelandt (2009) for a thorough discussion of all the assumptions, and Andrews and Didelez (2021) for more insights on assumption (d) and possible alternatives.

Proposition 1.2.6. *If assumptions (a)-(d) hold, the natural mediational effects are identified and are given by*

$$\begin{aligned} NDE &= \mathbb{E}[Y(x, M(x^*)) - Y(x^*, M(x^*))] \\ &= \sum_m \{\mathbb{E}[Y \,|\, x, m] - \mathbb{E}[Y \,|\, x^*, m]\} \, P(m \,|\, x^*) \end{aligned} \tag{1.19}$$

$$\begin{aligned} NIE &= \mathbb{E}[Y(x, M(x)) - Y(x, M(x^*))] \\ &= \sum_m \mathbb{E}[Y \,|\, x, m] \, \{P(m \,|\, x) - P(m \,|\, x^*)\} \, . \end{aligned} \tag{1.20}$$

Proof. First, we prove that $\mathbb{E}[Y(x, M(x^*))]$ is identified.

$$\begin{aligned} &\mathbb{E}[Y(x,M(x^*))] \\ &\quad = \sum_m \mathbb{E}[Y(x, m) \,|\, M(x^*) = m] P(M(x^*) = m) \qquad \text{by iterated expectations} \\ &\quad = \sum_m \mathbb{E}[Y(x, m)] P(M(x^*) = m \,|\, x^*) \qquad \text{by (d) and (c)} \\ &\quad = \sum_m \mathbb{E}[Y(x, m) \,|\, x] P(M = m \,|\, x^*) \qquad \text{by (a) and consistency} \\ &\quad = \sum_m \mathbb{E}[Y(x, m) \,|\, x, m] P(M = m \,|\, x^*) \qquad \text{by (b)} \\ &\quad = \sum_m \mathbb{E}[Y \,|\, x, m] P(M = m \,|\, x^*) \qquad \text{by consistency.} \end{aligned}$$

From this formula, it is easy to derive $\mathbb{E}[Y(x, M(x))]$ and $\mathbb{E}[Y(x^*, M(x^*))]$, and, as a consequence

$$NDE = \sum_m \{\mathbb{E}[Y \,|\, x, m] - \mathbb{E}[Y \,|\, x^*, m]\} \, P(m \,|\, x^*)$$

$$NIE = \sum_m \mathbb{E}[Y \,|\, x, m] \, \{P(m \,|\, x) - P(m \,|\, x^*)\} \, . \qquad \square$$

Expression (1.20) is known as mediational g-formula (Pearl 2012). If (a)-(d) hold, along with Normality of variables and linearity of models (1.3) - (1.4), it can be proved (VanderWeele 2015) that mediational effects in the counterfactual framework are also

parametrically identified and given by

$$NDE = \gamma_X(x - x^*); \qquad NIE = \beta_X \gamma_M (x - x^*). \tag{1.21}$$

Notice that these effects are equal to the path-analytic ones if X is binary or x and x^* differ of one unit. The formulas in (1.21) become much more complicated if linearity of regression models cannot be assumed, for example, if interactions are present or regression models require a link function different from identity (VanderWeele 2015).

So far, we have written assumptions not conditioning on any additional covariate and we have found expressions for marginal mediational effects, i.e. not conditional on a set of covariates C. Clearly, both assumptions (a)-(d) and formulas (1.19)-(1.20) can be rewritten conditioning on C.

As already mentioned, assumptions (a)-(d) are only sufficient for identifying causal mediational effects, not necessary. Shpitser and VanderWeele (2011) proposed a complete graphical criterion for the identifiability of natural mediational effects, which links assumptions (a)-(d) to the adjustment criterion in Definition 1.2.8.

Theorem 1.2.7. *On any causal DAG $\mathcal{G}$, assumptions (a)-(d) hold for all NPSEMs inducing $\mathcal{G}$ if and only if C satisfies the adjustment criterion relative to $(X \cup M, Y)$ and to (X, M).*

Theorem 1.2.8. *The adjustment formula for natural direct and indirect effects holds if and only if C satisfies the adjustment criterion relative to $(X \cup M, Y)$ and to (X, M).*

The last result makes perfect sense, since

$$P(Y(x, M(x^*)) \,|\, C) = \sum_m P(Y(x, m) \,|\, C) P(M(x^*) \,|\, C)$$

by (d) and

1. if C satisfies the adjustment criterion relative to $(X \cup M, Y)$, then

$$P(Y(x, m) \,|\, C = c) = P(Y \,|\, X = x,\, M = m,\, C = c),$$

2. if C satisfies the adjustment criterion relative to (X, M), then

$$P(M(x^*) \,|\, C = c) = P(M \,|\, X = x^*,\, C = c).$$

In other words, since the adjustment criterion relative to $(X \cup M, Y)$ and (X, M) makes the two terms in the product $P(Y(x, m) \,|\, C) P(M(x^*) \,|\, C)$ separately identified, then their product, i.e. the interventional distribution $P(Y(x, M(x^*)) \,|\, C)$ is identified

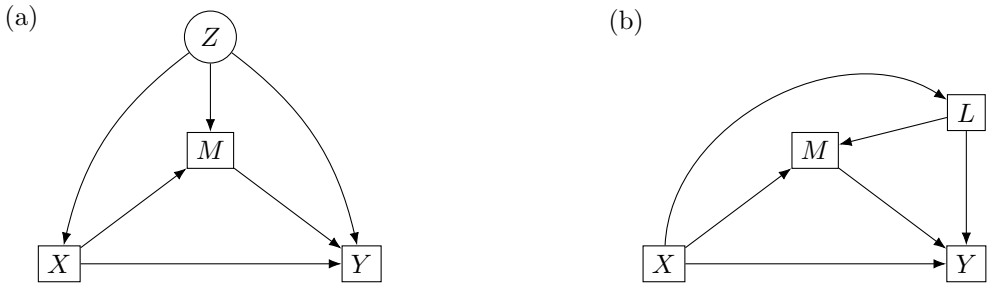

Figure 1.9: Example of DAGs where natural mediational effects are not identified.

as well. Notice that the adjustment criterion allows to identify the entire distribution, not just the expectation required to estimate the natural effects.

Figure 1.9 shows two DAGs which violate assumptions (a)-(d). In the left panel, Z is unobserved and confounds the exposure-mediator, exposure-outcome and mediator-outcome relationship, so it violates (a)-(c), while in the right panel L violates assumption (d), since it confounds the mediator-outcome relationship and is affected by the exposure. Confounders like L are called exposure-induced or post-treatment confounders, and they are very frequent in observational studies, especially in longitudinal ones, as we will see in the next chapter.

Note that L violates (d) even when it is observed. Indeed, either conditioning and non-conditioning induces bias and makes the indirect effect of X on Y through M unidentifiable. Conditioning on L blocks a part of the direct effect of X on Y, i.e. the path $X \to L \to Y$. Not conditioning, however, makes the mediator-outcome relationship confounded. Then, any choices lead to bias. VanderWeele et al. (2014) propose three methods to overcome the problem.

The first solution is to consider the couple (L, M) jointly as the mediator. Let $L(x)$ denote the value of L under an intervention setting X to x and $Y(x, l, m)$ the value of the outcome if X have been set to x, M to m and L to l. The natural direct and indirect effects become

$$NDE = \mathbb{E}[Y(x, L(x^*), M(x^*)) - Y(x^*, L(x^*), M(x^*))]$$
$$NIE = \mathbb{E}[Y(x, L(x), M(x)) - Y(x, L(x^*), M(x^*))],$$

where the direct effect is the effect of X on Y not through either M or L, while the indirect effect is the effect through L, M or both. Their identification requires a modification of assumptions (a)-(d), which can be rewritten as follows

(a*) $Y(x, l, m) \perp\!\!\!\perp X$

(b*) $Y(x, l, m) \perp\!\!\!\perp (L, M) \,|\, X$

(c*) $(L(x), M(x)) \perp\!\!\!\perp X$

(d*) $Y(x, l, m) \perp\!\!\!\perp (L(x^*), M(x^*))$

Notice that in Figure 1.9, assumption (d*) is satisfied, since there are no confounders of the L-M relationship affected by X. If (a*)-(d*) hold, then

$$NDE = \sum_{l,m} \{\mathbb{E}[Y \mid x, l, m] - \mathbb{E}[Y \mid x^*, l, m]\} P(l, m \mid x^*)$$
$$NIE = \sum_{l,m} \mathbb{E}[Y \mid x, l, m] \{P(l, m \mid x) - P(l, m \mid x^*)\}.$$

The other two methods entail changing the target of inference: with path-specific effects (Avin et al. 2005), it is possible to identify some of the paths from X to Y, while interventional effects require weaker assumptions than natural effects, and allow for exposure-induced confounders. We will see each approach in turn.

Path-specific effects

Consider Figure 1.9 (b). The cross-world counterfactual $Y(x, M(x^*))$ can be written as $Y(x, L(x), M(x^*, L(x^*)))$, which contains two counterfactual expressions for L with X set to different values. This feature is what prevents the identification of the indirect effect of X on Y through M in such settings: L should block the path $X \to L \to Y$, since it conveys a part of the direct effect, but at the same time it should not block the path $X \to L \to M \to Y$, which is part of the indirect effect, so L would assume its natural value under $X = x$. Clearly, L cannot be a blocking and a non-blocking node at the same time, for this reason L is called a *recanting witness*.

Definition 1.2.9 (Recanting witness)**.** Consider a DAG $\mathcal{G}(\boldsymbol{V})$ and two sets of nodes $\boldsymbol{X}, \boldsymbol{Y} \subset \boldsymbol{V}$. Let π denote a set of proper causal paths between $\boldsymbol{X}$ and $\boldsymbol{Y}$ and $\bar{\pi}$ its complement. A node $L \in \text{ch}(X)$, $X \in \boldsymbol{X}$, is called a recanting witness for π if there exist a path of the form $X \to L \to \cdots \to Y$ in π and a path $X \to L \to \cdots \to Y'$ in $\bar{\pi}$, with $Y, Y' \in \boldsymbol{Y}$ not necessarily coincident.

Avin et al. (2005) proved that, in a NPSEM without hidden variables, if there are no recanting witnesses for a path π, the π-specific effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ can be identified and the distribution of the corresponding counterfactual, denoted by $Y(\pi, \boldsymbol{x}, \boldsymbol{x}^*)$, $Y \in \boldsymbol{Y}$, can be expressed as

$$P(Y(\pi, \boldsymbol{x}, \boldsymbol{x}^*) = y) = \sum_{\boldsymbol{V} \setminus \{\boldsymbol{X} \cup \boldsymbol{Y}\}} \prod_{V_k \in \boldsymbol{V} \setminus \boldsymbol{X}} P(V_k = v_k \mid \text{pa}_k^{\pi} \cap \boldsymbol{X} = \boldsymbol{x}, \text{pa}_k^{\bar{\pi}} \cap \boldsymbol{X} = \boldsymbol{x}^*, \text{pa}_k \setminus \boldsymbol{X}) \quad (1.22)$$

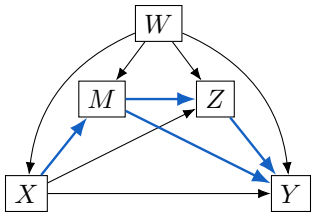

Figure 1.10: Example of identification of path-specific effects. The paths of interest are in blue.

where pa_k^π is the set of parents of V_k along paths in π. Formula (1.22) is known as *edge g-formula* (Shpitser and Tchetgen Tchetgen 2016) and it basically says that, for each vertex $V \in \boldsymbol{V}$, X can assume value x or x^* depending on whether the path $X \to V$ is in π or in its complement.

For example, consider the DAG in Figure 1.10 (Shpitser 2019). In such a graph there are no recanting witnesses for $\pi = \{X \to M \to Y,\, X \to M \to Z \to Y\}$, and the counterfactual distribution $P(Y(\pi,\, x,\, x^*))$ is identified as

$$\sum_{w,m,z} P(y \,|\, x^*, w, m, z) P(z \,|\, x^*, w, m) P(m \,|\, x, w) P(w).$$

In a graph like that in Figure 1.9 (b), L is a recanting witness which prevents the identification of path-specific effects relative to $\pi_D = \{X \to Y,\, X \to L \to Y\}$ and $\pi_I = \{X \to M \to Y,\, X \to L \to M \to Y\}$. Nonetheless, it is possible to identify the path-specific effects relative to $\pi_1 = X \to Y$, $\pi_2 = X \to M$ and $\pi_3 = \{X \to L \to M \to Y,\, X \to L \to Y\}$. In counterfactual terms, they can be expressed as

$$\begin{aligned}
\mathbb{E}_{\pi_1} &= \mathbb{E}[Y(x, L(x^*), M(x^*)) - Y(x^*, L(x^*), M(x^*))] \\
\mathbb{E}_{\pi_2} &= \mathbb{E}[Y(x, L(x^*), M(x, L(x^*))) - Y(x^*, L(x^*), M(x^*))] \\
\mathbb{E}_{\pi_3} &= \mathbb{E}[Y(x, L(x), M(x)) - Y(x, L(x^*), M(x, L(x^*)))]
\end{aligned}$$

and, under assumptions (a*)-(d*) they are identified as

$$\begin{aligned}
\mathbb{E}_{\pi_1} &= \sum_{l,m} \{\mathbb{E}[Y \,|\, x, l, m] - \mathbb{E}[Y \,|\, x^*, l, m]\}\, P(l, m \,|\, x^*) \\
\mathbb{E}_{\pi_2} &= \sum_{l,m} \mathbb{E}[Y \,|\, x, l, m]\, \{P(m \,|\, x, l) - P(m \,|\, x^*, l)\}\, P(l \,|\, x^*) \\
\mathbb{E}_{\pi_3} &= \sum_{l,m} \mathbb{E}[Y \,|\, x, l, m] P(m \,|\, x, l)\, \{P(l \,|\, x) - P(l \,|\, x^*)\}\,.
\end{aligned}$$

The results proved so far hold true in DAGs with only observed variables. When a DAG includes also latent variables, some difficulties arise and a new kind of graphs

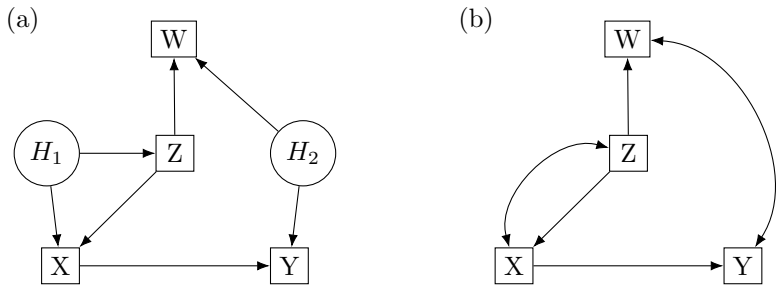

Figure 1.11: A DAG with unobserved variables and its latent projection.

is required.

Definition 1.2.10 (ADMG)**.** An acyclic directed mixed graph (ADMG) is a triple $\{\boldsymbol{V}, \boldsymbol{E}, \boldsymbol{B}\}$, where $\{\boldsymbol{V}, \boldsymbol{E}\}$ is a DAG and $\boldsymbol{B}$ is a collection of unordered pairs of vertices, known as bidirected edges.

Definition 1.2.11 (Latent projection)**.** Let $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$ a DAG with a set of observed nodes $\boldsymbol{V}$ and unobserved nodes $\boldsymbol{H}$. The latent projection of $\mathcal{G}$ is the induced subgraph $\mathcal{G}' = \mathcal{G}(\boldsymbol{V})$ obtained replacing edges of the form $V_i \leftarrow U \rightarrow V_j, V_i \leftrightarrow U \rightarrow V_j$, with $V_i, V_j \in \boldsymbol{V}$, $U \in \boldsymbol{H}$, in $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$ with bidirected edges $V_i \leftrightarrow V_j$.

Figure 1.11 shows a DAG $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$, with $\boldsymbol{V} = \{W, X, Y, Z\}$ and $\boldsymbol{H} = \{H_1, H_2\}$ (a), and its latent projection (b).

Definition 1.2.12. A bidirected path is a path containing only bidirected edges.

Definition 1.2.13 (District)**.** In an ADMG $\mathcal{G}$, the district of a node Z, denoted by $Dis_{\mathcal{G}}(Z)$, is the set of nodes reachable from Z via bidirected paths.

The set of observed nodes of a DAG $\mathcal{G}$ can then be partitioned into disjoint districts. This allows us to factorise the distribution of observed variables as follows. Let $\mathcal{D}(\mathcal{G})$ denote the set of all districts in the latent projection of $\mathcal{G}$. The marginal distribution of observed variables in $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$ is

$$P(\boldsymbol{V}) = \prod_{\boldsymbol{S} \in \mathcal{D}(\mathcal{G})} Q[\boldsymbol{S}], \tag{1.23}$$

where:

$$Q[\boldsymbol{S}] = \sum_{\mathrm{pa}_{\boldsymbol{H}}(\boldsymbol{S})} \prod_{V \in \boldsymbol{S}} P(V = v \,|\, \mathrm{pa}_{\boldsymbol{V}}(V), \mathrm{pa}_{\boldsymbol{H}}(V)) P(\mathrm{pa}_{\boldsymbol{H}}(\boldsymbol{S})), \tag{1.24}$$

$\mathrm{pa}_{\boldsymbol{V}}(V)$ and $\mathrm{pa}_{\boldsymbol{H}}(V)$ denote the set of observed and unobserved parents of V in $\mathcal{G}(\boldsymbol{V} \cup \boldsymbol{H})$, respectively, and the factors $Q[\boldsymbol{S}]$ are called kernels.

District factorisation extends Markov factorisation to DAGs including unobserved variables. Consider the simple DAG in Figure 1.12, where there are two districts,

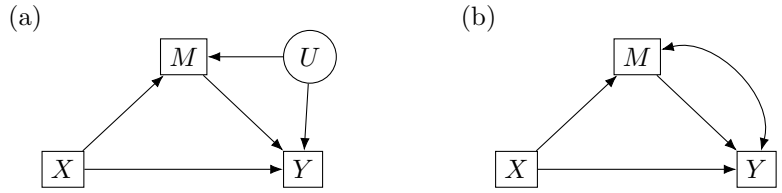

Figure 1.12: Example of the kernel factorisation.

$\{X\}$ and $\{M, Y\}$. The joint density over the observed variables factorises as

$$P(x, m, y) = P(Q[\{X\}])P(Q[\{M, Y\}]) = P(x) \underbrace{\sum_u P(y \mid x, m, u)P(m \mid x, u)P(u)}_{P(y \mid x, m)P(m \mid x)}$$

Analogously to what happens in DAGs with only observed variables, the counterfactual distribution $P(Y(x))$ can be expressed as a truncated factorisation of districts (Steen and Vansteelandt 2019)

$$P(Y(x) = y) = \sum_{\{v \mid V \in \boldsymbol{Y}^* \setminus Y\}} \prod_{\mathbf{D} \in \mathcal{D}^*} P(v \mid \text{pa}(\mathbf{D}) \cap X = x, \text{pa}(\mathbf{D}) \setminus X) \tag{1.25}$$

where $\boldsymbol{Y}^* = \text{an}(Y)$ in the subgraph $\mathcal{G}'(\boldsymbol{V} \setminus X)$ and $\mathcal{D}^* = \mathcal{D}(\mathcal{G}'_{\boldsymbol{Y}^*})$. Notice that the product runs across all districts $\mathbf{D} \in \mathcal{D}^*$ and the sum over all the possible values assumed by nodes in such districts, except for the outcome.

Considering again the graph in Figure 1.12, $\boldsymbol{Y}^* = \{M, Y\}$, which is also the only proper district in the latent projection shown in (b). $P(Y(x) = y)$ is then identified and given by

$$\sum_{u,m} P(y \mid x, m, u)P(m \mid x, u)P(u) = \sum_m P(y \mid x, m)P(m \mid x) = P(y \mid x).$$

In contrast, the counterfactual distribution $P(Y(x, M(x^*)))$ cannot be identified, since M and Y are in the same district and X should be set to x for path $\pi = X \to Y$, to x^* for path $\bar{\pi} = X \to M \to Y$. Shpitser (2013) extended the recanting witness criterion to graphs with unobserved variables, showing that cross-world distributions like $P(Y(x, M(x^*)))$ can be identified only in the absence of a particular kind of districts.

Definition 1.2.14 (Recanting district)**.** Consider an ADMG $\mathcal{G}$ and a set of proper causal paths π from a set of nodes $\boldsymbol{X}$ to a set $\boldsymbol{Y}$. Let $\boldsymbol{Y}^* = \text{an}(\boldsymbol{Y})$ in $\mathcal{G}(\boldsymbol{V} \setminus \boldsymbol{X})$. A district $\boldsymbol{D}$ in the ADMG $\mathcal{G}_{\boldsymbol{Y}^*}$ is said *recanting* for π if there exist Z_i, Z_j (possibly $Z_i = Z_j$) in $\boldsymbol{D}$, such that there is $X \to Z_i \to \cdots \to Y$ in π and $X \to Z_j \to \cdots \to Y$

in $\overline{\pi}$.

Theorem 1.2.9. *Let $\mathcal{G}$ be an ADMG. Let $\boldsymbol{X}, \boldsymbol{Y}$ be sets of nodes in $\mathcal{G}$, and π a subset of proper causal paths which start with a node in $\boldsymbol{X}$ and end with a node in $\boldsymbol{Y}$ in $\mathcal{G}$. Then the π-specific effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is expressible as a functional of interventional densities if and only if there does not exist a recanting district for this effect.*

A path-specific counterfactual distribution can be then expressed as the product

$$P(Y(\pi,\boldsymbol{x},\boldsymbol{x}^*)=y) = \sum_{V\in\boldsymbol{Y}^*\setminus\boldsymbol{Y}}\prod_{\boldsymbol{D}\in\mathcal{D}^*} P(V(\mathrm{pa}^{\pi}(\boldsymbol{D})\cap\boldsymbol{X}=\boldsymbol{x},\mathrm{pa}^{\overline{\pi}}(\boldsymbol{D})\cap\boldsymbol{X}=\boldsymbol{x}^*,\mathrm{pa}(\boldsymbol{D})\setminus\boldsymbol{X})) \tag{1.26}$$

where $\mathcal{D}^* \equiv \mathcal{D}(\mathcal{G}_{\boldsymbol{Y}^*})$. Notice that the product terms are counterfactuals. Their identifiability is not in general guaranteed due to the presence of latent variables, but, if $P(Y(x)=y)$ is identifiable through the algorithm proposed by Tian and Pearl (2003), it logically follows that also the kernels involved in (1.26) are identified and the counterfactual distributions can be expressed in terms of observed variables. This result is stated in the following theorem

Theorem 1.2.10. *Let $\mathcal{G}$ be an ADMG. Let $\boldsymbol{X}, \boldsymbol{Y}$ be sets of nodes in $\mathcal{G}$, and π a subset of proper causal paths which start with a node in $\boldsymbol{X}$ and end with a node in $\boldsymbol{Y}$ in $\mathcal{G}$. Assume there does not exist a recanting district for the π-specific effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$. Then the counterfactual representing the π-specific effect of $\boldsymbol{X}$ on $\boldsymbol{Y}$ is expressible in terms of the observed data if and only if the total effect $P(Y(x))$ is identifiable. Moreover, the functional expressing the counterfactual is obtained from Equation (1.26) by replacing each interventional term by a functional of the observed data identifying that term given by Tian's identification algorithm (Tian and Pearl 2003, Shpitser and Pearl 2008).*

Then, the recanting district criterion allows for the passage from cross-world densities to single world interventional densities, which can subsequently be identified using Tian and Pearl (2003) algorithm.

Figure 1.13 shows the same graph in Figure 1.10 with the inclusion of a latent variable U affecting Z and Y (a) and its latent projection (b). The π-specific effect relative to path $\pi = X \to M \to Y$, in blue, is identified. Indeed, $\boldsymbol{Y}^* = \{W, X, M, Z\}$ and $\mathcal{D}^* = \mathcal{D}(\mathcal{G}_{\boldsymbol{Y}^*}) = \{\{W\}, \{M\}, \{Z, Y\}\}$. There are no recanting districts relative to π, since M is the only element in its district. Since, in addition, the distribution $P(Y(x))$ is identifiable, the π-specific effect of X on Y can be identified as well and

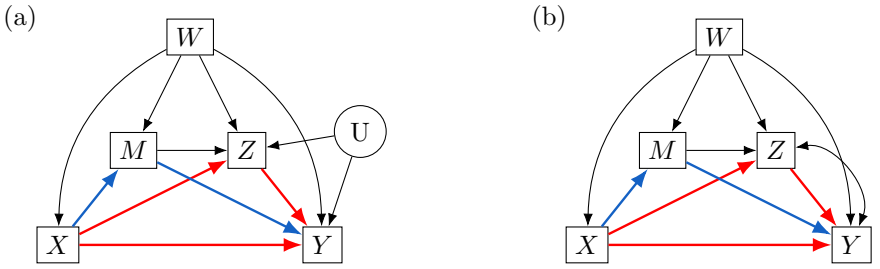

Figure 1.13: Example of identification of path-specific effects in the presence of a latent variable. The path-specific effect in blue is identified, the one in red shows why π' cannot be identified.

it is given by

$$P(Y(\pi, x, x^*)) = \sum_{w,z,m} P(Y \mid x^*, w, m, z)P(z \mid x^*, w, m)P(m \mid x, w)P(w).$$

In contrast, if one is interested in the effect along paths $\pi' = \{X \to Y,\ X \to M \to Y\}$, the π'-specific effect is not identified, since $X \to Y$ is in π, $X \to Z \to Y$ is in $\bar{\pi}'$ (see paths in red), and Y and Z share the same district, which is then recanting.

Interventional effects

The third approach proposed by VanderWeele et al. (2014) to overcome the problems related to an exposure-induced confounder is to change the target of inference, so that its identification does not require cross-world assumptions. This is the idea behind *interventional effects* (Didelez et al. 2006, Geneletti 2007)

Consider again Figure 1.9(b) and assume there are other observed covariates Z. Let $m^d(x|z)$ denote a random draw (the d superscript stands for draw) from the distribution of the mediator with exposure fixed at x conditional on $Z = z$. The *interventional* direct and indirect effects can be defined as

$$IDE = \mathbb{E}[Y(x,\ m^d(x^*|z))] - \mathbb{E}[Y(x^*,\ m^d(x^*|z))] \tag{1.27}$$

$$IIE = \mathbb{E}[Y(x,\ m^d(x|z))] - \mathbb{E}[Y(x,\ m^d(x^*|z))], \tag{1.28}$$

respectively, where x and x^* are two values of X. The IDE is the difference between the outcome if X were set to two different values and the mediator were randomly drawn from the distribution of the population when the exposure is at its baseline, conditional on covariates. The IIE is the difference in the outcome setting the exposure to the same value and drawing the mediator from two different distributions, one with $X = x$ and the other one with $X = x^*$, conditioning on covariates. It is easy to see that the total effect $\mathbb{E}[Y(x,\ m^d(x|z))] - \mathbb{E}[Y(x^*,\ m^d(x^*|z))]$ can be written as

the sum of the two. Notice that this definition of mediational effects differs from the usual natural one since the mediator is not fixed at the subject-specific value it would take if X were set to x, but it is a draw from the distribution of the mediator among subjects with a particular exposure, conditional on certain values of the covariates.

VanderWeele et al. (2014) show that these effects are identified under

(a′) $Y(x,m) \perp\!\!\!\perp X|Z$

(b′) $Y(x,m) \perp\!\!\!\perp M|X,Z,L$

(c′) $M(x) \perp\!\!\!\perp X|Z$

only. Assumption (d) is not required anymore, and, for this reason, interventional effects can be identified also in causal models relying on weaker sets of assumptions than NPSEMs, like FFRCISTGs. If the causal structure in Figure 1.9 holds and additional observed confounders Z are included, the IDE and IIE are given by (VanderWeele et al. 2014)

$$IDE = \sum_{z,l,m} \{\mathbb{E}[Y|x,l,m,z]P(l\,|x) - \mathbb{E}[Y|x^*,l,m,z]P(l\,|x^*)\}\, P(m|x^*,z)P(z)$$

$$IIE = \sum_{z,l,m} \mathbb{E}[Y|x,m,l,z]P(l\,|x)\,\{P(m|x,z) - P(m|x^*,z)\}\, P(z).$$

The longitudinal version of interventional effects will be discussed in Section 2.3.4.

Separable effects

A different approach is that proposed by Robins and Richardson (2011). They address the issue of identifiability of natural direct effects in four different causal models: the agnostic model (Spirtes et al. 1993), the minimal counterfactual model (MCM, Robins and Richardson 2011), the FFRCISTG and the NPSEM by Pearl (2009). These models are listed from the less restrictive, which does not even assume the existence of counterfactuals, to the one requiring the most restrictive assumptions. An extensive discussion on each of these models goes beyond the scope of this section: see Robins and Greenland (1992), Robins (2003) and Robins and Richardson (2011) for further details.

The contribution by Robins and Richardson, relevant to our discussion, is the introduction of expanded graphs which avoids that the NDE definition depends on nested counterfactuals. The authors propose an alternative definition of mediational effects, based on expanded graphs, which make them identifiable in all four counterfactual models.

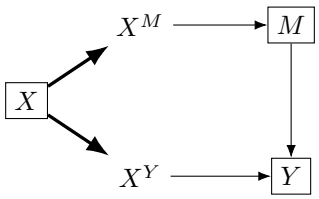

Figure 1.14: Separable effects via exposure decomposition.

Definition 1.2.15 (Expanded graph)**.** Given a DAG $\mathcal{G}$ with a single treatment or exposure variable X, the corresponding expanded graph $\mathcal{G}^{exp}$ for X is constructed by adding p new variables $\{X^1, X^2, \ldots, X^p\}$ representing separate components of X. In $\mathcal{G}^{exp}$, X has no other children than $\{X^1, X^2, \ldots, X^p\}$, and each $V_i \in \text{ch}(X)$ in $\mathcal{G}$ has a subset of $\{X^1, X^2, \ldots, X^p\}$ as parents in $\mathcal{G}^{exp}$.

Figure 1.14 shows a possible expanded graph for a simple mediation model, where we renamed X^1 and X^2 as X^M and X^Y, respectively. The bold arrows from X to its two components indicate a deterministic relationship, i.e. $X \equiv X^M \equiv X^Y$ in observed data.

In this framework, the cross-world parameter $\mathbb{E}[Y(x, M(x^*))]$, target of inference in NPSEMs and a non-manipulable quantity, is equivalent to $\mathbb{E}[Y(X^M = x^*, X^Y = x)]$, which is instead single-world and manipulable. As a consequence, the direct and indirect effects, which, using the terminology introduced by Stensrud et al. (2020), we will call *separable* (SDE and SIE), are defined, respectively, as follows

$$SDE = \mathbb{E}[Y(X^M = x^*, X^Y = x) - Y(X^M = x^*, X^Y = x^*)] \tag{1.29}$$

$$SIE = \mathbb{E}[Y(X^M = x, X^Y = x) - Y(X^M = x^*, X^Y = x)]. \tag{1.30}$$

and their sum yields the separable total effect $\mathbb{E}[Y(X^M = x, X^Y = x) - Y(X^M = x^*, X^Y = x^*)]$. Notice that these definitions do not involve any intervention on the mediator, and then, any cross-world intervention; thus, these effects can be identified also in FFRCISTGs. The identification of SDE and SIE can be proved by identifying $\mathbb{E}[Y(X^M = x^*, X^Y = x)]$, which is common to both effects. The identifiability of this expectation can, in turn, be proved starting from the counterfactual distribution $P(Y(X^M = x^*, X^Y = x, M(X^M = x^*, X^Y = x))) \equiv P_{x^M = x^*, x^Y = x}(y, m)$. Indeed,

$$\begin{aligned} P_{x^M = x^*, x^Y = x}(y, m) &= P(y \,|\, X^Y = x, m) P(m \,|\, X^M = x^*) \qquad \text{(g-formula)} \\ &= P(y \,|\, X^Y = x, X^M = x, m) P(m \,|\, X^Y = x^*, X^M = x^*) \\ &= P(y \,|\, X = x, m) P(m \,|\, X = x^*) \qquad \text{(by determinism)} \end{aligned}$$

where the second equality holds by the conditional independence relationships en-

coded in Figure 1.14, $Y \perp\!\!\!\perp X^M \,|\, X^Y, M$ and $M \perp\!\!\!\perp X^Y \,|\, X^M$.

It easily follows that

$$\begin{aligned}\mathbb{E}[Y(X^M = x^*, X^Y = x)] &= \sum_{y,m} y P_{x^M = x^*,\, x^Y = x}(y, m) \\ &= \sum_{y,m} y P(y \,|\, X = x, m) P(m \,|\, X = x^*) \\ &= \sum_{m} \mathbb{E}[Y \,|\, X = x, m] P(m \,|\, X = x^*).\end{aligned}$$

Robins et al. (2020) construct an interventionist theory for mediation analysis, generalising the results discussed by Robins and Richardson (2011). In particular, they clearly state which assumptions are required for identification of separable effects and their specific role. Consider the simplest case, shown in Figure 1.2, and the corresponding expanded graph in Figure 1.14. First, as we have already said, the assumption of determinism is crucial. This allows us to state the independence of $X^Y(x)$ from $X^M(x^*)$, $x \neq x^*$, as detailed in Robins and Richardson (2011), Robins et al. (2020, footnote 21). Second, Robins et al. (2020) show that two sets of assumptions are needed for non-parametric identification of the SDE (the same line of reasoning can be easily applied to SIE). The first set of assumptions, corresponding to Equations (14)-(15) in their paper, states that

- the distribution of $M(x^M, x^Y)$ does not depend on x^Y, i.e. $P(M(X^M = x, X^Y = x) = m) = P(M(X^M = x, X^Y = x^*) = m)$;
- the distribution $Y(x^M, x^Y) \,|\, M(x^M, x^Y)$ does not depend on x^M, i.e. $P(Y(X^M = x, X^Y = x) | M(X^M = x, X^Y = x)) = P(Y(X^M = x^*, X^Y = x) | M(X^M = x^*, X^Y = x))$.

They are necessary to identify

$$\begin{aligned}\mathbb{E}[Y(X^M = x^*, X^Y = x)] &= \sum_{m} \mathbb{E}[Y \,|\, X^Y = x, M = m] P(m \,|\, X^M = x^*) \\ &= \sum_{m} \mathbb{E}[Y \,|\, X = x, M = m] P(m \,|\, X = x^*)\end{aligned}$$

the last equality following from determinism. The second set of assumptions states that X^Y does not affect M directly, i.e. $M(x^M) = M(x, x^M, x^Y)$ and that X^M does not affect Y directly, i.e. $Y(x^Y, m) = Y(x, x^M, x^Y, m)$. Such assumptions are used to prove the equality $Y(x, M(x^*)) = Y(X^M = x^*, X^Y = x)$.

The separable effects approach can accommodate exposure-induced mediator- outcome confounders. Consider Figure 1.9(b). There are three possible expanded graphs

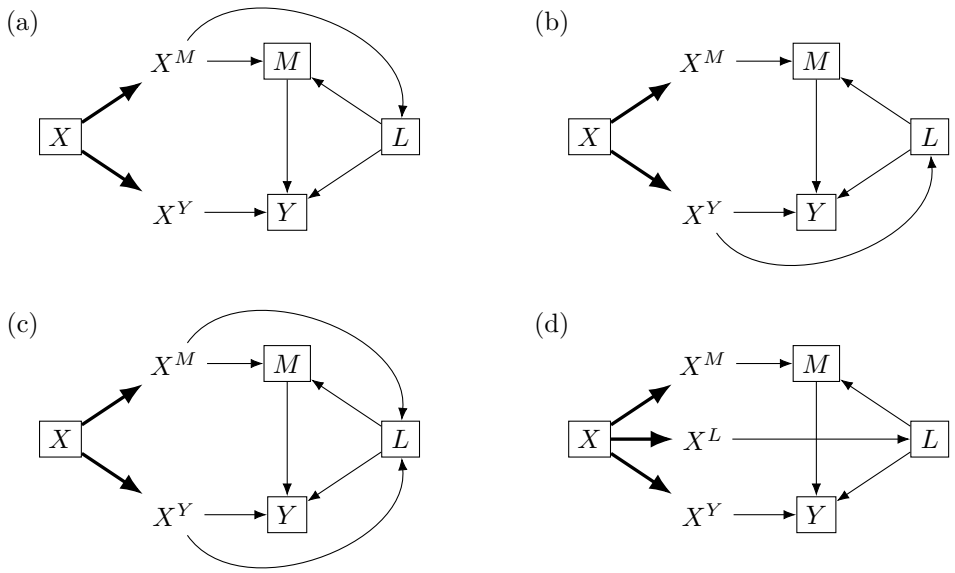

Figure 1.15: Expanded DAGs with an exposure-induced confounder.

compatible with this DAG, represented in 1.15(a-c). The parameter $\mathbb{E}[Y(x^Y = x, L(x^M = x), M(x^M = x^*, L(x^M = x^*))]$ is not identified under any causal model corresponding to these graphs, since, as we previously discussed, L is a recanting witness. Nonetheless, $\mathbb{E}[Y(x, L(x^*), M(x^*, L(x^*))]$, corresponding to the path $X \to Y$, and $\mathbb{E}[Y(x, L(x), M(x^*, L(x))]$, corresponding to the set of paths $\{X \to Y,\ X \to L \to Y,\ X \to L \to M \to Y\}$, can be identified, as we show below.

The parameter $\mathbb{E}[Y(x, L(x^*), M(x^*, L(x^*))]$ in DAG 1.9(b) corresponds to $\mathbb{E}[Y(x^Y = x, L(x^M = x^*), M(x^M = x^*, L(x^M = x^*))]$ in Figure 1.15(a). It is identified as

$$\mathbb{E}[Y(x^Y = x, L(x^M = x^*),\ M(x^M = x^*, L(x^M = x^*))] = \sum_{l,m} \mathbb{E}[Y \,|\, X = x, m, l] P(m \,|\, X = x^*, l) P(l \,|\, X = x^*) \tag{1.31}$$

since

$$\begin{aligned}
&P_{x^M = x^*,\, x^Y = x}(y, m, l) \\
&= P(y \,|\, X^Y = x, m, l] P(m \,|\, X^M = x^*, l) P(l \,|\, X^M = x^*) \\
&= P(y \,|\, X^Y = x, X^M = x, m, l] P(m \,|\, X^Y = x^*, X^M = x^*, l) P(l \,|\, X^Y = x^*, X^M = x^*) \\
&= P(y \,|\, X = x, m, l] P(m \,|\, X = x^*, l) P(l \,|\, X = x^*).
\end{aligned}$$

Analogously, the counterfactual $\mathbb{E}[Y(x, L(x), M(x^*, L(x))]$ in DAG 1.9(b) corresponds to $\mathbb{E}[Y(x^Y = x, L(x^Y = x), M(x^M = x^*, L(x^Y = x))]$ in Figure 1.15(b). It is

identified as

$$\mathbb{E}[Y(x^Y = x, L(x^M = x),\ M(x^M = x^*, L(x^M = x))] = \sum_{l,m} \mathbb{E}[Y \,|\, X = x, m, l] P(m \,|\, X = x^*, l) P(l \,|\, X = x) \tag{1.32}$$

since

$$\begin{aligned} &P_{x^M = x^*,\, x^Y = x}(y, m, l) \\ &= P(y \,|\, X^Y = x, m, l] P(m \,|\, X^M = x^*, l) P(l \,|\, X^Y = x) \\ &= P(y \,|\, X^Y = x, X^M = x, m, l] P(m \,|\, X^Y = x^*, X^M = x^*, l) P(l \,|\, X^Y = x, X^M = x) \\ &= P(y \,|\, X = x, m, l] P(m \,|\, X = x^*, l) P(l \,|\, X = x). \end{aligned}$$

In contrast, the counterfactual $\mathbb{E}[Y(X^M = x^*,\ X^Y = X)]$ associated to graph in 1.15(c) does not correspond to any counterfactual in DAG 1.9(b), since the counterfactual $L(X^M = x^*, X^Y = x)$ is equal to neither $L(X = x)$ nor $L(X = x^*)$. This follows from the next more general result (Robins et al. 2020):

Lemma 1.2.11. *Under an FFRCISTG corresponding to an expanded graph $\mathcal{G}^{exp}$ for X, the intervention distribution $P(V(x^1 = \tilde{x}^1,\ x^2 = \tilde{x}^2, \ldots, x^p = \tilde{x}^p))$ is identified by the g-formula from data on $\mathcal{G}$ if, for every $V_i \in \mathrm{ch}_{\mathcal{G}}(X)$, $\mathrm{pa}_{\mathcal{G}^{exp}}(V_i)$ belonging to the components of X take the same value.*

It should be noticed that scenarios (a), (b) and (c) are mutually exclusive, since they correspond to different causal hypotheses, indeed the identification formulas in (1.31)-(1.32) differ. An alternative scenario is shown in panel (d), where an additional component X^L of X is assumed to affect only L. By Lemma 1.2.11, it can be proved that, in such a scenario, both effects $\mathbb{E}[Y(x^Y = x,\ L(x^M = x^*),\ M(x^M = x^*, L(x^M = x^*))]$ and $\mathbb{E}[Y(x^Y = x, L(x^M = x), M(x^M = x^*, L(x^M = x))]$ are simultaneously identified through (1.31) and (1.32), respectively. The former parameter corresponds to $\mathbb{E}[Y(X^Y = x, X^L = x^*, X^M = x^*]$, where X^L and X^M are set to the same values, the latter to $\mathbb{E}[Y(x^Y = x, L(x^M = x), M(x^M = x^*, L(x^M = x))]$, where X^L and X^M are set to the same values.

Figure 1.15(d) is a special kind of expanded DAG, called *edge expanded graph* for X, where $p = |\mathrm{ch}_{\mathcal{G}}(X)|$. $\mathcal{G}^{edge}$ graphs are obtained by their corresponding DAGs $\mathcal{G}$ by replacing each path $X \to V_i$, $V_i \in \mathrm{ch}_{\mathcal{G}}(X)$, with $X \to X^i \to V_i$. Clearly, $\mathcal{G}^{edge}$ is unique for each $\mathcal{G}$.

Corollary 1.2.11.1. *Under the assumptions of Lemma 1.2.11, if $\mathcal{G}^{exp} = \mathcal{G}^{edge}$ for X, then, for every assignment $\tilde{x}^1, \ldots, \tilde{x}^p$, $P(V(x^1 = \tilde{x}^1, \ldots, x^p = \tilde{x}^p))$ is identified from data on $\mathcal{G}$.*

Principal stratification

To conclude this section, we briefly address another approach that, although will not be an object of further discussion, we believe is worth being mentioned, since it can shed some light on mediational mechanisms and does not require any intervention on the mediator, i.e. Principal stratification (Frangakis and Rubin 2002).

Consider a general setting with a binary treatment X whose values are denoted by 0 and 1 for simplicity, an outcome Y and a post-treatment covariate S, not necessarily regarded as a mediator. A *principal stratification* is a partition of statistical units such that elements in the same subset are characterised by the same values of $(S(0), S(1))$. The subsets in the partition are called *principal strata*. For example, in an experimental setting where X is a binary treatment to which each subject is randomly assigned by the researcher and S is a binary variable indicating the actual treatment assumed by the subject, there exist four principal strata: compliers $PS_{01} = \{i : S_i(0) = 0, S_i(1) = 1\}$, never-takers $PS_{00} = \{i : S_i(0) = 0, S_i(1) = 0\}$, always-takers $PS_{11} = \{i : S_i(0) = 1, S_i(1) = 1\}$, and defiers $PS_{10} = \{i : S_i(0) = 1, S_i(1) = 0\}$. Notice that principal strata are not affected by the treatment assignment, then they can be considered as pre-treatment covariates. However, it is impossible to know to which stratum each subject belongs, therefore, principal strata are latent.

A principal causal effect (PCE) is a comparison between $Y_i(1)$ and $Y_i(0)$ within a principal stratum. One of the most widely used PCE is the expectation

$$PCE_{s_0,s_1} = \mathbb{E}[Y(1)\text{–}Y(0) \,|\, S(0) = s_0, S(1) = s_1].$$

Frangakis and Rubin (2002) called the effects within strata where $s_0 = s_1$ 'dissociative', and the effects within strata where $s_0 \neq s_1$ 'associative'. Following Mealli and Mattei (2012) and VanderWeele (2011), we briefly discuss the relationship between these effects and natural effects, trying to extend their considerations also to the other effects introduced so far.

Let us start from dissociative effects, which are defined within strata in which S is not affected by the value of X. In other words, X does not impact the value taken by S and therefore it conveys its effect on Y directly. The principal strata direct effects (PSDEs) assume then the forms:

$$PSDE(0) = \mathbb{E}[Y(1)\text{–}Y(0) \,|\, S(0) = S(1) = 0]$$
$$PSDE(1) = \mathbb{E}[Y(1)\text{–}Y(0) \,|\, S(0) = S(1) = 1].$$

Since the concept of direct effect is well defined in the principal stratification frame-

work and relies on dissociative effects, one may think that also the concept of indirect effect has an analogous in this framework, relying on associative effects. Although associative effects are defined within those strata where the value of $S(0)$ differs from that of $S(1)$, i.e. where the treatment does affect the intermediate variable, they do not encode the concept of indirect effect. As proved by VanderWeele (2011), associative effects can be written as the sum of both natural direct and indirect effects within a stratum.

The lack of a principal stratification analogous of the NIE has, as one of the first consequences, the impossibility to decompose the total causal effect of X on Y into the sum of a direct and indirect effect. Mealli and Mattei (2012) show that the total effect can be written as a weighted average of the PCE over principal strata, but this decomposition differs from those we have seen to hold for all the other kinds of mediational effects introduced so far. Another difference between the principal stratification framework and all the aforementioned ones is that PCEs are local, in the sense that they are defined for subgroups of the population, while natural, interventional and separable effects are population effects.

Among the mediational effects introduced so far, separable effects may seem the most similar to PCEs, since both are defined without requiring an intervention on the mediator; however, they also present some key differences. First, the concept of indirect effect is well defined only for the former. For PCEs, this implies the absence of a direct correspondence with natural effects: as we saw, there exists a PSDE, but, as discussed in Mealli and Mattei (2012), it is only one of the terms in the expression for the NDE. Indeed, a null PSDE does not imply a null NDE, while the converse holds true (VanderWeele 2008). In contrast, there exists a direct correspondence between natural and separable effects, since the latter prove equal to the former in an NPSEM.

Second, as already pointed out, separable and principal effects rely on different conceptual frameworks. The former entails a split of the treatment into components which are regarded as independent intervening variables and this helps in understanding the different pathways through which the direct and indirect effects flow. In contrast, principal stratification does not involve a separation of the treatment, but rather, the estimation of effects in subgroups of the sample under investigation.

Finally, a note on identifiability assumptions. Separable effects require some assumptions to be identified from observed data which have already been discussed and differ from those required by PCEs, briefly addressed in Mealli and Mattei (2012). Generally, PCEs are difficult to identify, since the stratum to which a subject belongs is latent and even restrictive assumptions leads only to partial identification, unless

one is willing to make distributional assumptions as well. A Bayesian approach seems, nonetheless, a promising approach to deal with such issues and carry out sensitivity analyses.

As remarked by VanderWeele (2011), principal stratification does not seem the most appropriate framework to address mediation. Nonetheless, we agree with Mealli and Mattei (2012), who claim that principal stratification can be useful in some circumstances and can still provide information about mediational settings. For example, they show that in very specific cases it is possible to derive the NIE as the difference between the total effect and the NDE, the latter obtained as weighted sum of PSDEs. More generally, comparing the magnitude of associative and dissociative effects can also provide some insights into mediational questions. Indeed, when associative effects are larger than dissociative ones, this indicates that the treatment has a more substantial impact on the outcome for those units for which the treatment also affects the mediator. Conversely, associative and dissociative effects of comparable magnitude suggest that the treatment has the same effect on the outcome regardless of the mediator's level, then its effect is conveyed by other variables different from the mediator under investigation.

1.3 Structure of the thesis

The thesis is divided into four chapters structured as follows.

The second chapter provides a literature review on longitudinal mediation analysis, addressing associational and causal approaches. We give particular relevance to model specification and the definition of mediational effects. This aspect has a crucial role in the causal framework, and we discuss some problems which may arise in a longitudinal setting.[3]

The third chapter focuses on associational models. We show how SEMs and multilevel models can be regarded as particular cases of a more general and unique model. We explain how the special features of SEMs can be exploited in longitudinal, and more general multilevel, mediational contexts through the development of an approach based on definition variables. The implications of such unification are discussed.[4]

In Chapter 4 we apply the separable effect approach to two popular latent variable

[3]Co-authors' individual contributions: Chiara Di Maria - Conceptualisation, Resources, Writing - Original draft, Reviewing and Editing; Antonino Abbruzzo - Supervision, Writing - Reviewing and Editing; Gianfranco Lovison - Supervision, Writing - Reviewing and Editing.

[4]Co-authors' individual contributions: Chiara Di Maria - Conceptualisation, Methodology, Writing - Original draft, Reviewing and Editing; Antonino Abbruzzo - Supervision, Writing - Reviewing and Editing; Gianfranco Lovison - Supervision, Writing - Reviewing and Editing.

models for longitudinal data, i.e. multilevel and latent growth models. We derive assumptions and g-formulas for the separable effects, and conduct a simulation study to evaluate how model misspecification can impact the estimates. Extensions to more complex settings including baseline and time-varying confounders are discussed.[5]

Chapter 5 provides a real-world application of some the approaches discussed in the previous chapters. We analyse data from the COVCO study, a longitudinal cohort study carried out in Basel during the Covid pandemic with the goal of monitoring the spread of the virus in Basel and population mental health. Within a separable effects approach, we investigate the direct effects of income on depression and its indirect effects mediated by worries concerning different aspects of life, in a phase of acute spreading of Covid-19.[6]

The thesis concludes with a summary of the main results and a discussion of future directions.

[5]Co-authors' individual contributions: Chiara Di Maria - Methodology, Software, Writing - Original draft, Reviewing and Editing; Vanessa Didelez - Conceptualisation, Supervision, Writing - Reviewing and Editing.

[6]Co-authors' individual contributions: Chiara Di Maria - Software, Formal analysis, Writing - Original draft, Reviewing and Editing; Antonino Abbruzzo - Supervision, Writing - Reviewing and Editing; Gianfranco Lovison - Supervision, Data curation, Writing - Reviewing and Editing.

Chapter 2

Literature Review

In this chapter, we discuss the main approaches proposed over the years to address longitudinal mediation analysis. In spite of the intrinsically causal nature of mediation, a great stream of literature assesses mediational mechanisms in associational terms. This is not surprising, since the origins of mediation analysis are rooted in regression-based methods which were not intended to provide causal explanations to phenomena. The constant development and refinement of causal inference techniques to deal with increasingly complex problems led to the development of a parallel literature addressing longitudinal mediation from a causal perspective.

Although we believe that mediation is related to causal mechanisms, at the same time we are firmly convinced of the noticeable role played by the associational literature in the development of mediation analysis. For this reason, this chapter provides a literature review addressing both associational and causal approaches for longitudinal mediation.

The chapter is divided into two main parts: in Section 2.2, we address associational models, i.e. SEMs and mixed-effect models, whereas Section 2.3 is devoted to causal approaches. Each section starts with the notation and the preliminary concepts to understand what follows, proceeds with the introduction of the relevant models or approaches, and ends with a discussion. Finally, in the last section, we draw some conclusions.

2.1 Introduction

For a long time, mediation analysis was primarily addressed in a cross-sectional setting. However, scientists have started warning against the use of cross-sectional data for detecting mediational effects since the early '80s. Judd and Kenny (1981) highlight the importance of longitudinal designs for studying mediation and emphasise

the bias that could occur from failing to control for prior assessments of the mediator and the outcome variable. Methodological articles on longitudinal mediation, with authors arguing that empirical investigations of mediation should take time into account, began to appear in the late 1990s (Collins et al. 1998, MacCallum and Austin 2000, Cole and Maxwell 2003, Kenny et al. 2003). However, as remarked by Maxwell et al. (2011), their cautions were and keep being largely ignored, not only by substantive researchers but also by methodologists. More recently, several authors (Maxwell and Cole 2007, Maxwell et al. 2011, O'Laughlin et al. 2018) have shown that cross-sectional designs fail to capture true mediational processes under numerous conditions. Maxwell and Cole (2007) focus on the case of complete mediation (i.e. null direct effect), considering two different models, an autoregressive model and a mixed-effect model, and show that the magnitude of bias in the estimates of mediational effects can be substantial in both cases. Maxwell et al. (2011) extend the analysis to partial mediation settings. Their findings are consistent with those of Maxwell and Cole (2007), since they prove that cross-sectional mediation[1] may lead to severely misleading results in terms of bias in the effect estimates. Furthermore, a variable appearing as a potential significant mediator in a cross-sectional analysis may turn out to be almost irrelevant in a longitudinal analysis or vice versa. Since the effect of the exposure on the mediator and the outcome, and that of the mediator on the outcome unfold over time, cross-sectional models are not able to detect the change and turn out to be misspecified. Thus, time plays a prominent role in the correct estimation of effects, and this is the reason why methodological developments of longitudinal mediation analysis appear of paramount importance.

Several approaches have been proposed to deal with longitudinal mediation analysis, and they can be divided into two main categories: discrete-time models and continuous-time models. Discrete-time mediation models form a wide class, including several models. Some are based on structural equation models (SEMs), like cross-lagged panel models (MacKinnon 2008, Chapter 8, Cole and Maxwell 2003), latent growth models (Cheong et al. 2003, von Soest and Hagtvet 2011) and latent difference score models (Selig and Preacher 2009, O'Laughlin et al. 2018), while others include random effects (Bauer et al. 2006, Bind et al. 2016). Another broad class of longitudinal mediation models deals with time-to-event outcomes (Lin, Young, Logan and VanderWeele 2017, Aalen et al. 2020, Vansteelandt et al. 2019). Other methods

[1] Throughout the thesis, we will use the expression "cross-sectional mediation analysis" referring to situations in which the data generating mechanism is cross-sectional. Although questionable, this expression is widely used and well established in the literature, see, for example, Maxwell and Cole (2007), O'Laughlin et al. (2018). By using the adjective *cross-sectional* we want to remark that the variables may refer to different periods, but they are collected at the same time and measured only once. This contrasts with a longitudinal design, where variables are measured at multiple occasions in time.

assume a dynamic perspective and consider variables as stochastic processes, such as dynamic path analysis (Fosen et al. 2006). Continuous-time models are more recent and the methodologies proposed are mainly based on derivatives (Ryan et al. 2018, Deboeck and Preacher 2016, Albert et al. 2019).

Reviews on the topic of longitudinal mediation analysis can already be found in the literature. Selig and Preacher (2009) and O'Laughlin et al. (2018) compare different SEMs focusing on model specification and definition of effects. The former analyse data using only the latent difference score model, while the latter provide an empirical application to exemplify all described methods, similarly to what was done by Goldsmith et al. (2017). Krull et al. (2016) give a wide overview on many issues concerning mediation and moderation, describing popular models for longitudinal data and how some of them can accommodate mediation and moderation. They also discuss causal longitudinal mediation and some estimation methods proposed in the literature, although they do not provide a formal description of these topics.

The review we provide in this chapter differs from the previous ones, since it addresses a wide range of approaches, both associational and causal. The main focus lies on model specification and the definition and interpretation of mediational effects, while we do not deepen issues related to statistical inference, like estimation and hypothesis testing. Furthermore, for causal approaches we discuss assumptions required to make mediational effects (non parametrically) identifiable. We also discuss if it is possible to extend these conditions to associational models, to make them causally interpretable, and if they suffice for identifying direct and indirect effects. In addition, considering the variety of models discussed in this review, we give another contribution by unifying the diverse notation present in the literature.

Figure 2.1 provides a graphical overview of this chapter and Table 2.1, at the end of the chapter, summarises the approaches reviewed: for associational models, it provides information about their implementability/implementation in statistical software, if any simulation studies have been carried out and which applications have been made; for causal approaches, it shows the estimands of interest, in which counterfactual models they result identified and in which software they can be implemented.

2.2 Associational framework

This section focuses on two classes of associational models for longitudinal mediational analysis. They have a long tradition and are quite widespread among applied researchers: SEMs, which date back to Wright (1934) and Haavelmo (1943) and mixed-effect models (Henderson 1973, Laird and Ware 1982).

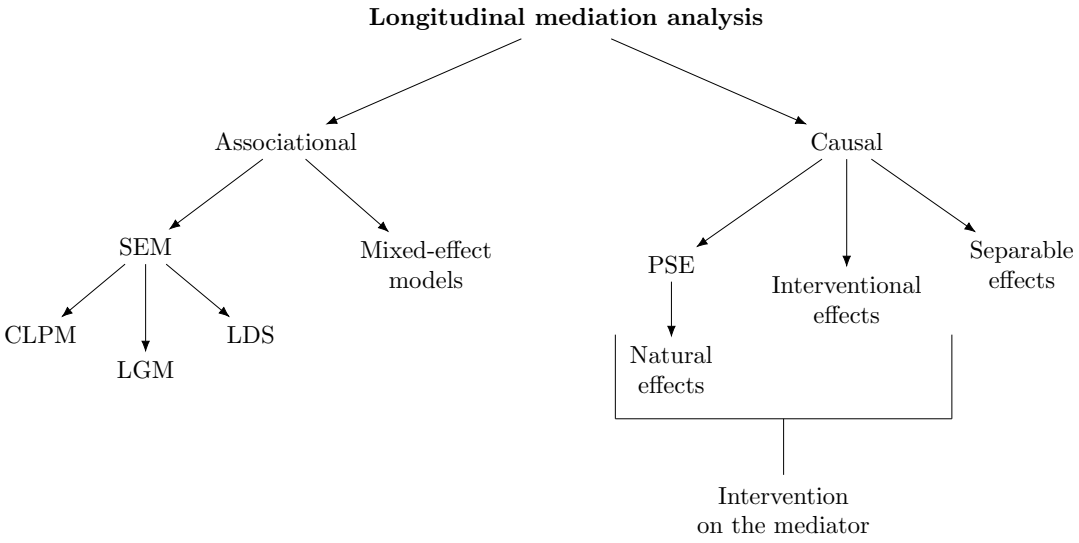

Figure 2.1: Summary of the approaches described in the review.

2.2.1 Basic concepts and notation

Here and throughout the chapter, if not differently specified, the exposure is denoted by X, the mediator by M and the response by Y. Generally, L and V indicate time-varying and time-fixed covariates, respectively. Since the main focus of this review is on longitudinal settings, the following notation is also introduced. The number of sample subjects is indicated by n, and variables are measured at T time occasions. Capital letters denote variables, lowercase letters their observed values. In particular, for any variable W, W_t is the variable at time t, assuming value w_t, while $W(t)$ denotes the variable considered as a stochastic process. $\overline{W}_t$ indicates the history of the variable up to time t, that is $(W_1, W_2, \ldots, W_t)$, $\forall\, t = 1, \ldots, T$. For $t = T$ we omit the subscript and indicate $\overline{W}_T$ simply by $\overline{W}$. For representing associational models, we will use path diagrams, where, as introduced in Chapter 1, if both observed and latent variables are present, the former will be represented as squares and the latter as circles. Arrows represent dependence relationships and are marked with the corresponding coefficients in the models. We do not include error terms, as traditionally used in the literature, to avoid clutter.

2.2.2 Structural Equation Models

In all the models introduced below we do not include covariates. This choice is motivated only by our desire to make the presentation as easy and clear as possible, but it is worth remarking that in real-world applications adjusting for covariates is necessary.

Cross-lagged panel models

A very commonly used model for longitudinal data is the *cross-lagged panel model* (CLPM). The most basic specification of the CLPM entails two variables measured at different time occasions, and each measure is assumed to depend both on itself and the other variable at previous times (Selig and Little 2012). The effects of a variable on itself at previous time occasions are called *autoregressive*, while *cross-lagged* effects refer to influences among different variables. The CLPM relies on three assumptions: *stability*, i.e. no difference between individuals; *stationarity*, that is, the within-person variances and covariance are invariant over time; *equilibrium*, i.e. the correlation among each couple of variables stays the same over time (MacKinnon 2008, Cole and Maxwell 2003, Preacher 2015).

Without loss of generality, consider a three-wave model for mediation analysis, i.e., a model where data are collected at three different time occasions, with three observed variables, one exposure, one mediator and one response. Variables at time $t = 1$ are treated as exogenous, and the model equations, for $t = 2$, 3, are

$$X_{it} = \alpha_0 + \alpha_X X_{it-1} + \varepsilon_{Xit} \tag{2.1}$$

$$M_{it} = \beta_0 + \beta_X X_{it-1} + \beta_M M_{it-1} + \varepsilon_{Mit} \tag{2.2}$$

$$Y_{it} = \gamma_0 + \gamma_X X_{it-1} + \gamma_M M_{it-1} + \gamma_Y Y_{it-1} + \varepsilon_{Yit}. \tag{2.3}$$

In this model, represented in Figure 2.2, the explanatory variable at time t depends on itself at the previous time $t-1$, the mediator on itself and on X at previous time, while the response depends on itself, X and M at time $t-1$. Notice that no cross-sectional effects are present, that is, there are no relationships between variables at the same time. This is just one of the possible CLPMs in a mediation setting: more generally, it is possible to consider two-unit lags or more cross-lagged relationships, for example, M_t may influence X_{t+1} (see MacKinnon 2008, Maxwell et al. 2011). Notice also that coefficients are not indexed, since they are assumed to be equal across subjects and time-invariant, although the last feature is not required and can be relaxed (Usami et al. 2019).

In this context there is not just a single mediated effect, but as many as the number of paths linking X to Y through M. Selig and Preacher (2009) distinguish the *time-specific* indirect effects from the *total* indirect effect. The former are the paths from X at a certain time occasion to Y at a subsequent time occasion passing through M, while the latter is the sum of all these paths. In a model like that in Figure 2.2 we can obtain the direct effect of X_t on Y_{t+1} as γ_X and the indirect effect of X_{t-1} on Y_{t+1}, mediated by M_t, as the product $\beta_X \gamma_M$. These effects concern observed variables and

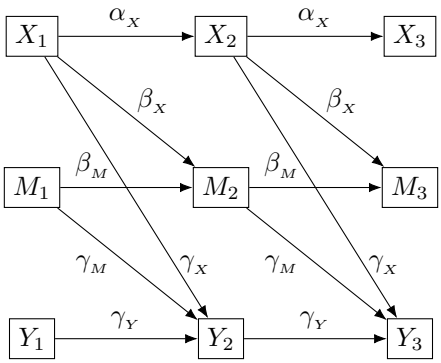

Figure 2.2: Cross-lagged panel mediation model for three waves.

have to be interpreted in the usual way: the direct effect as the extent to which a change in the exposure at time t directly affects the outcome at the subsequent time, the indirect effect as the extent to which a change in the exposure at time t affects the outcome measured two times later indirectly, through the mediator measured at $t+1$.

If a researcher suspects measurement errors in one or more variables, CLPMs can be applied to latent variables of which repeated measures are (fallible) indicators. In this formulation, the CLPM is generally addressed as *factor* CLPM (Usami et al. 2019) or *structural* CLPM (Ferrer and McArdle 2003), and each observed variable is split into two components, a latent true score and a measurement error. The outcome, for example, should be written as $Y_{it} = \tilde{Y}_{it} + \varepsilon_{Y_{it}}$. The autoregressive and cross-lagged relationships among variables shown in Equations (2.1) - (2.3) now involve the latent scores, so that, using the same notation also for the exposure and the mediator, the model can be written as follows

$$\tilde{X}_{it} = \alpha_{\tilde{X}} \tilde{X}_{it-1} + \delta_{\tilde{X}it} \tag{2.4}$$

$$\tilde{M}_{it} = \beta_{\tilde{X}} \tilde{X}_{it-1} + \beta_{\tilde{M}} \tilde{M}_{it-1} + \delta_{\tilde{M}it} \tag{2.5}$$

$$\tilde{Y}_{it} = \gamma_{\tilde{X}} \tilde{X}_{it-1} + \gamma_{\tilde{M}} \tilde{M}_{it-1} + \gamma_{\tilde{Y}} \tilde{Y}_{it-1} + \delta_{\tilde{Y}it}. \tag{2.6}$$

In this case the direct effect is given by $\gamma_{\tilde{X}}$ and the indirect effect by the product $\beta_{\tilde{X}}\gamma_{\tilde{M}}$, similarly to the case with only observed variables. However, the interpretation is different. Mediation happens at the latent score level, thus, the latent score of the exposure can affect directly the latent score of the response at a subsequent time, or indirectly through the latent score of the mediator. Observed variables are just indicators of the scores, which are latent, and if a mediational mechanism exists, it involves the latent scores.

The CLPM is commonly used in econometrics to analyse time-series, mainly due

to its simple implementation. However, it has several drawbacks. First, the time lag between measurements is not explicitly taken into account in the model. Second, since autoregressive and cross-lagged coefficients do not differ among subjects, this model is useful for interindividual change, but does not detect intraindividual change, so the selected variables need to show some degree of stability over time for each individual, a condition not always satisfied.

The CLPM can be implemented in STATA (sem), R (`lavaan`, `OpenMx` packages), SAS (CALIS procedure) and Mplus. We are not aware of any simulation study in a longitudinal mediation setting, whereas applications can be found in the behavioural and social sciences (Lee and Stone 2012, O'Laughlin et al. 2018).

❖ **Example** : Lee and Stone (2012) tried different CLPMs to study the relationship between internalising and externalising behavioral problems and the mediating role of negative self-concept in a sample of 2,844 Korean adolescents followed for four years. In psychology, internalising behaviours are a set of internally-focused behavioural symptoms including depression, anxiety and obsessive-compulsive disorder, while externalising disorders are a broad spectrum of behaviours expressing emotional distress, such as physical aggression, bullying and vandalism. Negative self-concept is the perception that an individual has of him/herself in negative terms, for example the extent to which he/she feels "wrong "or the degree of self-criticism.

The authors found that the relationship between externalising and internalising behaviours is entirely mediated by negative self-concept, so there are no significant direct effects, and the final model selected includes different cross-lagged effects, so it is more complex than that shown in Figure 2.2.

Latent growth curve models

The aim of *latent growth (curve) modelling* (LGC or LGM) is to estimate an underlying growth trajectory for each individual in the sample, highlighting "between-person differences in within-person change" (Curran et al. 2010). The basic (linear) latent growth model includes a response variable and two subject-specific factors, the *intercept factor*, constant for each individual over time, and the *slope factor*, representing the individual rate of change over time. Slightly modifying Bollen and Curran (2006) notation to make it consistent with our own, the *unconditional* (i.e. not including

explanatory variables) latent growth curve model can be written as follows:

$$\begin{aligned} Y_{it} &= \theta_{0i} + \theta_{1i}\lambda_t + \varepsilon_{Yit} \\ \theta_{0i} &= \mu_{\theta_0} + \zeta_{\theta_0 i} \\ \theta_{1i} &= \mu_{\theta_1} + \zeta_{\theta_1 i}, \end{aligned} \tag{2.7}$$

where Y_{it} is the value of the response variable for the i-th unit at time t, θ_{0i} and θ_{1i} are the intercept and the slope for the i-th observation, respectively, and are the sum of a fixed average (the μ's terms) and a stochastic component ζ, λ_t is a time indicator and ε represents the error term. Note that θ_0 and θ_1 may differ across subjects, allowing each individual to have his/her own trajectory. Moreover, they can depend on additional explanatory variables and in this case the LGM is called *conditional*. As regards λ, it is generally coded as a discrete variable, where $\lambda_1 = 0$ indicates the starting time point. Researchers can specify λ's fixing their values to assume a specific type of growth (e.g.: linear, quadratic), or they can estimate the λ's as free parameters in the model. LGMs rely on numerous assumptions, for further details see Duncan and Duncan (2004), Bollen and Curran (2006) and Preacher et al. (2008).

To extend LGM to mediational settings, it is necessary to fit two parallel growth processes, one for the mediator and one for the response, by using a structural equation modelling perspective. For each subject i, the processes can be written as

$$\begin{aligned} \mathbf{m}_i &= \boldsymbol{\Lambda}_m \boldsymbol{\theta}_{mi} + \boldsymbol{\varepsilon}_{mi} \\ \mathbf{y}_i &= \boldsymbol{\Lambda}_y \boldsymbol{\theta}_{yi} + \boldsymbol{\varepsilon}_{yi}, \\ \boldsymbol{\theta}_{mi} &= \boldsymbol{\beta}_0 + \mathbf{B}_x \mathbf{X} + \boldsymbol{\zeta}_{\theta_m i} \\ \boldsymbol{\theta}_{yi} &= \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}_x \mathbf{X} + \boldsymbol{\Gamma}_{\theta_m} \boldsymbol{\theta}_{mi} + \boldsymbol{\zeta}_{\theta_y i} \end{aligned} \tag{2.8}$$

where $\mathbf{m}_i$ and $\mathbf{y}_i$ are T-dimensional vectors including the repeated measures of the mediator M and the outcome Y for subject i, $\boldsymbol{\theta}_{mi}$ and $\boldsymbol{\theta}_{yi}$ are vectors of p- and q-dimensional latent factors characterising the mediator and the outcome trajectories, respectively, and $\boldsymbol{\Lambda}_m$ and $\boldsymbol{\Lambda}_y$ are the matrices of coefficients, made up of a column of 1's and columns containing the time components. Moreover, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are vectors of intercepts for the mediator and the outcome latent factors, respectively; $\mathbf{B}_x$, $\boldsymbol{\Gamma}_x$ and $\boldsymbol{\Gamma}_{\theta_m}$ are coefficient matrices expressing the relationships of the latent factors with themselves and with other explanatory variables $\mathbf{X}$.

An example of mediation analysis in an LGM framework is the study by Cheong et al. (2003), who carried out a mediation analysis in the context of longitudinal randomised prevention trials. Since they assume that time enters only linearly in

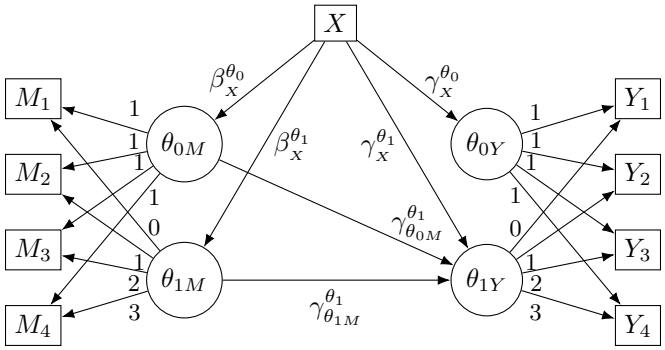

Figure 2.3: Latent growth mediation model for four waves.

the growth model, $\boldsymbol{\theta}_{mi}$ and $\boldsymbol{\theta}_{yi}$ are bidimensional and $\boldsymbol{\Lambda}_m = \boldsymbol{\Lambda}_y = \boldsymbol{\Lambda}$. The mediator process, in the extended form, can be written as

$$\begin{pmatrix} M_{i1} \\ M_{i2} \\ \vdots \\ M_{iT} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & T-1 \end{pmatrix} \begin{pmatrix} \theta_{0Mi} \\ \theta_{1Mi} \end{pmatrix} + \begin{pmatrix} \varepsilon_{Mi1} \\ \varepsilon_{Mi2} \\ \vdots \\ \varepsilon_{MiT} \end{pmatrix};$$

the outcome process can be written analogously. Since in the model of Cheong et al. (2003) there is only one explanatory variable X, the equations for $\boldsymbol{\theta}_{mi}$ and $\boldsymbol{\theta}_{yi}$ in scalar form are

$$\theta_{0Mi} = \beta_0^{\theta_0} + \beta_X^{\theta_0} X_i + \zeta_{\theta_{0M}i} \tag{2.9}$$

$$\theta_{1Mi} = \beta_0^{\theta_1} + \beta_X^{\theta_1} X_i + \zeta_{\theta_{1M}i} \tag{2.10}$$

$$\theta_{0Yi} = \gamma_0^{\theta_0} + \gamma_X^{\theta_0} X_i + \zeta_{\theta_{0Y}i} \tag{2.11}$$

$$\theta_{1Yi} = \gamma_0^{\theta_1} + \gamma_X^{\theta_1} X_i + \gamma_{\theta_{0M}}^{\theta_1} \theta_{0Mi} + \gamma_{\theta_{1M}}^{\theta_1} \theta_{1Mi} + \zeta_{\theta_{1Y}i}. \tag{2.12}$$

The final model entails thus four equations, one for each of the latent factors, for both the mediator and the outcome. Basically, the explanatory variable affects the growth trajectory of the outcome directly and also the trajectory of the mediator, which in turn affects the trajectory of the outcome, as shown in Figure 2.3.

von Soest and Hagtvet (2011) analyse three different model specifications, allowing the explanatory variable X to affect the outcome either through the mediator intercept or through the mediator slope, or through both, as in Cheong et al. (2003). The model including both the mediator latent intercept and slope shows significant differences in estimates with respect to models including only either, and the authors discuss the reasons for preferring one model to the other. O'Laughlin et al. (2018)

further extend the aforementioned works by allowing the exposure to vary over time, then modelling its process and use its latent factors as predictors in the mediator and outcome processes.

Selig and Preacher (2009) remark that modelling mediation through LGM leads to the specification of different indirect effects: those involving only intercepts, those involving only slopes, and those that involve intercepts and slopes. For example, the specification in Equations (2.9)-(2.12) allows us to define the mediated effect on the slope as a product of two coefficients, $\beta_X^{\theta_1}\gamma_{\theta_{1M}}^{\theta_1}$ as usual. The direct effect can be defined as that of the exposure on the outcome intercept, $\gamma_X^{\theta_0}$, or that on the outcome slope, $\gamma_X^{\theta_1}$. O'Laughlin et al. (2018) and von Soest and Hagtvet (2011) adopt a similar strategy for defining mediational effects, which is consistent with path analysis, since it entails the definition of pathways of interest and the computation of the relative effect as the product of coefficients corresponding to arrows in the path.

The interpretation of mediational effects is strongly related to the way time is coded. A common choice is to set $\lambda_t = t - 1$, $t = 1, \ldots, T$. This implies that the random intercept of each variable represents its expectation at the first measurement occasion, and the random slope represents the average change over time. For example, in the model depicted in Figure 2.3, both the initial level of mediator and its average growth mediate the relationship between the exposure and the outcome average growth rate. The conclusion to be drawn is that there is not a unique interpretation, but it depends on the research question and the time coding. If a researcher is not sure about the most appropriate time coding, some of the λ's can be freely estimated, as mentioned previously.

The main advantage of LGM is its noticeable flexibility since it allows us to model individual change trajectories, and the common shape chosen to model them can be of different kinds, not necessarily polynomial, such as exponential or periodic, using trigonometric functions. This model can accommodate both fixed and time-varying covariates. However, some drawbacks have to be taken into account: this model is not adequate if change is not systematically related to the passage of time. Moreover, the way time is codified may be an issue, especially if measurement occasions differ for mediator and outcome.

LGMs can be implemented in STATA (sem), R (`lavaan`, `OpenMx`), SAS (CALIS procedure) and Mplus. Cheong (2011) conducts a simulation study for assessing the accuracy of estimates and the statistical power of mediation in an LGM setting. Her findings show that, generally speaking, a large sample (at least 1,000 observations) is required to produce unbiased estimates and even a statistical power of 0.8 could be demanding in terms of sample size. LGM is widely used to analyse longitudinal data

and applications in mediational settings can be found in many fields: these models have been applied to evaluate the effectiveness of a prevention program (Cheong et al. 2003, Liu et al. 2009, Roesch et al. 2009), in behavioural psychology (von Soest and Hagtvet 2011) and social sciences (Ellwardt et al. 2013).

❖ **Example** : Roesch et al. (2009) analysed data from a randomised clinical trial for encouraging adolescents to do physical activity. Data were collected on a sample of 878 adolescents aged 11-15 years at three time occasions: baseline, after six and twelve months. Subjects were randomly assigned to the treatment or control group: the treatment group received a Teen guide and telephonic consulting to develop healthy behaviours related to diet and physical activity through cognitive and behavioural strategies. The control group received consulting about sun protection and the adoption of correct behaviours to protect skin. The outcome variable, physical activity, was measured as the sum per day of minutes spent in moderate or hard physical activities. Researchers were interested in testing the hypothesis that the treatment effect on physical activity is mediated by psychosocial measures, a composite indicator constructed combining change strategies, decisional balance, self efficacy, family/peer support. The authors fitted a model like that in Figure 2.3, and found a significant direct effect of treatment on physical activity, but no mediation was detected.

Latent difference score models

Another class of models employed for analysing longitudinal data is that of *latent difference score* (LDS) or latent change score models (McArdle 2001, McArdle and Hamagami 2001, Ferrer and McArdle 2003). These models are less popular than CLPM or LGM as regards mediation analysis, but recently they have gained some popularity as well.

In latent difference score models, each observed score is conceived as the sum of the true latent score and an independent error term as in factor CLPM. By defining the first difference in latent scores as $\Delta\tilde{Y}_{it} = \tilde{Y}_{it} - \tilde{Y}_{it-1}$, it follows that $\tilde{Y}_{it} = \tilde{Y}_{it-1} + \Delta\tilde{Y}_{it}$, so that trajectory equations rely on accumulation of latent changes $\tilde{Y}_{it} = \tilde{Y}_{i1} + \sum_{k=2}^{t} \Delta\tilde{Y}_{ik} + \delta_{Y_{it}}$. The latent differences can be modeled as well, for instance as a linear function of the outcome slope and its previous measurements (Ferrer and McArdle 2003). MacKinnon (2008) considers a setting where, in addition to the mediator and the outcome, also the exposure is time-varying. He suggests two ways to model mediation in this framework: modelling difference scores as dependent on variables at previous times or on other difference scores. In the first case, assuming, for example, three waves of data, each difference score depends on the corresponding variable at

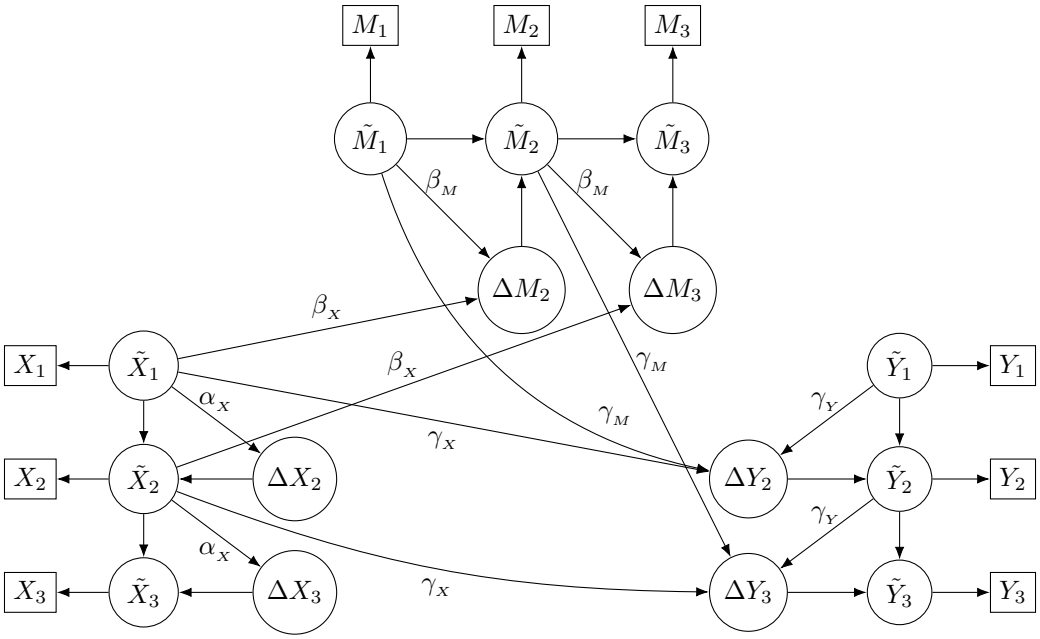

Figure 2.4: Latent difference score model for three measurement occasions.

the previous time occasion and on additional predictors, always measured at previous time, according to the mediation structure, for example

$$\Delta\tilde{X}_{it} = \tilde{X}_{it} - \tilde{X}_{it-1} = \alpha_{X}\tilde{X}_{it-1} + \delta_{\Delta\tilde{X}_{it}} \tag{2.13}$$

$$\Delta\tilde{M}_{it} = \tilde{M}_{it} - \tilde{M}_{it-1} = \beta_{X}\tilde{X}_{it-1} + \beta_{M}\tilde{M}_{it-1} + \delta_{\Delta\tilde{M}_{it}} \tag{2.14}$$

$$\Delta\tilde{Y}_{it} = \tilde{Y}_{it} - \tilde{Y}_{it-1} = \gamma_{X}\tilde{X}_{it-1} + \gamma_{M}\tilde{M}_{it-1} + \gamma_{Y}\tilde{Y}_{it-1} + \delta_{\Delta\tilde{Y}_{it}}. \tag{2.15}$$

A graphical representation for an LDS with three waves of data is given in Figure 2.4.

In the second case, each difference score will depend on difference scores relative to previous waves, for example, $\Delta\tilde{Y}_t$ may depend on $\Delta\tilde{M}_{t-1}$. Other kinds of models are possible, as shown in Selig and Preacher (2009). O'Laughlin et al. (2018) consider a model similar to the original formulation (Ferrer and McArdle 2003, McArdle and Hamagami 2001) since the model also includes a random slope for each variable.

Analogously to CLPM and LGM, mediated effects in an LDS can be obtained as products of coefficients along the paths of interest. Once again, the interpretation depends on which structure is hypothesised: for instance, in Figure 2.4, the exposure latent score at time t affects the latent difference between the outcome score at time $t+1$ and $t+2$ indirectly, through the latent difference of the mediator. This means that the latent construct of the exposure affects a successive change in the outcome through the mediator construct. In a model involving just latent differences, like the one discussed by O'Laughlin et al. (2018), the effects of changes in the exposure are

mediated by changes in the mediator, which in turn affect the change in the outcome.

LDS model is less used than LGM to carry out mediation analysis; however, if the trajectory of change is expected to modify from a time interval to the next, LDS is an alternative to nonlinear LGM (Selig and Preacher 2009).

The LDS has been implemented in STATA (sem), R (`lavaan`, `OpenMx`), SAS (CALIS) and Mplus. Simone and Lockhart (2019) conduct a Monte Carlo simulation study in Mplus to assess the sample size required to detect mediational effects. They simulate data considering ten LDS models and three effect sizes for the paths connecting the exposure to the mediator and the mediator to the outcome, with a total of 90 scenarios. They show that the complexity of the models highly affects the required sample sizes, and large effect size pairings require smaller samples. To the best of our knowledge, the only applications to longitudinal mediation settings can be found in social sciences (Selig and Preacher 2009, O'Laughlin et al. 2018).

❖ **Example** : Selig and Preacher (2009) analysed data from the National Institute for Child Health and Development's Study of Early Childcare and studied how maternal depression affects child's problem behaviour through maternal sensitivity. All variables are time-varying and the authors showed that mothers with higher depression levels are less sensitive and this negatively affects child's behaviour.

2.2.3 Mixed effects models

Mixed effects, multilevel or hierarchical models are generally employed for data obtained from nested designs, for example a sample of children nested in a sample of schools. In longitudinal settings, repeated measurements are nested within individuals. The lower level units (children or time occasions) are called level-1 units, while upper level units (schools or individuals in longitudinal settings) are called level-2 units.

The first articles about multilevel mediation analysis, not specifically focused on longitudinal contexts, appeared in the early 2000s. Krull and MacKinnon (1999, 2001) introduced the notation that is now well established in the literature for denoting multilevel mediation designs. A design can be identified as $l_X \rightarrow l_M \rightarrow l_Y$, where l_W, $W \in \{X, M, Y\}$, denotes the level at which variable W is measured: for example, if the exposure, the mediator and the outcome are all measured at a subject level, the design will be denoted as $1 \rightarrow 1 \rightarrow 1$, while if only the outcome is measured at the cluster level, the design will be denoted as $1 \rightarrow 1 \rightarrow 2$. In a longitudinal context, these designs correspond to the case where all variables are time-varying, and the

case of a response variable which does not change over time, respectively.

Krull and MacKinnon (1999, 2001) consider linear mixed models (LMMs) including only a random intercept, and Krull and MacKinnon (2001) claim that estimation is possible only if the mediator and the outcome are measured at the same or at a lower level than their predictors, so for example, there cannot be an outcome measured at the individual level (level 2) and a time-varying mediator (level 1).

Kenny et al. (2003) focus on more complex linear models including also random slopes for a $1 \rightarrow 1 \rightarrow 1$ design

$$\begin{aligned} M_{it} &= \pi_{0Mi} + \pi_{MXi} X_{it} + \varepsilon_{Mit} \\ \pi_{0Mi} &= \beta_0 + b_{0i} \\ \pi_{MXi} &= \beta_X + b_{Xi} \end{aligned} \tag{2.16}$$

$$\begin{aligned} Y_{it} &= \pi_{0Yi} + \pi_{YXi} X_{it} + \pi_{YMi} M_{it} + \varepsilon_{Yit} \\ \pi_{0Yi} &= \gamma_0 + g_{0i} \\ \pi_{YXi} &= \gamma_X + g_{Xi} \\ \pi_{YMi} &= \gamma_M + g_{Mi} \end{aligned} \tag{2.17}$$

where the π terms are subject-specific random coefficients, which can be written as the sum of a common mean and subject-specific deviations, and the ε terms are time-specific deviations. π are assumed to be from a multivariate Normal distribution, with possibly non-diagonal covariance matrix

$$\begin{pmatrix} \pi_{0Mi} \\ \pi_{MXi} \\ \pi_{0Yi} \\ \pi_{YXi} \\ \pi_{YMi} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \beta_0 \\ \beta_X \\ \gamma_0 \\ \gamma_X \\ \gamma_M \end{pmatrix}, \begin{pmatrix} \sigma^2_{b_0} & & & & \\ \sigma_{b_0 b_X} & \sigma^2_{b_X} & & & \\ \sigma_{b_0 g_0} & \sigma_{b_X g_0} & \sigma^2_{g_0} & & \\ \sigma_{b_0 g_X} & \sigma_{b_X g_X} & \sigma_{g_0 g_X} & \sigma^2_{g_X} & \\ \sigma_{b_0 g_M} & \sigma_{b_X g_M} & \sigma_{g_0 g_M} & \sigma_{g_X g_M} & \sigma^2_{g_M} \end{pmatrix} \right)$$

and ε terms are Normal with mean zero and are assumed to be uncorrelated.

Using the results proved by Goodman (1960), Kenny et al. (2003) show that the indirect effect can be obtained as

$$\mathbb{E}[\pi_{MXi}\, \pi_{YMi}] = \beta_X \gamma_M + \sigma_{b_X g_M}, \tag{2.18}$$

where $\sigma_{b_X g_M}$ is the covariance between b_X and g_M. The role of this term is important, since it can make an indirect effect non-null even if one of the two means β_X or γ_M is null. In addition, according to the sign of such covariance, it can attenuate or amplify the magnitude of the indirect effect.

Unfortunately, the estimation of $\sigma_{b_X g_M}$ is not straightforward in the traditional multilevel setting and these difficulties are reflected also on software implementation. Kenny et al. (2003) propose a two-step procedure for estimating this covariance: first they estimate the random coefficients of the mediator and outcome models for each level-2 unit, and then they use these estimates to compute the covariance. This strategy is definitely not optimal, see Kenny et al. (2003) for a discussion.

To overcome the problem, Bauer et al. (2006) propose the specification of a single model

$$Z_{it} = S_{M_{it}}(\pi_{0Mi} + \pi_{MXi}X_{it}) + S_{Y_{it}}(\pi_{0Yi} + \pi_{YXi}X_{it} + \pi_{YMi}M_{it}) + \varepsilon_{Zit},$$

where Z is a variable obtained by stacking M and Y, S_M and S_Y are indicator variables equal to 1 when Z refers to the mediator or to the outcome, respectively. In other words, this specification allows one to estimate the mediator and the outcome model simultaneously via a single model, making it possible to estimate the covariance between b_X and g_M. The authors provide formulas to quantify the precision of the estimates and conduct a simulation study to evaluate the performance of the proposed method.

A different approach was used by Preacher et al. (2010, 2011) within the multilevel structural equation framework, which allows researchers to deal with any kind of mediation design, even those including an upper-level response and a lower-level predictor. This framework will be discussed more deeply in the next chapter.

One of the characteristics of the multilevel SEM approach is the decomposition of effects into a within and a between component. This is related with the problem of centering, a very serious issue in the multilevel literature in general, not only in a mediation setting. Indeed, in multilevel models, choosing the mean with respect to which performing centering is not so straightforward, since there are at least three possibilities: no centering, centering variables with respect to their grand-mean (CGM) or centering each variable with respect to its cluster mean (CWC). The choice can have a severe impact on the estimates and their meaning. Reviews and comparisons of different centering strategies in multilevel mediation settings can be found in Asparouhov and Muthén (2019), Tofighi (2010) and Zigler and Ye (2019). We will not focus on centering since the topic is very broad and an extensive treatment would require an entire chapter. Thorough and enlightening articles about centering in multilevel models (not necessarily encompassing mediation) are Kreft et al. (1995), Paccagnella (2006) and Enders and Tofighi (2007). The implications of centering in multilevel longitudinal models are discussed in Curran et al. (2012), Curran and Bauer (2011), Wang and Maxwell (2015) and Hoffman and Stawski (2009).

Mixed effect models are quite common in longitudinal studies, and they have been implemented in different software: STATA (me), R (`lme4`, `nmle`, `glmmTMB`), SAS (GLIMMIX) and Mplus. Bauer et al. (2006) provide SAS codes to implement the methods proposed in their work.

The first simulation studies concerning multilevel mediation were carried out by Krull and MacKinnon (1999) and Krull and MacKinnon (2001). In the former, the authors aim to understand the extent to which the estimates of the indirect effect obtained through the product or the difference method differ, which estimation method should be preferred and how standard errors approximations based on large sample assumptions perform in small samples. Their results show that the average discrepancy between the two estimates is close to zero and that the two estimation methods are approximately equivalent in large sample sizes. The authors suggest, however, to prefer the product method since it is more informative than the difference method, especially in multiple mediator settings.

In the second study (Krull and MacKinnon 2001), the authors compared the performance of single-level and multilevel mediation models in estimating and testing mediated effects in clustered data, considering all admissible designs ($1 \to 1 \to 1$, $2 \to 1 \to 1$, $2 \to 2 \to 1$) and a wide variety of scenarios. They showed that single-level models underestimate the standard error of the indirect effect, as expected. This underestimation is often imputable to the underestimation of the standard error of a coefficient along a $2 \to 1$ path (β_X in $2 \to 1 \to 1$, γ_M in $2 \to 2 \to 1$).

Other simulation studies were conducted by Blood et al. (2010) and Blood and Cheng (2011), who compared the performance of LMMs with that of SEMs in different longitudinal mediation settings. They showed that in many cases the two models have similar performances. Applications of mixed models in longitudinal mediation settings have been provided in psychology (Bauer et al. 2006) and epidemiology (Blood and Cheng 2011).

❖ **Example** : Blood and Cheng (2011) analysed data from a cohort of HIV patients followed over four years. They were interested in the relationship between heavy alcoholism and CD4 cell count, mediated by adherence to the anti-retroviral therapy (ART), measured as the percentage of prescribed pills in the last three days. Each variable was assessed every six months. The authors fitted different models, including also a random slope for time, and compared the results, which did not show any significant relationship among variables.

2.2.4 Other models

The models discussed up to this point are the most common for studying mediation in the associational setting. Nonetheless, other longitudinal mediation models have been proposed in the literature, but they are difficult to classify into one of the categories previously proposed. In this section, we provide a brief overview of these other models.

Gunzler et al. (2014) propose a longitudinal SEM for mediation analysis by using *functional response models* (FRMs) to make inference. FRM is a class of distribution-free models, so they do not require any parametric assumptions on the data distribution, and they can easily handle the double role played by the mediator in the simultaneous equations. The authors extensively discuss inferential procedures, holding in both cases of complete and missing data (missing completely at random, MCAR, and missing at random, MAR). They carry out a simulation study to evaluate the performance of their approach for different sample sizes and considering scenarios with completed and missing data, in comparison with the traditional Maximum Likelihood (ML) approach. They show that efficiency of FRM approach is quite similar to that of ML under the complete data setting, but that it outperforms ML in the MAR case. An example from psychology is used to show a real-world application.

Starting from the concept of *ergodicity* in psychology, that is, the situation where the average of a single trajectory over time equals the average of all sample trajectories at a single time occasion, Gu et al. (2014) point out that researchers should rethink the way they conceive mediation. Indeed, ergodicity relies on both concepts of stationarity, which refers to a stochastic process having a time-invariant joint probability distribution, and homogeneity, meaning that all members in a given population are interchangeable and follow the same statistical model. Clearly, it is very unlikely that these assumptions hold in real-world psychological processes. Therefore, the authors suggest to focus on intra-individual variations, rather than on inter-individual ones, and to investigate, then, single-subject time series data. This method relies on the state-space model, i.e. an SEM where latent variables at time t depend on themselves at previous time $t-1$ via a transition matrix. Parameters are estimated through a recursive algorithm, the Kalman filter, and confidence intervals for the mediated effects are obtained with nonparametric or residual-based bootstrap. A simulation study and an application from behavioural psychology are provided.

Although not specifically targeted for longitudinal mediation, it is worth mentioning the paper by Zhang et al. (2009), where the authors consider three different multilevel mediation models and evaluate the bias produced by unobserved confounders, proposing some possible solutions. This model can find application in longitudinal

mediation settings.

2.2.5 Discussion

SEMs are very common and widely used among researchers, also due to the high number of software where they are implemented. They are a powerful tool to carry out longitudinal mediation analysis. However, they also have some limitations, the main being the assumption of Normality of the variables and the linearity of models. These two characteristics restrict the range of models which can be fitted and the outcome variables which can be analysed.

Although in the last years several scholars have started extending SEM estimation methods to accommodate non-Normal data, interactions among variables and other forms of non-linearity, see for example Wall (2009), Tsai et al. (2006), Lai (2018), Lee and Zhu (2002) and Mayer et al. (2017), these extensions have not been applied to mediation analysis yet.

Mixed-effect models are also well known, but they allow for more flexibility in the variables' distribution and regression models with link functions different from identity can be estimated. However, all the approaches discussed relies on linearity as well. One of the limitations of mixed-effect models is the impossibility of including latent variables. This prevents their application in some fields, such as psychology or social sciences, were measured variables are often fallible indicators of latent constructs. In addition, as discussed in Section 2.2.3, traditional multilevel settings cannot accommodate all types of multilevel mediation designs and the covariance among the random slope of the exposure in the mediator model and that of the mediator in the outcome model has to be estimated through *ad hoc* methods. We will propose a possible solution overcoming these issues in Chapter 3.

Finally, it is worth mentioning the issue of identification traditionally arising for SEM. It should not be confused with the identifiability issue typical of causal models, since that of SEMs concerns the difference between the number of known and unknown parameters. If this difference is negative, i.e if the number of unknown parameters exceeds the number of known ones, the model is not identified. We do not address this issue and refer the interested reader to Bollen and Curran (2006), Usami et al. (2019), McArdle and Nesselroade (2013), who provide conditions for this kind of identification.

2.3 Causal framework

In this second part of the chapter we shall focus on causal approaches to longitudinal mediation analysis. While in the associational framework we talked about models, here we shall refer to approaches, which provide different definitions of mediational effects, hence different estimands, and different assumptions for identifying them. This section addresses the different types of estimands discussed in Section 1.2.5 and extends them to a longitudinal setting, showing the advantages and drawbacks of each and providing some examples of application. First, we introduce some key concepts of the longitudinal causal literature.

2.3.1 Basic concepts and notation

Time-varying treatments and regimes

The definition of counterfactuals given in Section 1.2.3 needs to be extended to accommodate the temporal aspect. In a longitudinal setting, if the treatment is time-varying, the potential outcomes do not involve a comparison of just two different values of the treatment, but a comparison of two treatment strategies $\bar{x} = (x_1, x_2, \ldots, x_T)$ and $\bar{x}^* = (x_1^*, x_2^*, \ldots, x_T^*)$, where x_t and x_t^*, for $t = 1, \ldots, T$, are values assigned to treatment at time t. A treatment strategy, or a emphregime, is a rule to assign treatment at each time t. For example, in a RCT aimed to study the effectiveness of different strategies to cope with insomnia, a group of subjects can be randomly assigned to the treatment condition, i.e. taking sleeping tablets every night, or the control condition, i.e. doing relaxation exercises before sleeping every night. A sleep monitor can record the number of minutes spent before to fall asleep and how many hours a subject slept every night for two months, say.

A regime is called *static* when, at each time, it depends only on previous treatment values, *dynamic*, if, in addition, it depends also on previous values of the covariates $\overline{L}_t$. In the previous example, the two regimes compared would be static if all patients in the treatment group were told to take the same number of pills every day, and subjects in the control group were asked to do their relaxation exercise for thirty minutes every day before sleeping. In contrast, they would be dynamic if the number of pills to take or for how many minutes a subject should exercise depended on the type of physical activity done during the day. Indeed, physical activity is known to act as a stimulant and can make falling asleep more difficult. Individuals who do very intense physical activity could be asked to take more sleeping pills or to do relaxation exercises for longer time than that required for subjects practising very mild physical activity.

Regimes can also be distinguished into *deterministic*, when each subject receives a specific value at each time, and *random*, when each subject is assigned a probability of receiving a certain value of the treatment. See Young et al. (2014) and Hernán and Robins (2020) for more details on regimes.

The average total effect of a time-varying treatment X on an outcome Y measured at the end of the follow-up is then defined comparing two regimes $\bar{x} \neq \bar{x}^*$ as

$$\mathbb{E}[Y(\bar{x})] - \mathbb{E}[Y(\bar{x}^*)]. \tag{2.19}$$

If the treatment is binary, a comparison which is often of interest is that between individuals always treated, that is, individuals with $\bar{x} = \bar{1} = (1, \ldots, 1)$ and individuals never treated, that is, with $\bar{x}^* = \bar{0} = (0, \ldots, 0)$, but many other strategies are actually possible.

Analogously to cross-sectional contexts discussed in 1.2.3, the total causal effect in (2.19) needs an exchangeability assumption for being identified. In randomised experiments, the total effect is identified since one of the following conditions, which are called *sequential unconditional* and *sequential conditional exchangeability*, holds, according to whether there are time-varying confounders or not

$$Y(\bar{x}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1} = \bar{x}_{t-1} \tag{2.20}$$

$$Y(\bar{x}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1} = \bar{x}_{t-1}, \overline{L}_t. \tag{2.21}$$

It is worth remarking that these assumptions refer to *static sequential ignorability*, since they hold for static strategies. As discussed in Hernán and Robins (2020), there exist different forms of sequential exhangeability, and they allow one to identify the effects of some treatment strategies, while others remain unidentified. See Hernán and Robins (2020), Richardson and Robins (2013) for a discussion. In the following we will always refer to static strategies.

We will denote by $M(\bar{x})$ and $Y(\bar{x})$ the counterfactual values of the mediator and the outcome, respectively, under regime $\bar{x}$. Similar concepts apply to $\overline{m}$ and $Y(\bar{x}, \overline{m})$.

Time-varying confounders

Time-varying confounding is an issue which may clearly arise in both the associational and the causal frameworks, but in the causal one is particularly troubling.

Figure 2.5 shows a DAG including time-varying exposures X_t, mediators M_t and confounders L_t, $t = 1, \ldots, 3$, an outcome measured at the end of the follow-up period, Y, and an unobserved variable U. This DAG may represent a randomised sequential trial (notice the absence of arrows from U to X) where the treatment at time t is

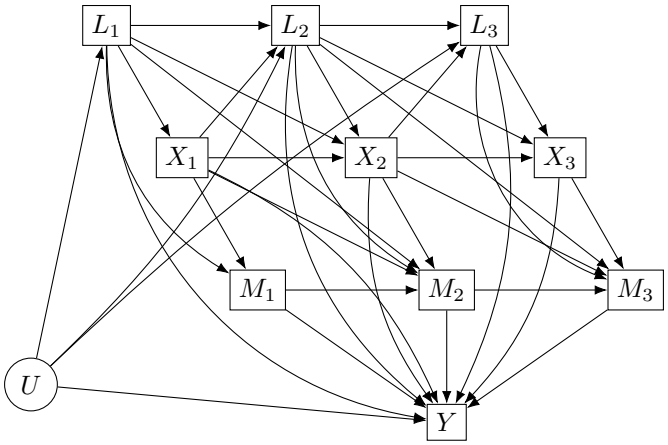

Figure 2.5: DAG including time-varying confounders.

assigned only on the basis of previous treatment and covariate history, $\overline{X}_{t-1}$ and $\overline{L}_t$. Notice that L_t confounds the relationship between M_t and Y and is affected by the exposure at prior time, X_{t-1} for $t > 1$. This violates assumption (d). Moreover, L_t is a confounder of the relationship between X_t and Y, for each t. It is defined a *time-varying confounder*, since it is not sufficient to condition on L at baseline ($t = 1$) to block the back-door paths between the exposure and the outcome, but at each time t the entire covariate history $\overline{L}_t$ is necessary (Hernán and Robins 2020, Chapter 19).

It is worth remarking that, even in the absence of L in Figure 2.5, the mediator can become itself a time-varying confounder, since, for example M_1 confounds the $M_2 - Y$ relationship and is affected by X_1.

In the presence of time-varying confounders, natural effects cannot be identified, and this calls for other types of mediational effects, relying on weaker assumptions. We are going to discuss these alternatives in the next sections.

Finally, although they do not regard the longitudinal setting specifically, we want to point out other problems which may affect any of the methods we are going to describe: measurement error of variables and model misspecification. Their discussion is beyond the scope of this chapter; for some useful references see VanderWeele et al. (2012b), le Cessie et al. (2012), Lutz et al. (2020).

2.3.2 Natural effects

Given the problems connected to the cross-world independence assumption, natural effects are not the most used in longitudinal settings.

Bind et al. (2016) approach makes use of mixed-effect models. They consider a setting where the exposure, the mediator and the outcome are time-varying. The exposure has a cross-sectional effect on the mediator and the outcome and an autoregres-

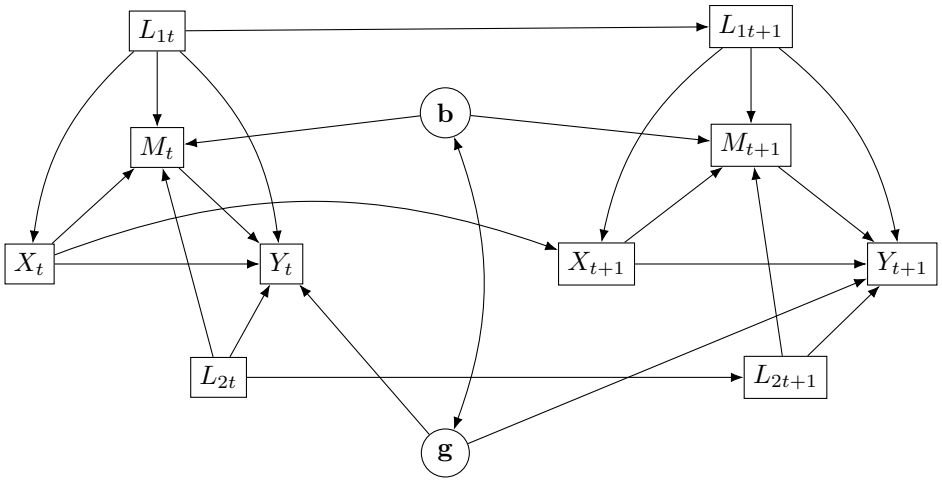

Figure 2.6: Model proposed by Bind et al. (2016).

sive effect on itself, while subsequent instances of the mediator and the outcome are not linked directly, but only via the random effects $\mathbf{u}_i = (\mathbf{b}_i, \mathbf{g}_i)'$, $i = 1, \ldots, n$. Additional observed covariates $\boldsymbol{L}$, possibly confounding the exposure-mediator, exposure-outcome and mediator-outcome relationships are allowed, see Figure 2.6.

The natural direct and indirect effects, conditional on covariates and random effects, are given by $\mathbb{E}[Y_{it}(x, M_{it}(x^*)) - Y_{it}(x^*, M_{it}(x^*)) \mid \boldsymbol{L}, \mathbf{u}_i]$ and $\mathbb{E}[Y_{it}(x, M_{it}(x)) - Y_{it}(x, M_{it}(x^*)) \mid \boldsymbol{L}, \mathbf{u}_i]$, respectively. They are identified under a modified version of (a)-(d) in Section 1.2.5, to accommodate the longitudinal setting and where the conditioning is also on random effects.

To estimate natural mediational effects, the authors model the mediator and the outcome through generalized linear mixed models, as shown below

$$h_1(\mathbb{E}[M_{it} \mid \boldsymbol{L}_{it} = \boldsymbol{l}, X_{it} = x, \mathbf{b}_i]) = (\beta_0 + b_{0i}) + (\beta_X + b_{Xi})x + \boldsymbol{\beta}_L' \boldsymbol{l}$$

$$\begin{aligned} &h_2(\mathbb{E}[Y_{it} \mid \boldsymbol{L}_{it} = \boldsymbol{l}, X_{it} = x, M_{it} = m, \mathbf{g}_i]) \\ &\quad = (\gamma_0 + g_{0i}) + (\gamma_X + g_{Xi})x + (\gamma_M + g_{Mi})m + (\gamma_{XM} + g_{XMi})xm + \boldsymbol{\gamma}_L' \boldsymbol{l}, \end{aligned}$$

where h_1 and h_2 are known link functions, and $\mathbf{b}_i$ and $\mathbf{g}_i$ are subject-specific random effects.

This approach is quite flexible since it allows for exposure-mediator interactions and multiple mediators, and can accommodate different types of mediators and outcomes. In the setting including a single time-varying mediator, if h_1 and h_2 are the identity link function, the natural mediational effects conditional on covariates and random effects are

$$NDE = (\gamma_X + g_{Xi})(x - x^*) + (\gamma_{XM} + g_{XMi})(x - x^*) \left[(\beta_0 + b_{0i}) + (\beta_X + b_{Xi})x^* + \boldsymbol{\beta}_L' \boldsymbol{l}\right]$$

$$NIE = (\gamma_M + g_{Mi})(\beta_X + b_{Xi})(x - x^*) + (\gamma_{XM} + g_{XMi})(\beta_X + b_{Xi})(x - x^*).$$

Effects conditional only on covariates can be obtained by integrating out the random effects.

Although quite general, a limitation of this approach is the cross-world independence assumption $Y_{it}(x, m) \perp\!\!\!\perp M_{it}(x^*) \,|\, \boldsymbol{L}_{it}, \mathbf{b}_i, \mathbf{g}_i$, for each subject i and time t. As discussed previously, this assumption may not be plausible, especially if the time lag between exposure and outcome measurements is wide. In this case, it holds because the authors considered only covariates not affected by the exposure, but this sounds unnatural or at least very unlikely in real contexts. Moreover, the exposure is assumed to be randomized, which is plausible when it is an environmental variable, as in the application provided in the paper, but may generally not be satisfied.

2.3.3 Path-specific effects

Path-specific effects reveal their usefulness in complex causal graphs, like those for multiple mediator models or longitudinal mediation models, where it is more likely that assumption (d) is not satisfied. In the former case, researchers can be interested in the effect conveyed by each mediator (Lin and VanderWeele 2017, Vansteelandt and Daniel 2017), while in the latter the focus is on the propagation of effects over time. Shpitser (2013) extended the results of Avin et al. (2005) to longitudinal mediation models characterised by the presence of unobserved variables, proposing a generalisation of the recanting witness criterion, as discussed in Chapter 1.

Consider the graph in Figure 2.7(a), where there are a time-varying treatment X, mediator M and observed confounder L and an outcome Y measured at the end of follow-up. The relationship between L and Y is confounded by the presence of an unobserved variable U. Panel (b) shows the corresponding latent projection. The set of paths in blue, $\pi = \{X_1 \to M_1 \to Y, X_2 \to M_2 \to Y, X_1 \to M_1 \to M_2 \to Y\}$, represents the π-specific effect we want to estimate. Consider again two levels of X, denoting treatment by x and no treatment (or treatment at a baseline level) by x^*. The potential outcome corresponding to π is

$$Y(\pi, x, x^*) = Y(x_1^*, x_2^*, L_1(x_1^*), L_2(x_1^*, x_2^*), M_1(x_1, L(x_1^*)), M_2(x_2, L_2(x_1^*, x_2^*)))$$

and, following Theorem 1.2.9, it can be expressed as a function of interventional densities, since there are no recanting districts for π. Indeed, $\boldsymbol{Y}^* = \{L_1, L_2, M_1, M_2, Y\}$ and $\mathcal{D}^* = \{\{M_1\}, \{M_2\}, \{L_1, L_2, Y\}\}$ and there are no elements of the same district lying on a blue and on a black path, then, there are no recanting districts. As a

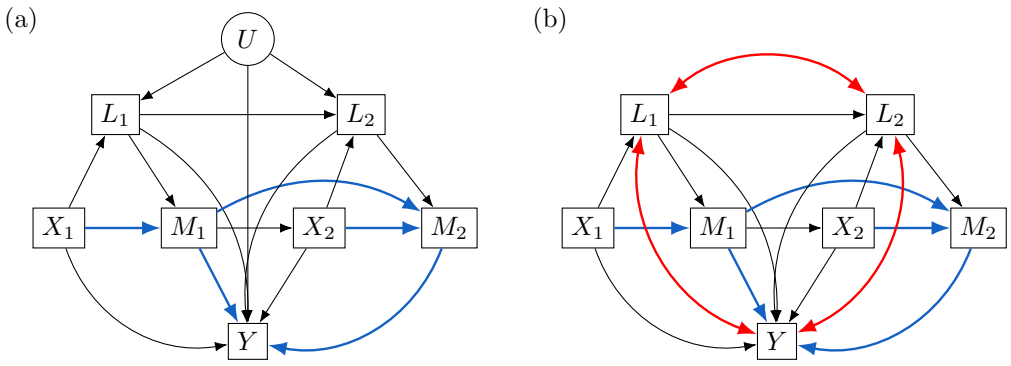

Figure 2.7: A DAG (a) and its corresponding ADMG (b). Blue edges denote paths of interest.

consequence, $Y(\pi, x, x^*)$ can be expressed as

$$\sum_{l_1,l_2,m_1,m_2} P(Y(x_1^*, x_2^*, l_1, l_2, m_1, m_2), L_1(x_1^*) = l_1, L_2(x_1^*, x_2^*) = l_2) \times$$
$$P(M_1(x_1, l_1) = m_1) P(M_2(x_2, l_2) = m_2)$$

consistently with Equation (1.26). Since $P(Y(x_1, x_2))$ is identified as well, the above product is a function of observed variables

$$\sum_{l_1,l_2,m_1,m_2} P(Y \mid x_1^*, x_2^*, l_1, l_2, m_1, m_2) P(m_2 \mid x_1, x_2, m_1, l_2) \times$$
$$P(l_2 \mid x_1^*, x_2^*, l_1) P(m_1 \mid x_1, l_1) P(l_1 \mid x_1^*).$$

The causal effect of X along π can be expressed similarly to natural effects, as the difference

$$\mathbb{E}[Y(\pi, x, x^*)] - \mathbb{E}[Y(x^*)] =$$
$$\sum_{l_1,l_2,m_1,m_2} P(Y \mid x_1^*, x_2^*, l_1, l_2, m_1, m_2) P(m_2 \mid x_1, x_2, m_1, l_2) P(l_2 \mid x_1^*, x_2^*, l_1) \times$$
$$P(m_1 \mid x_1, l_1) P(l_1 \mid x_1^*) - \sum_{m_1,l_1} \mathbb{E}[Y \mid m_1, l_1, x_1^*, x_2^*] P(m_1, l_1 \mid x_1^*).$$

This result is an illustration of the following theorem by Shpitser (2013)

Theorem 2.3.1. *Let $\mathcal{G}(\boldsymbol{V})$ be an ADMG representing a causal diagram with unobserved confounders. Let $\boldsymbol{X}$ be a set of nodes, Y a single node in $\mathcal{G}$, and π a subset of proper causal paths which start with a node in $\boldsymbol{X}$ and end in Y. If there does not exist a recanting district for the π-specific effect of $\boldsymbol{X}$ on Y, the counterfactual distribution $P(Y(x))$ is identified and $P(Y(\pi, x, x^*))$ can be expressed as in 1.26, then the path-specific effect along the set of paths π on the mean difference scale for active*

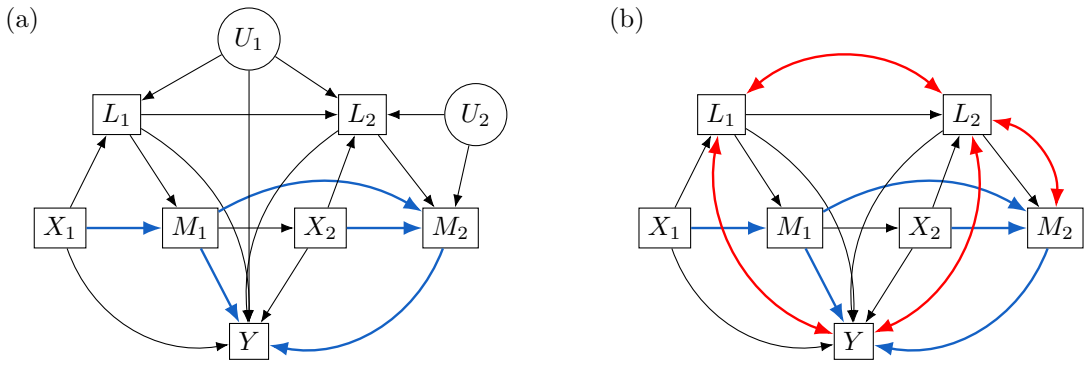

Figure 2.8: Causal DAG where the recanting district criterion is not satisfied.

value x *and baseline value* x^* *is equal to*

$$\mathbb{E}[Y(\pi, x, x^*)] - \mathbb{E}[Y(x^*)] \tag{2.22}$$

and the path specific effect along the set of paths $\bar{\pi}$ *complementary to* π *as*

$$\mathbb{E}[Y(x)] - \mathbb{E}[Y(\pi, x, x^*)]. \tag{2.23}$$

Figure 2.8(a) is an example where identifiability does not hold (Shpitser 2013). Districts in $\mathcal{D}^*$ are $\{M_1\}, \{L_1, L_2, M_2, Y\}$. The latter is a recanting district, since the path $X_2 \to M_2 \to Y$ is in π, $X_2 \to L_2 \to Y$ is in $\bar{\pi}$, and M_2 and L_2 belong to the same district, as can be seen from the latent projection in (b). The path-specific effect cannot be identified.

An application of PSE can be found in Mittinty and Vansteelandt (2020), who extend natural effects models (Lange et al. 2012, Vansteelandt et al. 2012) to longitudinal settings. The authors assume a structure including a baseline exposure X, time-varying mediators M and observed confounders L, and an outcome Y measured at the end of follow-up. They allow for unmeasured confounders between L and Y. They are interested in the direct effect $\mathbb{E}[Y_t(x, \overline{M}_t(x^*)) - Y_t(x^*, \overline{M}_t(x^*))]$, corresponding to PSE along paths linking X to Y directly, or through any of the time-varying covariates, and the indirect effect $\mathbb{E}[Y_t(x, \overline{M}_t(x)) - Y_t(x, \overline{M}_t(x^*))]$, encoding the PSE along paths connecting X to Y only through the mediators. Notice that the latter is only a portion of the indirect effect, since the combinations of paths of the form $X \to L \to M \to Y$ and $X \to M \to Y$ cannot be identified. The authors limit themselves to the estimation of the latter kind of paths, which under (a)-(d) are still identified. Modeling the potential outcome via a natural effect model, such as

$$h(\mathbb{E}[Y_t(x, \overline{M}_t(x^*))]) = \gamma_0 + \gamma_1 x + \gamma_2 x^* + \gamma_3 t + \gamma_4 t x + \gamma_5 t x^*,$$

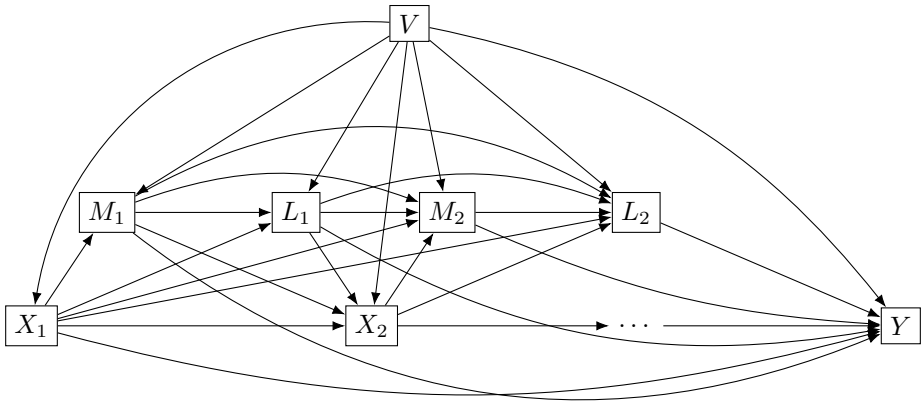

Figure 2.9: Longitudinal mediation model discussed in VanderWeele and Tchetgen Tchetgen (2017).

where h is a known link function, it can be proved that these effects correspond to combinations of model coefficients. Specifically, if h is the identity function, $x = 1, x^* = 0$, the direct effect is given by $\gamma_1 + \gamma_4 t$ and the indirect effect by $\gamma_2 + \gamma_5 t$.

2.3.4 Interventional effects

As we have seen, in longitudinal settings assumption (d) can easily be violated. In this section we show the extension of the interventional effects introduced in Section 1.2.5 to longitudinal settings and how they were employed in the literature.

VanderWeele and Tchetgen Tchetgen (2017) consider a setting which includes time-varying exposures X, and mediators M, a set of baseline and time-varying covariates V and L, respectively, and an outcome Y measured at the end of the follow-up. The L variables confound the relationship between the mediators and the outcome at each time, and they are affected by the exposure. Thus, natural direct and indirect effects cannot be identified, and one of the solutions is to resort to interventional effects.

Consistently with the notation introduced in Section 1.2.5, let $\overline{m}_t^d(\bar{x}|v)$ denote a random draw from the distribution of $\overline{M}_t$ that would have been observed if $\overline{X}$ were set to $\bar{x}$, with baseline covariates $V = v$. Identification of longitudinal interventional effects requires longitudinal analogues of ignorability assumptions (a)-(c), i.e.

(a″) $Y(\bar{x}, \overline{m}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1} = \bar{x}_{t-1}, \overline{M}_{t-1} = \overline{m}_{t-1}, \overline{L}_{t-1} = \dot{l}_{t-1}, V$

(b″) $Y(\bar{x}, \overline{m}) \perp\!\!\!\perp M_t \,|\, \overline{X}_t = \bar{x}_t, \overline{M}_{t-1} = \overline{m}_{t-1}, \overline{L}_{t-1} = \dot{l}_{t-1}, V$

(c″) $M(\bar{x}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1} = \bar{x}_{t-1}, \overline{M}_{t-1} = \overline{m}_{t-1}, \overline{L}_{t-1} = \dot{l}_{t-1}, V$

where variables with zero or negative subscripts are considered to be null. Under

these assumptions, the estimand of interest is identified by

$$Q_1(\bar{x}, \bar{x}^*) = \mathbb{E}[Y(\bar{x},\, \overline{m}^d(\bar{x}^*|v))|\, v] = \sum_{\overline{m}} \sum_{\dot{l}_{T-1}} \mathbb{E}[Y|\bar{x},\, \overline{m}, \dot{l},\, v] \prod_{t=1}^{T-1} P(\dot{l}_t \,|\, \bar{x}_t,\, \overline{m}_t,\, \dot{l}_{t-1}, v)$$
$$\times \sum_{\dot{l}^\dagger_{T-1}} \prod_{t=1}^{T} P(m_t|\bar{x}^*_t, \overline{m}_{t-1}, \dot{l}^\dagger_{t-1}, v)\, P(\dot{l}^\dagger_{t-1}|\bar{x}^*_{t-1}, \overline{m}_{t-1}, \dot{l}^\dagger_{t-2}, v). \tag{2.24}$$

Equation (2.24) is called *mediational g-formula*. The longitudinal interventional direct and indirect effects are then given by

$$IDE = \mathbb{E}[Y(\bar{x},\, \overline{m}^d(\bar{x}^*|v))|v] - \mathbb{E}[Y(\bar{x}^*,\, \overline{m}^d(\bar{x}^*|v))|v] = Q_1(\bar{x}, \bar{x}^*) - Q_1(\bar{x}^*, \bar{x}^*)$$
$$IIE = \mathbb{E}[Y(\bar{x},\, \overline{m}^d(\bar{x}|v))|v] - \mathbb{E}[Y(\bar{x}, \overline{m}^d(\bar{x}^*|v))|v] = Q_1(\bar{x}, \bar{x}) - Q_1(\bar{x}, \bar{x}^*)$$

The authors suggest to use marginal structural models and inverse probability weighting (Robins et al. 2000) for the estimation of each member of (2.24).

Zheng and van der Laan (2017) propose a modification of the approach just described, based on the following observation. VanderWeele and Tchetgen Tchetgen (2017) use marginal mediator distributions, i.e., they condition only on baseline covariates. Zheng and van der Laan point out that, in a survival setting, this method may lead to ill-defined counterfactuals since "a person who is still alive under $X = x$, would be allowed to draw the mediator value of someone under $X = x^*$ who has died". Instead, their proposal is based on conditional interventional counterfactuals, where the conditioning is on each subject's time-varying history. Consider the data structure

$$O = (V_0, X_1, L_1, M_1, V_1, \ldots, X_t, L_t, M_t, V_t, \ldots, X_T, L_T, M_T, V_T \supset Y_T) \sim D_0,$$

where V_0 are baseline covariates, L_t are covariates affected by the exposure X_t and which may affect the mediator M_t and V_t. Here, V_t are covariates affected by X_t, L_t, M_t, and, in particular, V_T, includes the final outcome $Y_T \equiv Y$.

The data structure O entails a certain time ordering, but the results obtained by the authors can be easily adapted to other choices of temporal order. Let $\bar{x}$ and $\bar{x}^*$ be two distinct regimens of the exposure and let

$$\Gamma_t^{\bar{x}^*}(m_t|\, \dot{l}_t, \overline{m}_{t-1},\, \bar{v}_{t-1}) \equiv$$
$$P(M_t(\bar{x}^*) = m_t \,|\, \overline{L}_t(\bar{x}^*) = \dot{l}_t, \overline{M}_{t-1}(\bar{x}^*) = \overline{m}_{t-1},\, \overline{V}_{t-1}(\bar{x}^*) = \bar{v}_{t-1})$$

indicate the conditional probabilities of the mediators at $t \geq 1$ if $\overline{X}$ has been set

to $\bar{x}$. At any time t, the mediator is a random draw from this distribution, denote it by $\overline{m}_t^d(\Gamma_t^{\bar{x}^*})$. From this formulation and sequential ignorability assumptions, the *conditional mediation formula* can be obtained as

$$Q_2(\bar{x}, \bar{x}^*) = \mathbb{E}\left[Y(\bar{x}, \overline{m}_t^d(\Gamma_t^{\bar{x}^*}))\right] =$$
$$\sum_{\dot{l},\bar{v},\overline{m}} yP(v_0) \prod_{t=1}^{T} P(l_t \,|\, \bar{x}_t, \dot{l}_{t-1}, \overline{m}_{t-1}, \bar{v}_{t-1}) P(m_t \,|\, \bar{x}_t^*, \dot{l}_t, \overline{m}_{t-1}, \bar{v}_{t-1}) P(v_t \,|\, \bar{x}_t, \dot{l}_t, \overline{m}_t, \bar{v}_{t-1}).$$

The interventional direct and indirect effects are identified and given by

$$IDE = \mathbb{E}\left[Y(\bar{x}, \overline{m}_t^d(\Gamma_t^{\bar{x}^*})\right] - \mathbb{E}\left[Y(\bar{x}^*, \overline{m}_t^d(\Gamma_t^{\bar{x}^*}))\right] = Q_2(\bar{x}, \bar{x}^*) - Q_2(\bar{x}^*, \bar{x}^*)$$
$$IIE = \mathbb{E}\left[Y(\bar{x}, \overline{m}_t^d(\Gamma_t^{\bar{x}}))\right] - \mathbb{E}\left[Y(\bar{x}, \overline{m}_t^d(\Gamma_t^{\bar{x}^*}))\right] = Q_2(\bar{x}, \bar{x}) - Q_2(\bar{x}, \bar{x}^*),$$

respectively. This definition leads to an additive decomposition of the total effect, as usual. The authors propose several methods for estimating the effects, specifically, they use efficient influence curves, and discuss three other methods, i.e. nested nontargeted substitution estimation, inverse probability weighting and targeted maximum likelihood estimation, see Zheng and van der Laan (2017) for more details.

Lin, Young, Logan and VanderWeele (2017) use a similar approach, specifically targeted for survival outcomes. They consider a setting akin to that described in Zheng and van der Laan (2017), including a vector V of initial baseline covariates and time-varying exposures, mediators and confounders, denoted by X, M and L, respectively. Y_t is a survival outcome taking the value 1 or 0 according to whether the patient is alive at time t or not. Notice that, differently from previous models, the outcome is not measured at the end of follow-up, but is itself time-varying. Since, also in this case, the presence of time-varying confounders violates assumption (d), the authors discuss why natural direct and indirect effects are not identifiable from data and they propose randomised interventional analogues of mediational effects, considering random draws from the mediator distribution, similarly to VanderWeele and Tchetgen Tchetgen (2017) and Zheng and van der Laan (2017).

At time $t = 1$, define $M_1^* = M_1(x_1^*)$, m_1^{d*} as a random draw of M_1^*, $Y_1^* = Y_1(x_1, m_1^{d*})$. Letting $M_t^* \equiv M_t(\bar{x}_t^*, \overline{m}_{t-1}^{d*}, \overline{Y}_{t-1}^*)$, where $\overline{m}_t^{d*} = (m_1^{d*}, m_2^{d*}, \ldots, m_t^{d*})$ and $Y_t^* \equiv Y_t(\bar{x}_t, \overline{m}_t^{d*}, \overline{Y}_{t-1}^*)$, the authors define an alternative mediation parameter[2] as $Q_3(\bar{x}, \bar{x}^*) = \mathbb{E}[Y_T^*]$. Under sequential ignorability and some consistency assumptions not reported here for the sake of brevity, the mediation parameter can be identified

[2]Here we borrow the notation used by VanderWeele and Tchetgen Tchetgen (2017).

as

$$Q_3(\bar{x}, \bar{x}^*) = \sum_{v,\overline{m}} \sum_{\dot{l}} \prod_{t=1}^{T} \mathbb{E}[Y_t \,|\, \bar{x}_t,\, \overline{m}_t,\, \dot{l}_t,\, Y_{t-1} = 1, v] \times \prod_{t=1}^{T-1} P(l_t \,|\, \bar{x}_t,\, \overline{m}_t,\, \dot{l}_{t-1},\, Y_{t-1} = 1, v)$$

$$\times \sum_{\dot{l}^{\dagger}_{T-1}} \prod_{t=1}^{T} P(m_t \,|\, \bar{x}^*_t,\, \overline{m}_{t-1},\, \dot{l}^{\dagger}_{t-1},\, \overline{Y}_{t-1} = 1, v)\, P(l^{\dagger}_{t-1} \,|\, \bar{x}^*_{t-1},\, \overline{m}_{t-1},\, \dot{l}^{\dagger}_{t-2},\, \overline{Y}_{t-2} = 1, v) P(v).$$

This expression is called the *survival mediational g-formula* and the interventional direct and indirect effects result to be identified and can be written as $Q_3(\bar{x}, \bar{x}^*) - Q_3(\bar{x}^*, \bar{x}^*)$ and $Q_3(\bar{x}, \bar{x}) - Q_3(\bar{x}, \bar{x}^*)$, respectively. The authors also suggest a parametric method to estimate the mediational g-formula based on maximum likelihood.

Interventional effects do not have the same interpretation as natural ones. A non null interventional indirect effect means that an intervention on the exposure generates a change in the distribution of the mediator, which, in turn, generates a change in the outcome. The interventional direct effect is the difference in the outcome obtained comparing its value under treatment versus its value under no treatment, setting the mediator to a random draw from its distribution with baseline or null exposure. See VanderWeele and Tchetgen Tchetgen (2017) for an example about racial health disparity. Moreover, as already remarked, VanderWeele and Tchetgen Tchetgen (2017) and Zheng and van der Laan (2017) use different distributions of the mediator from which to draw: the former consider the distribution of the mediator under the intervened exposure and given baseline covariates V, the latter use instead the distribution of the mediator under a certain intervention, conditional on the mediator history, i.e., taking into account all its predictors, for each time t.

2.3.5 Separable effects

In recent years, the separable effect approach has been applied to longitudinal mediation settings with survival outcomes (Didelez 2019*b*, Aalen et al. 2020).

Didelez (2019*b*) analyses the case of a time-to-event outcome, in particular the counting process $N(t) = \delta(Y \leq t)$, where Y is time to death; the mediator $M(t)$ is time-varying, while the exposure X is fixed at baseline and its two components are denoted by X^Y and X^M, the former affecting the outcome, the latter only the mediator.

Observationally, we have $X \equiv X^Y \equiv X^M$, but it is possible to conceive an intervention setting the two components to different values, say x and x^*. The target of inference is the interventional survival function $P(Y_{X^M=x^*, X^Y=x} > t)$[3] and the

[3]Just in this case, to avoid clutter, we write counterfactuals as subscripts instead of in brackets as done so far.

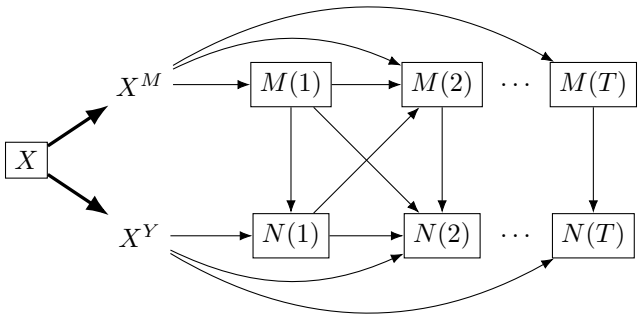

Figure 2.10: Longitudinal separable effects mediation model (Didelez 2019*b*).

separable mediational effects can be obtained as contrasts of

$$P(Y_{X^M=x^*,\,X^Y=x} > t) \;\text{ vs. }\; P(Y_{X^M=x^*,\,X^Y=x^*} > t)$$

for the direct effect and

$$P(Y_{X^M=x,\,X^Y=x} > t) \;\text{ vs. }\; P(Y_{X^M=x^*,\,X^Y=x} > t)$$

for the indirect effect.

Consider a simple scenario where the mediator and the outcome are assessed at T distinct time points. The expanded graph is shown in Figure 2.10.

For the effects to be identifiable, it is necessary to make strict assumptions about the independence of each exposure component from the variable it does not affect, that is,

A1. $M(t) \perp\!\!\!\perp X^Y \,|\, \overline{M}(t-1), \overline{N}(t-1) = 0, X^M = x^M$

A2. $N(t) \perp\!\!\!\perp X^M \,|\, \overline{M}(t), \overline{N}(t-1) = 0, X^Y = x^Y$.

It is required that X^M is independent of $N(t)$ and X^Y of $M(t)$, given past values of the mediator and the outcome count process. Assuming also that the treatment is randomised and that $P(W_{X^M=j,\,X^Y=j}) = P(W_{X=j})$ for each variable W and $\forall\, j = x, x^*$, holds (property P1 in Didelez 2019*b*), one can derive the following

$$\begin{aligned} P(Y_{X^M=x^*,\,X^Y=x} > t') = \sum_{\overline{m}(t')} \prod_{t=1}^{t'} & P(N(t) = 0 \,|\, X = x, \overline{N}(t-1) = 0, \overline{M}(t) = \overline{m}(t)) \\ & \times P(M(t) = m(t) \,|\, X = x^*, \overline{N}(t-1) = 0, \overline{M}(t-1) = \overline{m}(t-1)), \end{aligned}$$

which allows us to compute the separable effects of interest.

This approach can be extended to include exposure-induced confounders and time-varying covariates. As noted by Didelez (2019*b*), if the treatment is randomised, ad-

justing for covariates is not necessary to make the relationship between X^M and X^Y unconfounded, since this holds by design, but it is useful to make $M(t)$ and $N(t)$ independent of the opposite treatment component. Then, these variables should not be regarded as proper confounders, but as variables which may prevent the separability of the treatment components.

Aalen et al. (2020) apply these results in a survival setting including covariates. They assume a linear model for the mediators and an additive hazard model for the outcome and derive analytical expressions for survival and cumulative direct and indirect effects.

2.3.6 Other approaches

As done for the associational models, in this section, we provide a brief overview of other causal approaches which cannot be included in any of the categories previously discussed.

Some recent papers (Aalen and Frigessi 2007, Aalen et al. 2012, 2016) have pointed out the dynamic nature of mediation, as a process operating in continuous time. These authors discuss and compare the mechanistic and the interventionist conceptualisation of mediation, and suggest to model stochastic processes instead of variables. As reported in Aalen and Frigessi (2007), "When thinking of direct and indirect effects, one has in mind some process. [...] Hence, there is a need for a statistical framework that models the effects that stochastic processes have on one another". A method proposed to take into account the dynamic aspect of mechanisms is *dynamic path analysis*. Introduced by Fosen et al. (2006), dynamic path analysis considers a set of covariates $Z_1, \ldots, Z_p$ and an outcome process $Y(t)$. The relationships among variables are represented by a set of time-indexed DAGs, one for each recorded time occasion, where the vertex set is constant over time, while the edge set changes, since some links connecting variables may appear or disappear. Direct and indirect effects are defined as integrals of products of coefficients, somewhat similar to standard path analysis. Under the assumptions of no-unmeasured exposure-outcome and mediator-outcome confounders, these effects can be interpreted in a causal way, although their definition does not rely on a counterfactual framework (Strohmaier et al. 2015).

Vansteelandt et al. (2019) discuss a methodology to address longitudinal mediation in a survival setting and considering a counterfactual framework. The formula to compute the estimand of interest is partly similar to the g-formula, but, as the authors themselves remark, it does not follow from that theory. They show that their approach is a generalisation of dynamic path analysis since it can accommodate more general models for the mediator and the outcome and the inclusion of time-varying

confounders.

Finally, in recent years Bayesian inference has seen a rise, also in the field of mediation. Yuan and MacKinnon (2009) describe the advantages of this approach and show how to apply it in single-level and multilevel models. Indeed, Bayesian methods allow one to make more accurate estimates by incorporating prior information into mediation parameters, and they make it easier to compute confidence intervals for the mediated effect, since it is possible to obtain the entire posterior distribution of the effect. Miočević et al. (2018) and Daniels et al. (2012) propose methods to address causality in a Bayesian mediation framework, which is further extended by Kim et al. (2019), who explicitly address a longitudinal mediation setting through a Bayesian dynamic model.

2.3.7 Discussion

The main difference between the approaches addressed in this section is the estimand of interest and, as a consequence, the corresponding assumptions which ensure its identifiability. Bind et al. (2016) are interested in natural effects, indeed their assumptions are a modified version of (a)-(d), to accommodate the presence of random effects. Thus, they assume cross-world independence, which holds in the graph in Figure 2.6 (Fig. 1 in their paper), since there are no mediator-outcome confounders affected by the exposure. We have already discussed that, in some contexts, this assumption can be quite implausible. It is worth noting that these assumptions suffice to ensure non-parametrical identifiability just in the case of non-correlated random effects (see Appendix A for more details). PSE are a generalisation of natural effects, indeed they were developed in a NPSEM framework, the only one which permits the identification of cross-world counterfactuals.

VanderWeele and Tchetgen Tchetgen (2017), Zheng and van der Laan (2017) and Lin, Young, Logan and VanderWeele (2017) consider settings including time-varying confounders affected by the exposure, then, natural effects cannot be identified. All of them use interventional effects and, as in Bind et al. (2016), they consider static regimes. Each approach relies on its own assumptions, although the underlying logic is similar. VanderWeele and Tchetgen Tchetgen (2017) assume static sequential ignorability (a″)-(c″), conditioning also on baseline covariates V. Zheng and van der Laan (2017) consider a setting a bit more complex and require no unmeasured confounders of the relationship between X_t and all its subsequent covariates, including mediators, conditional on observed history, and no unmeasured confounders of the relationship between the mediator at each time M_t and all its subsequent covariates, conditional on observed history. They also include some positivity assumptions. Fi-

nally, the assumptions proposed by Lin, Young, Logan and VanderWeele (2017) are specific for settings with a time-to-event outcome measured at each time, and are probably more controversial than those in the other two works. The outcome is an intervening variable as well, on which it is necessary to intervene to ensure that a subject survives up to $t-1$.

Natural, path-specific and interventional effects entail an intervention on the mediator, either at individual or population level. In contrast, the separable effect approach does not require interventions on the mediators, just on the separate components of the exposure(s). This approach has only recently been applied to longitudinal settings and it seems a promising ground for future extensions. Robins et al. (2020) provide intriguing contributions by extending the previous work by Robins and Richardson (2011) in many directions: they consider the case of multiple treatments, which is a straightforward extension of the simpler case, and discuss how to identify PSE in the presence of hidden variables, proposing a two-step algorithm based on the recanting district criterion introduced by Shpitser (2013). Their extensions to multiple treatment can provide insights into the issue of time-varying treatments, since the setting is similar.

As remarked by Robins et al. (2020), the separable effects theory is very advantageous from many points of view: not only it allows for identification of effects in a wide variety of causal models, but separable components of X also have a proper causal meaning and, for substantive researchers, it may be easier to conceive interventions on these separable components than on the mediator.

However, this approach is not free from drawbacks. First, one should be willing to assume that the exposure acts along two or more different pathways. This assumption is plausible for some biological processes, but there are cases in which this decomposition is not practically nor even hypothetically conceivable. Second, each child of the exposure in the original graph should be affected by only one component of X in the expanded graph, an assumption which can be questionable in some applied contexts. Furthermore, in Didelez (2019*b*) it is assumed that the exposure is randomised, a strong assumption difficult to satisfy in observational studies. Finally, the independence of X^Y from $M(t)$ and of X^M from $N(t)$ conditionally on history could be violated in the presence of unobserved variables, for example, if an unobserved variable affects both $M(t)$ and $N(t)$. Didelez (2019*b*) remarks that an insufficient number of measurements of $M(t)$ could be another source of violation of the assumptions.

In some of the discussed approaches, the effects can be non-parametrically identified by a version of the g-formula (Robins 1986). The g-formula is one of the

g-methods, which enable us to estimate the causal effect of a time-varying exposure on an outcome of interest taking into account time-varying confounders. The g-formula has been implemented in many pieces of software, although some of them are not specific for mediation. The packages currently available are gformula in Stata (Daniel et al. 2011) and `gfoRmula` in R (Lin et al. 2020). VanderWeele and Tchetgen Tchetgen (2017) and Lin, Young, Logan and VanderWeele (2017) implement SAS codes for their g-mediation formula, the former is specific for the data set used in the paper, the latter is more general and implements the algorithm described in Section 4 of Lin, Young, Logan and VanderWeele (2017). Aalen et al. (2020) do not develop an R package, but they provide codes to replicate their study.

2.4 General discussion and conclusions

In this chapter, we have discussed the associational models and the causal approaches more frequently employed for addressing longitudinal mediation analysis in discrete time. We focused mainly on the definition of mediational effects, their interpretation, and assumptions required to identify them when analysed in causal frameworks. In this regard, we addressed two issues, violation of the cross-world independence assumption and time-varying confounders, which may hinder identification, and review alternative assumptions to overcome them.

Drawing causal inferences from observational data is not straightforward, and in the longitudinal setting, this task may become quite complicated. Several assumptions have to be made to ensure identifiability of mediational effects. These assumptions must take into account the temporal dependencies among variables and carefully ponder on which covariates to include in the model. The issue of time-varying confounders, as remarked many times, is of paramount importance. There exist several definitions of causal mediational effects, and each of them requires different sets of assumptions for being identifiable. Researchers who want to draw causal inferences from their longitudinal mediation analyses should be aware of the assumptions needed to identify the effects of interest and decide if they are plausible in the setting under study.

In Section 2.2 we have discussed some models which are not inherently causal. However, many researchers use SEMs or mixed-effect models for their analyses and comment results in causal terms without stating explicitly conditions which ensure such an interpretation. Usami et al. (2019) analyse the assumptions necessary to interpret SEM parameters as causal in the presence of two variables affecting each other reciprocally. Their considerations can be extended to the mediational setting

and models discussed so far, in the light of the considerations on identifiability made before. Notice, that the identifiability issues discussed at the end of Section 2.2 are not the same we are now addressing. Here, instead, we want to suggest which assumptions are needed for interpreting associational parameters causally and making them identifiable, parametrically or not.

It can be demonstrated that, for the CLPM shown in Figure 2.2, the natural and the separable mediational effects are nonparametrically identified (proofs are given in Appendix A). However, the LGM and LDS models include several latent variables, which prevent nonparametric identification. For these models, parametric assumptions are needed to ensure effects to be expressible as functions of observed data. Technical details can be found in Appendix A.

Another interesting aspect emerging from this review is how differently the associational and causal approach address latent variables. In SEMs and mixed-effect models, latent variables play a key role in defining the trajectories of change or encoding the dependence among temporal occasions. In causal models, latent variables are mainly a source of problems, since they often prevent the identification of mediational effects. In recent years, some scholars have started addressing the issue of estimating mediational effects in models including latent variables. Loeys et al. (2014), Loh et al. (2020) show how to estimate the controlled direct effect when variables are measured with errors. Muthén and Asparouhov (2015), Albert et al. (2016), Valeri et al. (2014) provide estimation methods for the natural mediational effects when the mediator is measured with error. Among the aforementioned studies, only the one by Loh et al. (2020) is longitudinal. In addition, most articles use latent variables as a way to tackle measurement error. However, given the flexibility and the richness of SEM, it would be interesting to endow longitudinal mediation SEMs with a causal interpretation. This issue will be addressed in Chapter 4.

To the best of our knowledge, this is the first review giving a rich overview of methods to tackle longitudinal mediation analysis. The contribution of the chapter is twofold: it uniforms the heterogeneous notation present in the literature on the topic and addresses different issues related to longitudinal mediation, showing various approaches to deal with them. However, it is worth noting some limitations. We did not address inferential issues, such as estimation methods and hypothesis testing. Moreover, we have discussed methods considering time as discrete and, in the causal framework, we focused only on static regimes. Future work will be necessary to investigate inference methods and to provide a review of continuous-time longitudinal mediation.

Table 2.1: Summary of associational models and causal approaches. For the former, we indicate whether the model has been implemented in any pieces of software, if any simulations have been carried out, and which applications have been done in the literature. For the latter, we provide the kind of effects, the corresponding estimands, the counterfactual models where identification can be achieved, and which dedicated software allow one to estimate the effects.

Associational				**Causal**			
Model	**Simulations**	**Applications**	**Software**	**Type of effects**	**Estimand**	**Model**	**Software**
CLPM	No	Psychology/Social sciences (Selig and Preacher, 2009; O'Laughlin et al., 2018)	Stata (sem) R (lavaan, OpenMx) SAS (Calis) Mplus	Natural	$\mathbb{E}[Y_t(\bar{x}, M_t(\bar{x}^*))]$	NPSEM	R (code by Mittinty and Vansteelandt 2020) SAS (code by Bind et al. 2016)
LGM	Cheong (2011)	Social sciences (Cheong et al., 2003; von Soest and Hagtvet, 2011)	Stata (sem) R (lavaan, OpenMx) SAS (Calis) Mplus	Path-specific	$\mathbb{E}[Y(\pi, \bar{x}, \bar{x}^*)]$	NPSEM	Stata (gformula) R (gfoRmula) SAS (Gformula)
LDS	Simone and Lockhart (2019)	Social science (Selig and Preacher, 2009; O'Laughlin et al., 2018)	Stata (sem) R (lavaan, OpenMx) SAS (Calis) Mplus	Interventional	$\mathbb{E}[Y_t(\bar{x}, \overline{m}^d(\bar{x}^*))]$	FFRCISTG	Stata R (intmed) SAS (macros by Lin, Young, Logan and VanderWeele 2017)
Mixed effects	Krull and MacKinnon (1999, 2001) Blood and Cheng (2011)	Psychology (Bauer et al. 2006)	Stata (me) R (lme4, nmle, glmmTMB) SAS (GLIMMIX) Mplus	Separable	$\mathbb{E}[Y_t(X^M = x^*, X^Y = x)]$	NPSEM FFRCISTG MCM Agnostic	Stata (gformula) R (gfoRmula) SAS (Gformula)

Chapter 3

Structural and multilevel mediation models: a unification

In the previous chapter we reviewed the main approaches to deal with longitudinal mediation analysis, distinguishing between the associational and the causal framework. In this chapter, we focus on the former and show that the models we addressed in Chapter 2, i.e. SEMs and mixed-effect models, can be seen as instances of a more general unified model. In the first section we prove that mixed-effect models can be written as structural equation models; in the second one we introduce the RAM notation which will be useful to write a general model encompassing SEMs and linear mixed-effect models. In Section 3 we describe the unified model and show how CLPMs, LGMs, LDS models and mixed-effect models can be formalised within this framework. Section 4 focuses on how longitudinal mediation models can be addressed in the multilevel SEM framework through definition variables. In Section 5 we discuss inferential aspects and in Section 6 we draw some conclusions.

3.1 Mixed-effect models as SEMs

Structural equation models and mixed-effect models have traditionally been addressed as separate and distinct models. They were developed at different times and stem from different traditions.

SEMs have their roots in several disciplines, and their origins can be traced back to Sewall Wright's method of path analysis, devised to address genetic questions (Wright 1920, 1921, 1934), to the advances of factor analysis in psychology (Lawley 1940, Harman 1960, Jöreskog 1969), and to the simultaneous equation models developed in econometrics by Haavelmo (1943) and Koopmans et al. (1950). The '70s represented a fundamental decade for the development of SEMs, thanks to the con-

tributions of many scholars, above all Jöreskog. In some landmark papers (Jöreskog 1970, 1973, 1978), he formalised covariance analysis and specified a model encompassing all the different instances proposed in the previous years, unifying simultaneous equation models with the possibility of including latent variables. He also developed an estimation procedure for SEM based on maximum likelihood and implemented it in a specific program called LISREL (LInear Structural RELations, Jöreskog and Sörbom 2001), which became so popular that its name ended up denoting the model specification in addition to the software.

Mixed-effect models, also known as multilevel or hierarchical models, can be seen as a generalisation of traditional regression models which can accommodate the interdependence among subjects belonging to the same group, for example people living in the same neighbourhood. These models are suited for the cases in which it is possible to distinguish two or more levels of aggregation for units, for example in a two-stage sampling, where first J level-2 units are randomly drawn (schools, departments, neighborhoods) and from each of them n_j level-1 units are drawn (students, employees, citizens). Mixed-effect models were introduced in the early '50s by Charles Henderson, an expert of genetics and animal breeding, who was trying to improve the quantitative methods to produce the best genetic characteristics in breeds (Henderson 1953). About thirty years later, Goldstein (1979) and Laird and Ware (1982) proposed mixed-effect models for longitudinal data in the context of social sciences and biometrics, respectively. The key characteristic of these models is the presence of one or more random coefficients differing across clusters, which capture the between component of units' heterogeneity. Over the decades, researchers have developed the theory underlying multilevel models to address more complex settings, like unbalanced designs and non-Gaussian variables. Reviews on the topic can be found in Verbeke and Molenberghs (2000), Raudenbush and Bryk (2002), Bickel (2007), Gelman and Hill (2007) and Snijders and Bosker (2012).

Although SEMs and mixed-effect models may appear quite different at first sight, they have a lot in common. The similarities between them have been known to scholars since the end of the last century. Meredith and Tisak (1984, 1990) were the first to acknowledge that multilevel models can be fitted in the SEM framework, focusing on the case of LGMs. In the next years, the same topic was investigated by several other researchers (Chou et al. 1998, Rovine and Molenaar 1998, 2000, 2001). Bauer (2003) and Curran (2003) extensively discuss how to translate multilevel models into the SEM notation, showing that this is possible not only for balanced data, but even for unbalanced designs through an appropriate data managing procedure. In what follows we summarise the main results in Bauer (2003) and Curran (2003),

which will be useful throughout this chapter.

Consider a basic linear mixed model with a random intercept and a random slope

$$Y_{ij} = \pi_{0j} + \pi_{1j} X_{ij} + \varepsilon_{Y_{ij}}, \tag{3.1}$$

where i and j denote level-1 (subjects) and level-2 (clusters) units, respectively. π_{0j} and π_{1j} are the random intercept and random slope, respectively, which vary for each cluster j and can be written as the sum of their expected (fixed) values and a random deviation u, as shown below:

$$\begin{aligned} \pi_{0j} &= \kappa_0 + u_{0j} \\ \pi_{1j} &= \kappa_1 + u_{1j}. \end{aligned} \tag{3.2}$$

Combining these equations yields the traditional formulation

$$Y_{ij} = (\kappa_0 + u_{0j}) + (\kappa_1 + u_{1j}) X_{ij} + \varepsilon_{Y_{ij}}. \tag{3.3}$$

Combining equations shown in (2.7) for LGMs, where subjects represent the clusters and measurement occasions the individuals, the resulting equation is

$$Y_{it} = (\mu_{\theta_0} + \zeta_{\theta_0 i}) + (\mu_{\theta_1} + \zeta_{\theta_1 i}) \lambda_t + \varepsilon_{Y_{it}},$$

from which it immediately follows the isomorphism between mixed-effect models and LGMs.

For the moment, let us assume that the design matrix $\mathbf{X}$ of the mixed-effect model contains only variables which take the same values for each subject, for example $Time$ in a longitudinal balanced case. In this setting $\mathbf{X}$ is invariant across subjects and does not need to be indexed. We will relax this assumption later. Rewriting both models in matrix form, the mixed-effect model is given by the following equations:

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\pi}_j + \boldsymbol{\varepsilon}_{yj} \quad j = 1, \ldots, J \tag{3.4}$$

$$\boldsymbol{\pi}_j = \boldsymbol{\kappa} + \mathbf{u}_j \qquad j = 1, \ldots, J \tag{3.5}$$

which combined give

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\kappa} + \mathbf{X}\mathbf{u}_j + \boldsymbol{\varepsilon}_{yj}. \tag{3.6}$$

The covariance structure implied by Equation (3.6) is

$$\boldsymbol{\Sigma}_y = \mathbf{X}\boldsymbol{\Phi}\mathbf{X}' + \boldsymbol{\Theta}_\varepsilon, \tag{3.7}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}_\varepsilon$ are the covariance matrices of $\mathbf{u}_j$ and $\boldsymbol{\varepsilon}_{yj}$.

An LGM can analogously be written as (cfr. Section 2.2.2, Chapter 2)

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\delta}_{yi} \tag{3.8}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\zeta}_i \tag{3.9}$$

which combined yield

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\mu} + \boldsymbol{\Lambda}\boldsymbol{\zeta}_i + \boldsymbol{\delta}_{yi} \tag{3.10}$$

and the implied covariance structure is

$$\boldsymbol{\Sigma}_y = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}_\delta, \tag{3.11}$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}_\delta$ are the covariance matrices of $\boldsymbol{\zeta}_i$ and $\boldsymbol{\delta}_{yi}$.

Comparing Equations (3.5) and (3.7) with Equations (3.9) and (3.11), it is evident that the two models share the same structure. Specifically, $\mathbf{X}$ corresponds to $\boldsymbol{\Lambda}$, $\boldsymbol{\kappa}$ to $\boldsymbol{\mu}$, $\boldsymbol{\Phi}$ to $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}_\varepsilon$ to $\boldsymbol{\Theta}_\delta$. These correspondences inform us on the way the parallelism can be implemented. The fact that the design matrix $\mathbf{X}$ corresponds to the factor loadings matrix $\boldsymbol{\Lambda}$ tells us that the variables assuming the role of predictors in mixed-effect models become loadings in LGMs. The vector of fixed effect $\boldsymbol{\kappa}$ in mixed models corresponds to the mean of the latent variables $\boldsymbol{\mu}$, and the vector of random effects $\mathbf{u}_j$ is analogous to the $\boldsymbol{\eta}$'s vector of errors $\boldsymbol{\zeta}_i$. A concrete example will be provided in Section 3.3.

In the balanced case, multilevel models can be written as SEMs quite straightforwardly. However, as soon as the number of level-1 predictors increases, the data managing process becomes harder and more tedious, see Curran (2003). In addition, including a continuous level-1 predictor may be troublesome, since it would imply to treat all the values assumed by this variable, but not observed in the sample, as missing (Curran 2003, Bauer 2003).

The difficulties increase when the design is unbalanced, as the equivalence between mixed-effect models and SEMs is more complex to express, although not impossible, as shown by Bauer (2003) and Curran (2003). There are two methods to address unbalanced designs, the missing-data approach and the case-varying factor loadings approach.

The former relies on the assumption that the number of subjects in each cluster equals that of the most numerous one. If a cluster has less units, the remaining values are treated as missing. For example, suppose to carry out a study on the quality of hospital care in a given city. To do this, one may sample J city hospitals and draw from each a number of patients ranging from $n_{\min}$ to $n_{\max}$, who are asked to judge the quality of their hospitalisation. The maximum number of patients sampled is

$n_{\max}$, then the missing-data approach assumes that $n_{\max}$ patients are sampled from each hospital and, if hospital j presents less patients, say $n_{\min} < n_j < n_{\max}$, the information related to the remaining $n_{\max} - n_j$ is missing. The advantage of this approach is that it entails a unique factor loading matrix, common to all clusters, and the model can be fitted by using a single covariance matrix and a single mean vector. Nonetheless, the inclusion of a continuous covariate in the model can make the estimation computationally intensive if not impossible.

To overcome this problem, it would be necessary to allow the factor loading matrix to vary over clusters, which is exactly what the case-varying factor loadings approach does. In order to accommodate the unbalanced design, each cluster can have its own factor loading matrix, which therefore is no longer common to all level-2 units. As a consequence there are as many $\mathbf{\Lambda}$'s as the number of clusters: each $\mathbf{\Lambda}_j$, $j = 1, \ldots, J$, has the dimension of the cluster and each of its column contains the values assumed by the corresponding predictor in cluster j. Considering the previous example about hospitalisation quality, it is plausible to think that the opinion on quality may depend on the severity of a patient's conditions. Then, if the model is assumed to include a random intercept and a random slope for the patients' health status, each $\mathbf{\Lambda}_j$ will include a unit column and a column with cells containing the severity of condition score of each patient sampled from hospital j. Level-1 predictors having an effect assumed as random and which are included in $\mathbf{\Lambda}$ as loadings are called *definition variables*. This concept was introduced by Mehta and West (2000), to deal with LGMs where measurement occasions differ across subjects, and by Mehta and Neale (2005) in the more general framework of multilevel SEM, which will be discussed later in this chapter.

Since we showed that mixed-effect models can be written in an SEM fashion in both balanced and unbalanced cases, we propose to see CLPMs, LGMs, LDS and mixed-effect models as special instances of a unified model that will be described in Section 3.3. First, we need to introduce the RAM notation, which is crucial to understand the rest of the chapter.

3.2 Introduction to the RAM notation

The Reticular Action Model (RAM) was developed by McArdle at the end of the '70s with the aim to generalise LISREL (Jöreskog and Sörbom 2001) and COSAN (McDonald 1978) notations, by using a compact matrix representation (McArdle 1978, 1979*a*,*b*). The key advantage of RAM over LISREL and COSAN is that it relies only on three matrices, and it has been proved that LISREL and COSAN are

subcases of RAM. A formal treatment of RAM and its algebraic properties is given in McArdle and McDonald (1984) and McArdle (2005). This section is based on these two references, but the notation is slightly changed to be consistent with the previous equations and those coming next.

Let $\boldsymbol{\eta}$ be a vector of r random variables and let

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\zeta}, \tag{3.12}$$

where $\boldsymbol{\Gamma}$ is a square matrix of coefficients γ_{ij} representing the influence of column variable $\boldsymbol{\eta}_j$ on row variable $\boldsymbol{\eta}_i$, and $\boldsymbol{\zeta}$ is a vector of residuals. We define the following $r \times r$ matrices

$$\boldsymbol{\Phi} = \mathbb{E}[\boldsymbol{\eta}\boldsymbol{\eta}'] \qquad\qquad \boldsymbol{\Psi} = \mathbb{E}[\boldsymbol{\zeta}\boldsymbol{\zeta}'] \tag{3.13}$$

the elements of which, φ_{ij} and ψ_{ij}, represent the symmetric overall association between $\boldsymbol{\eta}_i$ and $\boldsymbol{\eta}_j$, and the structural relationship between the residual variables $\boldsymbol{\zeta}_i$ and $\boldsymbol{\zeta}_j$, respectively.

Equation (3.12) can be rewritten as

$$\boldsymbol{\zeta} = \boldsymbol{\eta} - \boldsymbol{\Gamma}\boldsymbol{\eta} = (\mathbf{I}_r - \boldsymbol{\Gamma})\boldsymbol{\eta},$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix. If $(\mathbf{I}_r - \boldsymbol{\Gamma})$ is non-singular, the previous equation can equivalently be written as

$$\boldsymbol{\eta} = (\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta},$$

from which it follows that $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ can be re-expressed as

$$\begin{aligned}\boldsymbol{\Phi} &= (\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}\boldsymbol{\Psi}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1'} \\ \boldsymbol{\Psi} &= (\mathbf{I}_r - \boldsymbol{\Gamma})\boldsymbol{\Phi}(\mathbf{I}_r - \boldsymbol{\Gamma})'.\end{aligned}$$

The $\boldsymbol{\eta}$ vector can be partitioned into two subvectors $\mathbf{o}$ of p components and $\mathbf{l}$ of q components, corresponding to observed and latent variables, respectively, so that $r = p + q$.

Defining a $p \times r$ matrix $\mathbf{F}$ partitioned into a $p \times p$ identity matrix and a $p \times q$ null matrix

$$\mathbf{F} = [\mathbf{I}_p \,|\, \mathbf{0}],$$

the vector of observed variables can be obtained as

$$\mathbf{o} = \mathbf{F}\boldsymbol{\eta} = [\mathbf{I}_p \,|\, \mathbf{0}] \begin{bmatrix}\mathbf{o} \\ \mathbf{l}\end{bmatrix}. \tag{3.14}$$

$\mathbf{F}$ is a fixed matrix containing only 0 and 1, used as a *filter* to select only the observed variables from $\boldsymbol{\eta}$.

To complete the RAM model specification we can rewrite the vector of observed variables as

$$\mathbf{o} = \mathbf{F}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}\boldsymbol{\zeta} \tag{3.15}$$

and its model-implied covariance matrix as

$$\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{o}\mathbf{o}'] = \mathbf{F}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}\boldsymbol{\Psi}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1'}\mathbf{F}'.$$

Notice that, at first glance, it may seem that this specification rules out the possibility to include variable means or intercepts. In fact, they can be easily accommodated by including a constant unit vector in $\boldsymbol{\eta}$, as discussed in McArdle (2005). However, it is often easier to consider the vector of means/intercepts as separate from the factor loadings matrix, for this reason Equations (3.12) and (3.15) are rewritten as follows

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\eta} + \boldsymbol{\zeta} \\ \mathbf{o} &= \mathbf{F}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}(\boldsymbol{\mu} + \boldsymbol{\zeta}) \end{aligned} \tag{3.16}$$

where $\boldsymbol{\mu}$ is the mean/intercept vector. This is the notation adopted by the R library `OpenMx`, for example. The model-implied mean vector is then written as

$$\mathbb{E}[\mathbf{o}] = \boldsymbol{\mu}^o = \mathbf{F}(\mathbf{I}_r - \boldsymbol{\Gamma})^{-1}\boldsymbol{\mu}. \tag{3.17}$$

The matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Psi}$ and $\mathbf{F}$ are the building blocks of the RAM notation.[1] $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are patterned matrices, since their elements can be freely estimated, constrained to assume fixed values or to be functions of other parameters. In contrast $\mathbf{F}$ is fixed by the researcher, who uses it to indicate which model variables are observed.

To clarify these concepts, let us consider a one-factor model as that represented in Figure 3.1. The model includes a single latent factor ω and its three observed indicators, Y_1, Y_2 and Y_3. Equation (3.12) is given by

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \omega \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \lambda_1 \\ 0 & 0 & 0 & \lambda_2 \\ 0 & 0 & 0 & \lambda_3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \omega \end{pmatrix} + \begin{pmatrix} \zeta_{Y1} \\ \zeta_{Y2} \\ \zeta_{Y3} \\ \zeta_{\omega} \end{pmatrix}$$

[1] In the original notation proposed by McArdle and McDonald (1984) $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ are called $\mathbf{A}$ and $\mathbf{S}$, respectively. These are also the names they take in specialised software or packages, like the `OpenMx` and `sem` packages in `R`.

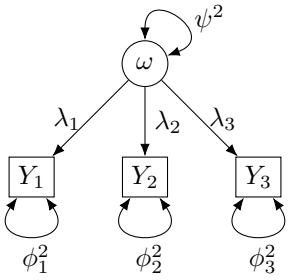

Figure 3.1: One-factor CFA model.

and $\mathbf{\Psi}$ by

$$\begin{pmatrix} \phi_1^2 & 0 & 0 & 0 \\ 0 & \phi_2^2 & 0 & 0 \\ 0 & 0 & \phi_3^2 & 0 \\ 0 & 0 & 0 & \psi^2 \end{pmatrix}.$$

In this case $\mathbf{o} = (Y_1 \;\; Y_2 \;\; Y_3)'$ and $\mathbf{l}$ contains only ω, therefore $r = 4$, the sum of $p = 3$ and $q = 1$. The filter matrix is then

$$\mathbf{F} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right).$$

As already mentioned, RAM notation was proved to be equivalent to LISREL and COSAN ones (McArdle and McDonald 1984, McArdle 2005). This allows one to rely on the same estimation theory, exploiting the RAM algebraic simplicity. Let $\mathbf{S}$ denote the sample covariance matrix and let $\boldsymbol{f}(\mathbf{S}, \mathbf{\Sigma})$ be a loss function. The parameter estimates for $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ belong to the set of solutions of the following first-order partial derivative equations

$$\begin{aligned} \partial \boldsymbol{f}(\mathbf{S}, \mathbf{\Sigma})/\partial \mathbf{\Gamma} &= \mathbf{I} - 2\,\mathrm{Vec}\left\{ [\mathbf{F}(\mathbf{I}_r - \mathbf{\Gamma})]' \mathbf{V} [\mathbf{F}(\mathbf{I}_r - \mathbf{\Gamma})\mathbf{\Psi}(\mathbf{I}_r - \mathbf{\Gamma})] \right\} = 0 \\ \partial \boldsymbol{f}(\mathbf{S}, \mathbf{\Sigma})/\partial \mathbf{\Psi} &= \mathrm{Vec}\left\{ [\mathbf{F}(\mathbf{I}_r - \mathbf{\Gamma})\mathbf{\Psi}]' \mathbf{V} [\mathbf{F}(\mathbf{I}_r - \mathbf{\Gamma})\mathbf{\Psi}] \right\} = 0, \end{aligned} \tag{3.18}$$

where $\mathrm{Vec}(\cdot)$ is the column operator and $\mathbf{V} = \partial \boldsymbol{f}/\partial \mathbf{\Sigma}$. Following McDonald and Swaminathan (1972), it is possible to write a general quadratic form of a Newton-based optimization scheme whose k-th step is

$$(\hat{\mathbf{\Gamma}}, \hat{\mathbf{\Psi}})_{k+1} = (\hat{\mathbf{\Gamma}}, \hat{\mathbf{\Psi}})_k - \alpha_k \mathbf{W}^{-1} \nabla_k$$

where $(\hat{\mathbf{\Gamma}}, \hat{\mathbf{\Psi}})_k$ is a vector of estimates for $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ at step k, α_k is a constant in the interval $(0, 1)$ to prevent descent divergence, $\mathbf{W}$ is a square symmetric positive-definite

matrix and ∇ is the gradient vector. If $\mathbf{W}$ is the Hessian matrix, i.e. $\mathbf{W} = \partial^2 \boldsymbol{f} / \partial \boldsymbol{\Sigma}^2$, the iterative procedure reduces to the traditional Newton-Raphson algorithm. A concrete example of how the procedure works can be found in McArdle (2005).

One of the most well known loss functions is that obtained from the log-likelihood of a multivariate Normal, i.e. when all variables are assumed to be Normally distributed, and is given by

$$\boldsymbol{f}(\mathbf{S}, \boldsymbol{\Sigma}) = F_{ML} = \log |\boldsymbol{\Sigma}| + \text{tr}(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log |\mathbf{S}| - (p + q), \tag{3.19}$$

where $|\cdot|$ and $\text{tr}(\cdot)$ are the determinant and the trace of a matrix, respectively. Bollen (1989) shows where this function stems from and illustrates its application to a very simple model, see Appendix 4A and 4C of his book. We will come back to Equation (3.19) in the next sections.

3.3 Unified model

Having introduced the RAM notation, we are now ready to show how SEMs and mixed-effect models can be seen as special instances of a general unified model.

The CLPM, LGM and LDS models, described in the last chapter, are widespread in the analysis of longitudinal data and the choice of the most appropriate depends on the type of data and on the interest of the researcher. Pairwise comparisons between them have been made in several papers (Serang et al. 2019, Curran and Bollen 2001, Usami et al. 2015, 2016), which have highlighted common characteristics and differences. A separate mention should be made to mixed models, which, as previously said, have traditionally been conceived as a class of models different from SEMs. For example, McNeish and Matta (2018) point out the differences between LGMs and mixed-effect models.

Indeed, all these models present strengths and drawbacks and, for this reason, researchers have tried to combine them, proposing new models. Curran and Bollen (2001), Bollen and Zimmer (2010) introduce the *autoregressive latent trajectory* model (ALT), which combines the flexibility and the capability of modelling individual trajectories of LGM with the possibility of including dependencies among variables over time as in CLPM. Hamaker et al. (2015) propose the *random intercept cross-lagged panel model* (RI-CLPM), which incorporates latent variables into the factor CLPM. Specifically for the mediation setting, Zhang et al. (2018) and Wu et al. (2018) introduce similar models, called *multilevel autoregressive mediation models* (MAMMs) and *random effects cross-lagged panel models* (RE-CLPM), respectively. They are basically autoregressive models where coefficients can be decomposed into a fixed and

a random part, as in mixed models. Zhang and Phillips (2018) further extend this idea to clustered data.

In this section, we propose a general unified model which encompasses CLPM, LGC, LDS and mixed models. Usami et al. (2019) propose a unified framework for many longitudinal models, but not for mixed effect models and not in the mediational case. The idea developed in this section partly relies on that in Bollen and Curran (2004) and Bollen and Zimmer (2010), but we extend their framework to the mediational setting and include also LDS and mixed effect models[2]. Although we present the unified model in a longitudinal framework, clearly the results are quite general and the proposed model can be used in any context with clustered data, not necessarily including repeated measures.

Let us assume a balanced design, so that the exposure, the mediator and the outcome are measured at the same T time occasions and rewrite Equations (3.14)-(3.16) in a subject-specific fashion, for each subject $i = 1, \dots, n$:

$$\boldsymbol{\eta}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i \tag{3.20}$$

$$\mathbf{o}_i = \mathbf{F}\boldsymbol{\eta}_i. \tag{3.21}$$

In extended form, these equations become

$$\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \mathbf{z}_i \\ \tilde{\mathbf{x}}_i \\ \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{y}}_i \\ \boldsymbol{\theta}_{0i} \\ \boldsymbol{\theta}_{1i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_m \\ \boldsymbol{\mu}_y \\ \boldsymbol{\mu}_z \\ \boldsymbol{\mu}_{\tilde{x}} \\ \boldsymbol{\mu}_{\tilde{m}} \\ \boldsymbol{\mu}_{\tilde{y}} \\ \boldsymbol{\mu}_{\theta_0} \\ \boldsymbol{\mu}_{\theta_1} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\Gamma}_{xx} & \boldsymbol{\Gamma}_{xm} & \boldsymbol{\Gamma}_{xy} & \boldsymbol{\Gamma}_{xz} & \boldsymbol{\Gamma}_{x\tilde{x}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{x\theta_0} & \boldsymbol{\Gamma}_{x\theta_1} \\ \boldsymbol{\Gamma}_{mx} & \boldsymbol{\Gamma}_{mm} & \boldsymbol{\Gamma}_{my} & \boldsymbol{\Gamma}_{mz} & \boldsymbol{\Gamma}_{m\tilde{x}} & \boldsymbol{\Gamma}_{m\tilde{m}} & \mathbf{0} & \boldsymbol{\Gamma}_{m\theta_0} & \boldsymbol{\Gamma}_{m\theta_1} \\ \boldsymbol{\Gamma}_{yx} & \boldsymbol{\Gamma}_{ym} & \boldsymbol{\Gamma}_{yy} & \boldsymbol{\Gamma}_{yz} & \boldsymbol{\Gamma}_{y\tilde{x}} & \boldsymbol{\Gamma}_{y\tilde{m}} & \boldsymbol{\Gamma}_{y\tilde{y}} & \boldsymbol{\Gamma}_{y\theta_0} & \boldsymbol{\Gamma}_{y\theta_1} \\ \boldsymbol{\Gamma}_{zx} & \boldsymbol{\Gamma}_{zm} & \boldsymbol{\Gamma}_{zy} & \boldsymbol{\Gamma}_{zz} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{x}\tilde{x}} & \boldsymbol{\Gamma}_{\tilde{x}\tilde{m}} & \boldsymbol{\Gamma}_{\tilde{x}\tilde{y}} & \boldsymbol{\Gamma}_{\tilde{x}\theta_0} & \boldsymbol{\Gamma}_{\tilde{x}\theta_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{m}\tilde{x}} & \boldsymbol{\Gamma}_{\tilde{m}\tilde{m}} & \boldsymbol{\Gamma}_{\tilde{m}\tilde{y}} & \boldsymbol{\Gamma}_{\tilde{m}\theta_0} & \boldsymbol{\Gamma}_{\tilde{m}\theta_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{y}\tilde{x}} & \boldsymbol{\Gamma}_{\tilde{y}\tilde{m}} & \boldsymbol{\Gamma}_{\tilde{y}\tilde{y}} & \boldsymbol{\Gamma}_{\tilde{y}\theta_0} & \boldsymbol{\Gamma}_{\tilde{y}\theta_1} \\ \boldsymbol{\Gamma}_{\theta_0 x} & \boldsymbol{\Gamma}_{\theta_0 m} & \boldsymbol{\Gamma}_{\theta_0 y} & \boldsymbol{\Gamma}_{\theta_0 z} & \boldsymbol{\Gamma}_{\theta_0\tilde{x}} & \boldsymbol{\Gamma}_{\theta_0\tilde{m}} & \boldsymbol{\Gamma}_{\theta_0\tilde{y}} & \boldsymbol{\Gamma}_{\theta_0\theta_0} & \boldsymbol{\Gamma}_{\theta_0\theta_1} \\ \boldsymbol{\Gamma}_{\theta_1 x} & \boldsymbol{\Gamma}_{\theta_1 m} & \boldsymbol{\Gamma}_{\theta_1 y} & \boldsymbol{\Gamma}_{\theta_1 z} & \boldsymbol{\Gamma}_{\theta_1\tilde{x}} & \boldsymbol{\Gamma}_{\theta_1\tilde{m}} & \boldsymbol{\Gamma}_{\theta_1\tilde{y}} & \boldsymbol{\Gamma}_{\theta_1\theta_0} & \boldsymbol{\Gamma}_{\theta_1\theta_1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \mathbf{z}_i \\ \tilde{\mathbf{x}}_i \\ \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{y}}_i \\ \boldsymbol{\theta}_{0i} \\ \boldsymbol{\theta}_{1i} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_{xi} \\ \boldsymbol{\varepsilon}_{mi} \\ \boldsymbol{\varepsilon}_{yi} \\ \boldsymbol{\varepsilon}_{zi} \\ \boldsymbol{\delta}_{\tilde{x}i} \\ \boldsymbol{\delta}_{\tilde{m}i} \\ \boldsymbol{\delta}_{\tilde{y}i} \\ \boldsymbol{\zeta}_{\theta_0 i} \\ \boldsymbol{\zeta}_{\theta_1 i} \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \mathbf{z}_i \end{pmatrix} = \begin{pmatrix} \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_v & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \mathbf{z}_i \\ \tilde{\mathbf{x}}_i \\ \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{y}}_i \\ \boldsymbol{\theta}_{0i} \\ \boldsymbol{\theta}_{1i} \end{pmatrix}$$

[2] Actually, also the other aforementioned models, ALT, MAMMs, etc., could be written in this way, but they are not the main focus of the discussion.

where $\mathbf{x}_i$, $\mathbf{m}_i$ and $\mathbf{y}_i$ are $T \times 1$ vectors of observed exposures, mediators and outcomes, $\mathbf{z}_i$ is a $v \times 1$ vector of observed covariates which may depend on the other observed variables or may interact among themselves[3], variables with tilde are the latent true scores of the observed variables and $\boldsymbol{\theta}_{0i}$ and $\boldsymbol{\theta}_{1i}$ are vectors of random intercepts and slopes, respectively, for the exposures, mediators and outcomes. For each pair of variables $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in $\boldsymbol{\eta}_i$, $\boldsymbol{\Gamma}_{w_1 w_2}$ is the matrix of coefficients describing how w_2 influences w_1. Finally, $\mathbf{I}_T$ and $\mathbf{I}_v$ are identity matrices of dimension T and v, respectively, and they are used to pick out the observed variables.

This general formulation allows us to express the first three models described in the previous chapter and linear mixed-effect models by appropriately specifying $\boldsymbol{\eta}_i$ and $\boldsymbol{\Gamma}$. The assumption of balanced design is made just to make the treatment of the topic easier to follow, but it is not necessary, as will be discussed at the end.

3.3.1 CLPM

Let us start with the CLPM including only observed variables described in Equations (2.1) - (2.3). In the proposed unified framework it can be obtained as follows

$$\underbrace{\begin{pmatrix}\mathbf{x}_i\\ \mathbf{m}_i\\ \mathbf{y}_i\end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix}\boldsymbol{\alpha}_0\\ \boldsymbol{\beta}_0\\ \boldsymbol{\gamma}_0\end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix}\boldsymbol{\Gamma}_{xx} & \mathbf{0} & \mathbf{0}\\ \boldsymbol{\Gamma}_{mx} & \boldsymbol{\Gamma}_{mm} & \mathbf{0}\\ \boldsymbol{\Gamma}_{yx} & \boldsymbol{\Gamma}_{ym} & \boldsymbol{\Gamma}_{yy}\end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix}\mathbf{x}_i\\ \mathbf{m}_i\\ \mathbf{y}_i\end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix}\boldsymbol{\varepsilon}_{xi}\\ \boldsymbol{\varepsilon}_{mi}\\ \boldsymbol{\varepsilon}_{yi}\end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where $\boldsymbol{\eta}_i$ is a $3T \times 1$ vector containing only observed variables, $\boldsymbol{\mu}$ has the same dimension of $\boldsymbol{\eta}_i$ and each entry is a T-dimensional vector $\rho \mathbf{1}_T$, $\rho \in \{\alpha_0, \beta_0, \gamma_0\}$, respectively. $\boldsymbol{\Gamma}$ is a $3T \times 3T$ factor loading matrix, common to all subjects. Examples of its submatrices are

$$\boldsymbol{\Gamma}_{xx} = \begin{pmatrix}0 & 0 & 0 & \dots & 0 & 0\\ \alpha_X & 0 & 0 & \dots & 0 & 0\\ 0 & \alpha_X & 0 & \dots & 0 & 0\\ & & \ddots & & & \\ 0 & 0 & 0 & \dots & \alpha_X & 0\end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{mx} = \begin{pmatrix}0 & 0 & 0 & \dots & 0 & 0\\ \beta_X & 0 & 0 & \dots & 0 & 0\\ 0 & \beta_X & 0 & \dots & 0 & 0\\ & & \ddots & & & \\ 0 & 0 & 0 & \dots & \beta_X & 0\end{pmatrix},$$

where it can be noticed that the first rows are null, since X_1, M_1 and Y_1 are assumed to be exogenous. It is also worth mentioning that, in the model considered, $\boldsymbol{\Gamma}_{xm}, \boldsymbol{\Gamma}_{xy}$ and $\boldsymbol{\Gamma}_{my}$ are null matrices. However, in more complex models including cross-lagged effects of the mediator or the outcome on the exposure, these matrices could be non-null.

[3] If the variables in $\mathbf{z}$ are assumed to be exogenous it is sufficient to replace the non-null matrices $\boldsymbol{\Gamma}_{z\cdot}$ in the fourth row of $\boldsymbol{\Gamma}$ with null matrices.

For the factor CLPM in Equations (2.4) - (2.6), the model is instead

$$\underbrace{\begin{pmatrix}\mathbf{x}_i\\ \mathbf{m}_i\\ \mathbf{y}_i\\ \tilde{\mathbf{x}}_i\\ \tilde{\mathbf{m}}_i\\ \tilde{\mathbf{y}}_i\end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\\ \mathbf{0}\\ \boldsymbol{\mu}_{\tilde{x}}\\ \boldsymbol{\mu}_{\tilde{m}}\\ \boldsymbol{\mu}_{\tilde{y}}\end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix}\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_T&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_T&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{I}_T\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{\tilde{x}\tilde{x}}&\mathbf{0}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{\tilde{m}\tilde{x}}&\boldsymbol{\Gamma}_{\tilde{m}\tilde{m}}&\mathbf{0}\\ \mathbf{0}&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{\tilde{y}\tilde{x}}&\boldsymbol{\Gamma}_{\tilde{y}\tilde{m}}&\boldsymbol{\Gamma}_{\tilde{y}\tilde{y}}\end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix}\mathbf{x}_i\\ \mathbf{m}_i\\ \mathbf{y}_i\\ \tilde{\mathbf{x}}_i\\ \tilde{\mathbf{m}}_i\\ \tilde{\mathbf{y}}_i\end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix}\boldsymbol{\varepsilon}_{xi}\\ \boldsymbol{\varepsilon}_{mi}\\ \boldsymbol{\varepsilon}_{yi}\\ \boldsymbol{\delta}_{\tilde{x}i}\\ \boldsymbol{\delta}_{\tilde{m}i}\\ \boldsymbol{\delta}_{\tilde{y}i}\end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where $\boldsymbol{\eta}_i$ includes the latent score vectors in addition to the observed ones and $\boldsymbol{\mu}$ contains the latent score means. The loading matrices relating observed variables with each others are null, since the relationships are at the true scores level. $\boldsymbol{\Gamma}$ contains $\boldsymbol{\Gamma}_{x\tilde{x}}$, $\boldsymbol{\Gamma}_{m\tilde{m}}$ and $\boldsymbol{\Gamma}_{y\tilde{y}}$ which are $T \times T$ identity matrices linking observed variables to their latent counterparts, and non null matrices expressing the relationships between the latent scores, such as

$$\boldsymbol{\Gamma}_{\tilde{m}\tilde{x}} = \begin{pmatrix}0&0&0&\dots&0&0\\ \beta_{\tilde{X}}&0&0&\dots&0&0\\ 0&\beta_{\tilde{X}}&0&\dots&0&0\\ &&\ddots&&&\\ 0&0&0&\dots&\beta_{\tilde{X}}&0\end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{\tilde{y}\tilde{m}} = \begin{pmatrix}0&0&0&\dots&0&0\\ \gamma_{\tilde{M}}&0&0&\dots&0&0\\ 0&\gamma_{\tilde{M}}&0&\dots&0&0\\ &&\ddots&&&\\ 0&0&0&\dots&\gamma_{\tilde{M}}&0\end{pmatrix}.$$

The error vector contains both the residuals of observed variables and the disturbances of latent true scores.

3.3.2 LGMs

As regards LGC models, let us start with the model in (2.9)-(2.12), which in the unified framework can be written as

$$\underbrace{\begin{pmatrix}X_i\\ \mathbf{m}_i\\ \mathbf{y}_i\\ \boldsymbol{\theta}_{0i}\\ \boldsymbol{\theta}_{1i}\end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix}\mu_X\\ \mathbf{0}\\ \mathbf{0}\\ \boldsymbol{\mu}_{\theta_0}\\ \boldsymbol{\mu}_{\theta_1}\end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix}0&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ 0&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{m\theta_0}&\boldsymbol{\Gamma}_{m\theta_1}\\ 0&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{y\theta_0}&\boldsymbol{\Gamma}_{y\theta_1}\\ \boldsymbol{\Gamma}_{\theta_0 x}&\mathbf{0}&\mathbf{0}&\mathbf{0}&\mathbf{0}\\ \boldsymbol{\Gamma}_{\theta_1 x}&\mathbf{0}&\mathbf{0}&\boldsymbol{\Gamma}_{\theta_1\theta_0}&\boldsymbol{\Gamma}_{\theta_1\theta_1}\end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix}X_i\\ \mathbf{m}_i\\ \mathbf{y}_i\\ \boldsymbol{\theta}_{0i}\\ \boldsymbol{\theta}_{1i}\end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix}\varepsilon_{Xi}\\ \boldsymbol{\varepsilon}_{mi}\\ \boldsymbol{\varepsilon}_{yi}\\ \boldsymbol{\zeta}_{\theta_0 i}\\ \boldsymbol{\zeta}_{\theta_1 i}\end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where $\boldsymbol{\theta}_{0i} = (\theta_{0Mi} \quad \theta_{0Yi})'$ and $\boldsymbol{\theta}_{1i} = (\theta_{1Mi} \quad \theta_{1Yi})'$, $\boldsymbol{\mu}_{\theta_0} = (\beta_0^{\theta_0} \quad \gamma_0^{\theta_0})'$ and $\boldsymbol{\mu}_{\theta_1} = (\beta_0^{\theta_1} \quad \gamma_0^{\theta_1})'$. The submatrices in $\boldsymbol{\Gamma}$ assume the following forms:

$$\boldsymbol{\Gamma}_{m\theta_0} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix} \quad \boldsymbol{\Gamma}_{m\theta_1} = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ T-1 & 0 \end{pmatrix}; \quad \boldsymbol{\Gamma}_{y\theta_0} = \begin{pmatrix} 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \quad \boldsymbol{\Gamma}_{y\theta_1} = \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & T-1 \end{pmatrix},$$

having dimension $T \times 2$, and

$$\boldsymbol{\Gamma}_{\theta_0 x} = \begin{pmatrix} \beta_X^{\theta_0} \\ \\ \gamma_X^{\theta_0} \end{pmatrix} \quad \boldsymbol{\Gamma}_{\theta_1 x} = \begin{pmatrix} \beta_X^{\theta_1} \\ \\ \gamma_X^{\theta_1} \end{pmatrix}; \quad \boldsymbol{\Gamma}_{\theta_1\theta_0} = \begin{pmatrix} 0 & 0 \\ \\ \gamma_{\theta_{0M}}^{\theta_1} & \gamma_{\theta_{0Y}}^{\theta_1} \end{pmatrix} \quad \boldsymbol{\Gamma}_{\theta_1\theta_1} = \begin{pmatrix} 0 & 0 \\ \\ \gamma_{\theta_{1M}}^{\theta_1} & 0 \end{pmatrix}.$$

$\boldsymbol{\varepsilon}_{mi}$ and $\boldsymbol{\varepsilon}_{yi}$ are $T \times 1$ error vectors for the mediator and the outcome, while $\boldsymbol{\zeta}_{\theta_0 i}$ $\boldsymbol{\zeta}_{\theta_1 i}$ are the vectors of factor disturbances. Notice that some of these $\boldsymbol{\Gamma}$ submatrices correspond to submatrices of $\boldsymbol{\Lambda}$ matrices in Equations 2.8.

Now, let us consider the more general case in which also the exposure is time-varying, for example as in the model described by O'Laughlin et al. (2018):

$$\begin{aligned}
X_{it} &= \theta_{0X_i} + \theta_{1X_i}\lambda_t + \varepsilon_{Xit} \\
\theta_{0Xi} &= \alpha_0^{\theta_0} + \zeta_{\theta_{0X}{}^i} \\
\theta_{1Xi} &= \alpha_0^{\theta_1} + \zeta_{\theta_{1X}{}^i} \\
M_{it} &= \theta_{0M_i} + \theta_{1M_i}\lambda_t + \varepsilon_{Mit} \\
\theta_{0Mi} &= \beta_0^{\theta_0} + \zeta_{\theta_{0M}{}^i} \\
\theta_{1Mi} &= \beta_0^{\theta_1} + \beta_{\theta_{0X}}^{\theta_1}\theta_{0Xi} + \zeta_{\theta_{1M}{}^i} \\
Y_{it} &= \theta_{0Y_i} + \theta_{1Y_i}\lambda_t + \varepsilon_{Yit} \\
\theta_{0Yi} &= \gamma_0^{\theta_0} + \zeta_{\theta_{0Y}{}^i} \\
\theta_{1Yi} &= \gamma_0^{\theta_1} + \gamma_{\theta_{0x}}^{\theta_1}\theta_{0Xi} + \gamma_{\theta_{1m}}^{\theta_1}\theta_{1Mi} + \zeta_{\theta_{1Y}{}^i}.
\end{aligned}$$

In matrix form the model is

$$\underbrace{\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \boldsymbol{\theta}_{0i} \\ \boldsymbol{\theta}_{1i} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\mu}_{\theta_0} \\ \boldsymbol{\mu}_{\theta_1} \end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{x\theta_0} & \boldsymbol{\Gamma}_{x\theta_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{m\theta_0} & \boldsymbol{\Gamma}_{m\theta_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{y\theta_0} & \boldsymbol{\Gamma}_{y\theta_1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\theta_1\theta_0} & \boldsymbol{\Gamma}_{\theta_1\theta_1} \end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \boldsymbol{\theta}_{0i} \\ \boldsymbol{\theta}_{1i} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_{xi} \\ \boldsymbol{\varepsilon}_{mi} \\ \boldsymbol{\varepsilon}_{yi} \\ \boldsymbol{\zeta}_{\theta_0 i} \\ \boldsymbol{\zeta}_{\theta_1 i} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

which is basically the same as before, except for the fact that $\mathbf{x}_i$ is now a vector of

repeated measures and the first row of $\mathbf{\Gamma}$ is no longer null. In addition, $\boldsymbol{\theta}_{0i}, \boldsymbol{\theta}_{1i}, \boldsymbol{\mu}_{\theta_0}$ and $\boldsymbol{\mu}_{\theta_0}$ are now 3×1 vectors, since they include the exposure latent factors and their means. The matrices relating the observed variables to the corresponding latent factors are

$$\mathbf{\Gamma}_{x\theta_0} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{m\theta_0} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 1 & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{y\theta_0} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{\Gamma}_{x\theta_1} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ T-1 & 0 & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{m\theta_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & T-1 & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{y\theta_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & T-1 \end{pmatrix},$$

while the latent factor loading matrices are

$$\mathbf{\Gamma}_{\theta_1\theta_0} = \begin{pmatrix} 0 & 0 & 0 \\ \beta^{\theta_1}_{\theta_{0X}} & 0 & 0 \\ \gamma^{\theta_1}_{\theta_{0X}} & 0 & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{\theta_1\theta_1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \gamma^{\theta_1}_{\theta_{1M}} & 0 \end{pmatrix}.$$

A graphical representation of the model for $T = 4$ is given in Figure 3.2.

3.3.3 LDS

As regards the LDS models, there are two ways to express them in the unified notation. If equations for the latent score of each variable, as discussed at the beginning of Section 2.2.2, are combined with the equations for the latent differences (2.13) - (2.15), one obtains

$$\tilde{X}_{it} = (1 + \alpha_{\tilde{X}})\tilde{X}_{t-1} + \delta_{\tilde{X}} \tag{3.22}$$

$$\tilde{M}_{it} = \beta_{\tilde{X}}\tilde{X}_{t-1} + (1 + \beta_{\tilde{M}})\tilde{M}_{it-1} + \delta_{\tilde{M}} \tag{3.23}$$

$$\tilde{Y}_{it} = \gamma_{\tilde{X}}\tilde{X}_{t-1} + \gamma_{\tilde{M}}\tilde{M}_{it-1} + (1 + \gamma_{\tilde{Y}})\tilde{Y}_{it-1} + \delta_{\tilde{Y}}, \tag{3.24}$$

which are quite similar to the factor CLPM equations, and can be written as already shown.

Alternatively, one can consider the equations as they are, including both latent

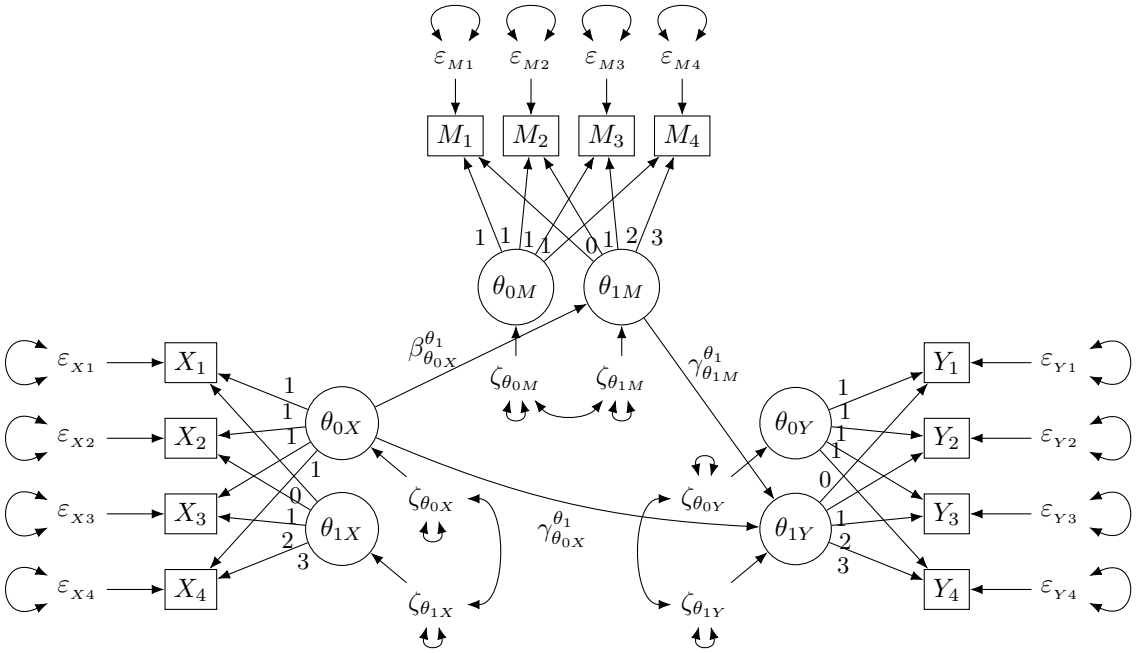

Figure 3.2: Three-process LGM with four waves.

scores and latent differences in $\boldsymbol{\eta}$, as follows:

$$\underbrace{\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \tilde{\mathbf{x}}_i \\ \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{y}}_i \\ \boldsymbol{\Delta}_x \\ \boldsymbol{\Delta}_m \\ \boldsymbol{\Delta}_y \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\mu}_{\tilde{x}} \\ \boldsymbol{\mu}_{\tilde{m}} \\ \boldsymbol{\mu}_{\tilde{y}} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_T & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{x}\tilde{x}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{x}\Delta_x} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{m}\tilde{m}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{m}\Delta_m} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{y}\tilde{y}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\tilde{y}\Delta_y} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\Delta_x\tilde{x}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\Delta_x\tilde{x}} & \boldsymbol{\Gamma}_{\Delta_m\tilde{m}} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{\Delta_x\tilde{x}} & \boldsymbol{\Gamma}_{\Delta_m\tilde{m}} & \boldsymbol{\Gamma}_{\Delta_y\tilde{y}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix} \mathbf{x}_i \\ \mathbf{m}_i \\ \mathbf{y}_i \\ \tilde{\mathbf{x}}_i \\ \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{y}}_i \\ \boldsymbol{\Delta}_x \\ \boldsymbol{\Delta}_m \\ \boldsymbol{\Delta}_y \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_{xi} \\ \boldsymbol{\varepsilon}_{mi} \\ \boldsymbol{\varepsilon}_{yi} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{\delta}_{\Delta\tilde{x}i} \\ \boldsymbol{\delta}_{\Delta\tilde{m}i} \\ \boldsymbol{\delta}_{\Delta\tilde{y}i} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where $\boldsymbol{\Delta}_x, \boldsymbol{\Delta}_m$ and $\boldsymbol{\Delta}_y$ are $(T-1)\times 1$ vectors containing the latent differences for X, M and Y. As in the factor CLPM, observed variables are linked to the corresponding latent scores through identity matrices; $\boldsymbol{\Gamma}_{\tilde{x}\tilde{x}}, \boldsymbol{\Gamma}_{\tilde{m}\tilde{m}}$ and $\boldsymbol{\Gamma}_{\tilde{y}\tilde{y}}$ are $T \times T$ matrices which have the same structure, and the same holds true for the $T \times (T-1)$ matrices $\boldsymbol{\Gamma}_{\tilde{x}\Delta_x}, \boldsymbol{\Gamma}_{\tilde{m}\Delta_m}$ and $\boldsymbol{\Gamma}_{\tilde{y}\Delta_y}$:

$$\boldsymbol{\Gamma}_{\tilde{x}\tilde{x}} = \boldsymbol{\Gamma}_{\tilde{m}\tilde{m}} = \boldsymbol{\Gamma}_{\tilde{y}\tilde{y}} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \vdots \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \boldsymbol{\Gamma}_{\tilde{x}\Delta_x} = \boldsymbol{\Gamma}_{\tilde{m}\Delta_m} = \boldsymbol{\Gamma}_{\tilde{y}\Delta_y} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \ddots & 0 & 0 \\ 0 & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

The other matrices are $(T-1) \times T$ non null matrices expressing the effects of latent scores on the latent differences, for example

$$\mathbf{\Gamma}_{\Delta_m \tilde{m}} = \begin{pmatrix} \beta_M & 0 & 0 & \dots & 0 & 0 \\ 0 & \beta_M & 0 & \dots & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & 0 & \dots & \beta_M & 0 \end{pmatrix} \qquad \mathbf{\Gamma}_{\Delta_y \tilde{m}} = \begin{pmatrix} \gamma_M & 0 & 0 & \dots & 0 & 0 \\ 0 & \gamma_M & 0 & \dots & 0 & 0 \\ & & \ddots & & & \\ 0 & 0 & 0 & \dots & \gamma_M & 0 \end{pmatrix}.$$

3.3.4 Mixed-effect models

As regards mixed-effect models, we showed the parallelism between Equations (3.4)-(3.5) and (3.8)-(3.9), and that between (3.7) and (3.11), highlighting that, in the case of balanced design, it implies the equivalence between the design matrix and the factor loading matrix. In order to deepen this issue, since a balanced design is a very unlikely event in longitudinal frameworks, we leave aside longitudinal settings just for a moment.

Bauer (2003) illustrates how to fit mixed-effect models as SEMs by considering a study involving students randomly sampled from J schools, which are then regarded as clusters. The aim of the study is to investigate differences in students' language proficiency and how the proficiency scores vary across schools. Suppose that six students, three male and three female, are randomly drawn from each school. If the mixed-effect model is assumed to include a random intercept and a random slope for the variable Sex, the model equation can be written as

$$\begin{aligned} LP_{ij} &= \pi_{0j} + \pi_{1j} S_{ij} + \varepsilon_{ij} \\ \pi_{0j} &= \kappa_0 + u_{0j} \\ \pi_{1j} &= \kappa_1 + u_{1j}, \end{aligned} \tag{3.25}$$

where $i = 1, \dots, 6$ and $j = 1, \dots, J$ denote students and schools, respectively, LP_{ij} and S_{ij} are the language proficiency score and the sex of student i in school j, respectively, and π_{0j} and π_{1j} are the random coefficients varying at the school-level. The errors ε_{ij} have constant residual variance θ^2, while u_0 and u_1 have the following covariance matrix:

$$\mathbf{\Psi} = \begin{pmatrix} \psi_0^2 & \psi_{01} \\ \psi_{01} & \psi_1^2 \end{pmatrix}.$$

Suppose to arrange the students so that the first three are male ($S = 0$) and the other three are female ($S = 1$). In the unified notation proposed above, the model

can be written as

$$\underbrace{\begin{pmatrix} LP_{1j} \\ LP_{2j} \\ LP_{3j} \\ LP_{4j} \\ LP_{5j} \\ LP_{6j} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix}}_{\boldsymbol{\eta}_j} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \kappa_0 \\ \kappa_1 \end{pmatrix}}_{\boldsymbol{\mu}} + \underbrace{\left(\begin{array}{cccccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}} \underbrace{\begin{pmatrix} LP_{1j} \\ LP_{2j} \\ LP_{3j} \\ LP_{4j} \\ LP_{5j} \\ LP_{6j} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix}}_{\boldsymbol{\eta}_j} + \underbrace{\begin{pmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \\ \varepsilon_{5j} \\ \varepsilon_{6j} \\ u_{0j} \\ u_{1j} \end{pmatrix}}_{\boldsymbol{\zeta_j}}$$

Note that $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ are not indexed, since they are constant over schools, while $\boldsymbol{\eta}_j$ includes the language proficiency scores of the six students drawn from school j and the random factors corresponding to this cluster, and $\boldsymbol{\zeta}_j$ includes the residuals of subjects and the random components of latent factors. It is worth remarking that students of a given sex are arbitrarily ordered, or, in other words, they are interchangeable within sex. This translates into a diagonal residual covariance matrix, whose diagonal elements are all constrained to be equal to θ^2.

The dependence of the outcome variable on the random factors is expressed by the non null matrix in the top-right part of $\boldsymbol{\Gamma}$, where the first column is identically 1, since it refers to the random intercept π_{0j} and the second column contains the sex of each student. Thus, Sex, which would be regarded as an explanatory variable in the traditional formulation of mixed-effect models, assumes the role of a fixed coefficient in the SEM setting. The model can then be rewritten in a more compact way as

$$\begin{pmatrix} \mathbf{LP}_j \\ \boldsymbol{\pi}_j \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\kappa} \end{pmatrix} + \left(\begin{array}{c|c} \mathbf{0}_{6\times 6} & [\mathbf{1} : \boldsymbol{S}] \\ \hline \mathbf{0}_{2\times 6} & \mathbf{0}_{2\times 2} \end{array}\right) \begin{pmatrix} \mathbf{LP}_j \\ \boldsymbol{\pi}_j \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_j \\ \mathbf{u}_j \end{pmatrix},$$

where $\mathbf{LP}_j$ is the vector of proficiency scores for students in school j, $\boldsymbol{\pi}_j$ and $\boldsymbol{\kappa}$ are the latent coefficient vector and its mean vector, respectively. $[\mathbf{1} : \boldsymbol{S}]$ is a 6×2 matrix, where the first column contains only 1 and the second the sex of each student drawn from school j. Since the design is balanced, $[\mathbf{1} : \boldsymbol{S}]$ does not differ across schools. Figure 3.3 provides a graphical representation of the model.

This notation allows for the inclusion of level-2 predictors as well. For example, suppose to want to include the school size in the model. It can be simply added to $\boldsymbol{\eta}$ as another observed variable and it can affect either π_0 and π_1. In the former case

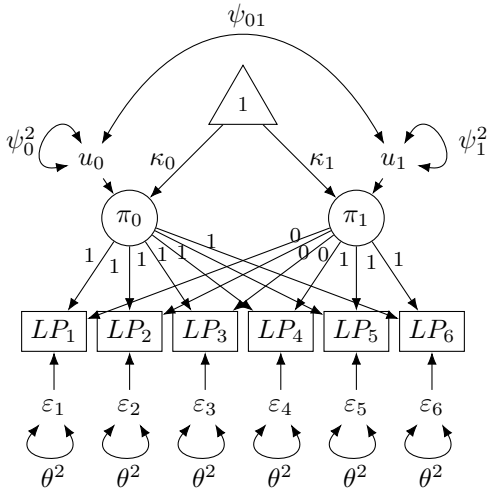

Figure 3.3: Model in Equations (3.25) represented as an SEM.

the equation for π_0 becomes

$$\pi_{0j} = \kappa_0 + \beta_0 \cdot size_j + u_{0j}$$

so that the reduced-form model for language proficiency is

$$LP_{ij} = (\kappa_0 + u_{0j}) + (\kappa_1 + u_{1j})S_{ij} + \beta_0 \cdot size_j + \varepsilon_{ij}.$$

Then, it is sufficient to include $size_j$ in $\boldsymbol{\eta}_j$ and β_0 in the factor loading matrix as follows

$$\begin{pmatrix} size_j \\ LP_{1j} \\ LP_{2j} \\ LP_{3j} \\ LP_{4j} \\ LP_{5j} \\ LP_{6j} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix} = \begin{pmatrix} \mu_{size} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \kappa_0 \\ \kappa_1 \end{pmatrix} + \left(\begin{array}{ccccccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ \hline \beta_0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \begin{pmatrix} size_j \\ LP_{1j} \\ LP_{2j} \\ LP_{3j} \\ LP_{4j} \\ LP_{5j} \\ LP_{6j} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix} + \begin{pmatrix} \varepsilon_{size} \\ \varepsilon_{1j} \\ \varepsilon_{2j} \\ \varepsilon_{3j} \\ \varepsilon_{4j} \\ \varepsilon_{5j} \\ \varepsilon_{6j} \\ u_{0j} \\ u_{1j} \end{pmatrix}.$$

Notice that *size* is regarded as an exogenous variable. It can analogously be included as a predictor of π_{1j}

$$\pi_{1j} = \kappa_1 + \beta_1 \cdot size_j + u_{1j}$$

and the reduced-form outcome model is

$$LP_{ij} = (\kappa_0 + u_{0j}) + (\kappa_1 + u_{1j}) \cdot S_{ij} + \beta_0 \cdot size_j + \beta_1 \cdot size_j \cdot S_{ij} + \varepsilon_{ij},$$

where it can be noticed the presence of a cross-level interaction between sex (level 1) and size (level 2).

As mentioned in Section 3.1, the inclusion of a continuous level-1 predictor makes the design unbalanced, which requires the case-varying factor loading approach. Continuing the example in Bauer (2003), suppose to add verbal intelligence in the model, as a predictor of language proficiency. Verbal intelligence, denoted by V, is a continuous score measured at the subject level. Moreover, consider a more general case in which the number of students sampled from each school varies, so that each cluster has its own dimension n_j. The model can be rewritten as

$$\underbrace{\begin{pmatrix} LP_{1j} \\ LP_{2j} \\ \vdots \\ LP_{n_jj} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix}}_{\boldsymbol{\eta}_j} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \kappa_0 \\ \kappa_1 \end{pmatrix}}_{\boldsymbol{\mu}_j} + \underbrace{\left(\begin{array}{cccccc|cc} 0 & 0 & 0 & 0 & 0 & 0 & 1 & V_{1j} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & V_{2j} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & V_{n_jj} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}_j} \underbrace{\begin{pmatrix} LP_{1j} \\ LP_{2j} \\ \vdots \\ LP_{n_jj} \\ \pi_{0j} \\ \pi_{1j} \end{pmatrix}}_{\boldsymbol{\eta}_j} + \underbrace{\begin{pmatrix} \varepsilon_{1j} \\ \varepsilon_{2j} \\ \vdots \\ \varepsilon_{n_jj} \\ u_{0j} \\ u_{1j} \end{pmatrix}}_{\boldsymbol{\zeta}_j}$$

where $\boldsymbol{\eta_j}$ and $\boldsymbol{\mu}_j$ have dimension $(n_j + 2) \times 1$ and V_{ij}, $i = 1, \ldots, n_j$, is the value of verbal intelligence for subject i in cluster j.

So far we have then proved that both balanced- and unbalanced-design multilevel models can be written as SEMs. Longitudinal data are very likely to show an unbalanced design, thus the definition variable approach is useful to deal with them in an SEM framework. Since the results obtained within SEMs are basically equivalent to those obtained through multilevel models, see Bauer (2003), Curran (2003), one may ask why the translation from mixed-effect to SEM should be worth the effort. Curran (2003) devotes a section of his paper, very appropriately entitled 'Why bother?', to answer this question. He proposes two reasons why the translation is meaningful: first, it helps us to better understand both modeling frameworks; second, it gives hints about the way to exploit the peculiarities of each approach. This latter motivation is particularly important in the mediational framework, which we have not addressed yet and which will be the focus of the next section. Indeed, mediation in the multilevel framework has some limitations, already highlighted in Section 2.2.3, i.e. the impossibility to model bottom-up effects and the difficulties connected to the estimation of the covariance between the mediator and the outcome random factors. These shortcomings can be effectively overcome in the SEM framework due to its capacity of incorporating latent variables in the model, as will be extensively discussed in the next section.

3.4 Multilevel SEM approach to mediation models through definition variables

Our proposal for unifying SEM and multilevel models is far from being the first. Skrondal and Rabe-Hesketh (2004) propose an extremely general framework capable of modeling response variables belonging to the natural exponential class, with the possibility of incorporating many hierarchical levels and either discrete or continuous latent variables. They developed a Stata package called GLLAMM based on their approach. More recently, inspired by Huber et al. (2004), Niku et al. (2017, 2019) develop an approach to address the nested structure of some kinds of ecological data, like species counts and biomass. Their model can address non-Normally distributed variables as well, and it is implemented in the `R` package `gllvm`.

However, the most widely known approach is that of multilevel SEM (MSEM), which capitalises on the work of McDonald and Goldstein (1989), Muthén and Satorra (1989) and Muthén (1989, 1991, 1994). This approach relies on the decomposition of each observed variable into two orthogonal components, a between and a within component. This decomposition is reflected also in the sample covariance matrix, which can analogously be written as the sum of a between and a within covariance matrix (Muthén 1994). Preacher et al. (2010) and Preacher et al. (2011) use this framework to construct a general approach to multilevel mediation which overcomes the traditional problems highlighted by Krull and MacKinnon (2001) about the impossibility of fitting models with *bottom-up effects* (level-1 variables affecting level-2 variables), and by Pituch et al. (2006, 2010) about the difficulties related to software implementation. The approach by Preacher et al. (2010) disentangles the within component of the indirect effect from the between component.

In the next section we will briefly introduce the MSEM framework and show how the approach to multilevel mediation by Preacher et al. works.

3.4.1 A brief introduction to MSEM and the approach to mediation proposed by Preacher

We will use the LISREL instead of the RAM notation, since it was the one originally employed when the MSEM approach was developed and it is that still adopted by textbooks and scientific papers on the topic.

The measurement model is defined as

$$\mathbf{y}_{ij} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{ij} + \boldsymbol{\varepsilon}_{ij} \tag{3.26}$$

where $\mathbf{y}_{ij}$ is a $p \times 1$ vector of observed variables for subject i in cluster j, $\boldsymbol{\eta}_{ij}$ is a $q \times 1$ vector of latent factors, $\boldsymbol{\Lambda}$ is a $p \times q$ factor loading matrix and $\boldsymbol{\varepsilon}_{ij}$ is the p-dimensional residual vector.

The structural model is

$$\boldsymbol{\eta}_{ij} = \boldsymbol{\nu} + \boldsymbol{\eta}_{Bj} + \boldsymbol{\eta}_{Wij}, \tag{3.27}$$

where $\boldsymbol{\nu}$ is the grand mean of $\boldsymbol{\eta}_{ij}$, $\boldsymbol{\eta}_{Bj}$ is a random factor characterising clusters and $\boldsymbol{\eta}_{Wij}$ is a random factor varying over individuals in cluster j. It can be proved that $\boldsymbol{\eta}_{Bj}$ and $\boldsymbol{\eta}_{Wij}$ are independent, therefore the covariance matrix of $\boldsymbol{\eta}_{ij}$ can be written as

$$V(\boldsymbol{\eta}_{ij}) = \boldsymbol{\Psi}_T = \boldsymbol{\Psi}_B + \boldsymbol{\Psi}_W, \tag{3.28}$$

and, in the same fashion, the residual covariance matrix can be decomposed into a between and a within component

$$V(\boldsymbol{\varepsilon}_{ij}) = \boldsymbol{\Theta}_T = \boldsymbol{\Theta}_B + \boldsymbol{\Theta}_W. \tag{3.29}$$

This decomposition makes the specification of separate models for each level quite natural, so that observed or latent variable can be included as predictors of either between or within factors. A general two-level model can then be reformulated as

$$\mathbf{w}_{ij} = \begin{pmatrix} \mathbf{x}_j \\ \\ \mathbf{y}_{ij} \end{pmatrix} = \boldsymbol{\omega}_j + \boldsymbol{\omega}_{ij} = \begin{pmatrix} \boldsymbol{\omega}_{xj} \\ \\ \boldsymbol{\omega}_{yj} \end{pmatrix} + \begin{pmatrix} 0 \\ \\ \boldsymbol{\omega}_{yij} \end{pmatrix}, \tag{3.30}$$

where $\mathbf{w}_{ij}$ may contain either cluster-level variables $\mathbf{x}_j$ or variables measured at the individual level $\mathbf{y}_{ij}$. Both kinds of variables can be seen as the sum of their between and within latent components, which in turn can be modeled as follows:

$$\boldsymbol{\omega}_j = \boldsymbol{\alpha}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\varepsilon}_{Bj} \tag{3.31}$$

$$\boldsymbol{\eta}_{Bj} = \boldsymbol{\nu}_B + \boldsymbol{\Gamma}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\zeta}_{Bj} \tag{3.32}$$

$$\boldsymbol{\omega}_{ij} = \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\varepsilon}_{Wij} \tag{3.33}$$

$$\boldsymbol{\eta}_{Wij} = \boldsymbol{\Gamma}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\zeta}_{Wij}. \tag{3.34}$$

Combining equations (3.31) and (3.33) a more general expression for $\boldsymbol{w}_{ij}$ can be written as:

$$\mathbf{w}_{ij} = \boldsymbol{\alpha}_B + \boldsymbol{\Lambda}_B \boldsymbol{\eta}_{Bj} + \boldsymbol{\varepsilon}_{Bj} + \boldsymbol{\Lambda}_W \boldsymbol{\eta}_{Wij} + \boldsymbol{\varepsilon}_{Wij} \tag{3.35}$$

The general mean and covariance structures are given by

$$\boldsymbol{\mu} = \boldsymbol{\alpha}_B + \boldsymbol{\Lambda}_B(\mathbf{I} - \boldsymbol{\Gamma}_B)^{-1}\boldsymbol{\nu}_B \tag{3.36}$$

$$\boldsymbol{\Sigma}_B = \boldsymbol{\Lambda}_B(\mathbf{I} - \boldsymbol{\Gamma}_B)^{-1}\boldsymbol{\Psi}_B(\mathbf{I} - \boldsymbol{\Gamma}_B)^{-1'}\boldsymbol{\Lambda}'_B + \boldsymbol{\Theta}_B \tag{3.37}$$

$$\boldsymbol{\Sigma}_W = \boldsymbol{\Lambda}_W(\mathbf{I} - \boldsymbol{\Gamma}_W)^{-1}\boldsymbol{\Psi}_W(\mathbf{I} - \boldsymbol{\Gamma}_W)^{-1'}\boldsymbol{\Lambda}'_W + \boldsymbol{\Theta}_W. \tag{3.38}$$

An extensive treatment of MSEM, with graphical representations and empirical examples can be found in Heck (2009) and Heck and Thomas 2020 (Chapter 7). See also Lee (1990), Kaplan and Elliott (1997) and Kaplan (2009) for a discussion on estimation methods.

Preacher et al. (2010) develop multilevel mediation analysis in the MSEM framework, highlighting its main advantage over the traditional multilevel setting, i.e. the possibility to address also outcome variables assessed at the cluster level, which Krull and MacKinnon (2001) had excluded. In addition, this allows researchers to distinguish the between part of the indirect effect from the within part. Below we will give a short introduction on this approach, changing the original notation used by Preacher et al. (2010) and Preacher et al. (2011) and employing that previously introduced in this thesis, in order to better distinguish the between and within part of the model.

Let us consider the 1-1-1 design, where the exposure, the mediator and the outcome are measured at level 1.

$$\begin{aligned}
\boldsymbol{\omega}_{ij} &= \boldsymbol{\Lambda}_{Wij}\boldsymbol{\eta}_{Wij} + \boldsymbol{\Lambda}_{Bj}\boldsymbol{\eta}_{Bj} \\
&= \begin{pmatrix} X_{ij} \\ M_{ij} \\ Y_{ij} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{pmatrix} \\
\boldsymbol{\eta}_{Wij} &= \boldsymbol{\Gamma}_W\boldsymbol{\eta}_{Wij} + \boldsymbol{\zeta}_{Wij} \\
&= \begin{pmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ \Gamma_{MX} & 0 & 0 \\ \Gamma_{YX} & \Gamma_{YM} & 0 \end{pmatrix} \begin{pmatrix} \eta_{Xij} \\ \eta_{Mij} \\ \eta_{Yij} \end{pmatrix} + \begin{pmatrix} \zeta_{\eta Xij} \\ \zeta_{\eta Mij} \\ \zeta_{\eta Yij} \end{pmatrix} \\
\boldsymbol{\eta}_{Bj} &= \boldsymbol{\nu}_B + \boldsymbol{\Gamma}_B\boldsymbol{\eta}_{Bj} + \boldsymbol{\zeta}_{Bj} \\
&= \begin{pmatrix} \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{pmatrix} = \begin{pmatrix} \nu_{\eta Xj} \\ \nu_{\eta Mj} \\ \nu_{\eta Yj} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ \gamma_{MX} & 0 & 0 \\ \gamma_{YX} & \gamma_{YM} & 0 \end{pmatrix} \begin{pmatrix} \eta_{Xj} \\ \eta_{Mj} \\ \eta_{Yj} \end{pmatrix} + \begin{pmatrix} \zeta_{\eta Xj} \\ \zeta_{\eta Mj} \\ \zeta_{\eta Yj} \end{pmatrix}
\end{aligned}$$

These equations present some particular features, which will be analysed in turn. The first equation shows that each observed variable is the sum of a within and a

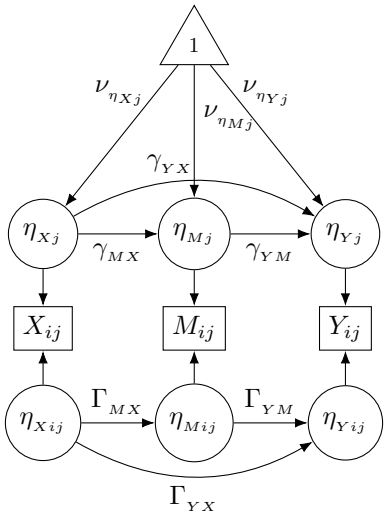

Figure 3.4: Graphical representation of 1-1-1 mediation in the approach of Preacher et al. (2010). Error terms have been omitted to avoid cluttering.

between latent component. The second and the third equations are the within and between structural models, respectively. Notice that the between measurement model is missing, since there are no observed variables at level 2. The within factors have zero mean and are linked to each other through the coefficients in matrix $\mathbf{\Gamma}_W$[4]. The between factors have intercept $\boldsymbol{\nu}_{Bj}$ and their reciprocal influences are expressed in matrix $\mathbf{\Gamma}_B$.

The between indirect effect is defined as the product $\gamma_{MX}\gamma_{YM}$. As regards the within indirect effect, if the loadings in $\mathbf{\Gamma}_W$ are fixed, it is simply the product $\Gamma_{MX}\Gamma_{YM}$. Vice versa, if the loadings are random, the within indirect effect is given by $\mu_{\Gamma MXj}\mu_{\Gamma YMj} + \psi_{\zeta_{MXj}-\zeta_{YMj}}$, where $\mu_{\Gamma MXj}$ and $\mu_{\Gamma YMj}$ are the means of Γ_{MXj} and Γ_{YMj}, respectively, and $\psi_{\zeta_{MXj}-\zeta_{YMj}}$ is the covariance between the random disturbances of Γ_{MXj} and Γ_{YMj}.

Other possible multilevel mediation designs are discussed in Preacher et al. (2010) and Pituch and Stapleton (2011).

3.4.2 The definition variable approach to longitudinal mediation

The approach proposed by Preacher et al. (2010) is very general and has the advantage to decompose clearly the indirect effect in its within and between components. Sometimes, however, this is not of interest for the researchers, which may prefer a model more similar to the classical multilevel ones, but keeping the strengths of SEMs. The definition variable approach seems to satisfy this need.

[4]Preacher et al. (2010) allow $\mathbf{\Gamma}_W$ to vary at the cluster level, i.e. to include random coefficients. For the sake of simplicity, we present a case where the within factor loadings are fixed.

As already said in Section 3.1, the most general way to translate a mixed-effect model into an SEM is to include the variables playing the role of predictors in multilevel models into the factor loading matrices. They become coefficients in factor loading matrices varying at the cluster level and are called definition variables (Mehta and Neale 2005). Basically, this entails a change of perspective: if in the traditional mixed-effect setting there are predictors the effect of which may vary over clusters, in the MSEM framework the latent variables, regarded as random coefficients in the mixed-effect framework, affect the outcome, and their effect is expressed by the predictors in the classic multilevel setting. To the best of our knowledge, this approach has never been proposed for mediation analysis, except for a brief mention in Preacher (2011), where it is used in a three-level setting, in the very specific case of a 1-1-1 mediation.

In this section we want to develop the definition variable approach for longitudinal mediation analysis, where the individuals are level-2 units and the measurement occasions are level-1 units. In longitudinal mediation, at least one between the mediator and the outcome changes over time; for this reason, traditional designs are 1-1-1, described before, 2-1-1, where the exposure is measured at baseline, and 2-2-1, where only the outcome changes over time. However, there are other plausible designs in a longitudinal context, such as 1-2-1 and 1-1-2, but they have never been addressed within the multilevel framework, since two variables playing the role of outcomes are measured at level 2. As we will show in the next sections, this issue can be overcome through MSEMs. In addition, although the primary focus of this thesis is longitudinal mediation, we will address the other multilevel mediation designs as well (2-1-2, 1-2-2), even if they are not common in longitudinal settings, in order to provide a general approach to multilevel mediation within the SEM framework. We will not address the 2-2-2 design, since it reduces to a simple non-longitudinal mediation model where the sample is made up of level-2 units.

In what follows, we will use the same notation employed in Chapter 2 and Section 3.3, denoting by β's the coefficients for the mediator model and by γ's the coefficients for the outcome model.

1-1-1 design

As already said, in this design all variables are measured over time. For example, consider an observational study carried out for ten years on a sample of 500 patients to measure the effects of air pollution on health. The average level of PM_{10} concentration is measured monthly in the area where the subjects live, and some researchers are interested in its effect on the levels of subjects' ICAM-1 protein through their ICAM-1

DNA methylation. This setting is similar to that described in Bind et al. (2016).

Let us consider the following models, for $i = 1, \ldots, n$ subjects and $t = 1, \ldots, T_i$ time occasions (if $T_i = T \; \forall \, i$, the design is balanced)

$$M_{it} = \underbrace{(\beta_0 + b_{0i})}_{\pi_{0Mi}} + \underbrace{(\beta_X + b_{Xi})}_{\pi_{MXi}} X_{it} + \varepsilon_{Mit} \tag{3.39}$$

$$Y_{it} = \underbrace{(\gamma_0 + g_{0i})}_{\pi_{0Yi}} + \underbrace{(\gamma_X + g_{Xi})}_{\pi_{YXi}} X_{it} + \underbrace{(\gamma_M + g_{Mi})}_{\pi_{YMi}} M_{it} + \varepsilon_{Yit}. \tag{3.40}$$

The mediator model includes a random intercept and a random slope for the exposure, while the outcome model includes a random intercept and two random slopes, one for the exposure and one for the mediator. In the traditional mixed-effect perspective, what can be read off from these models is that there are some subject-level predictors, the effects of which vary across individuals, as shown by the presence of random coefficients π's. From an SEM perspective instead, there are latent variables π's affecting the mediator and the outcome, the effects of which are either fixed (as is the case for random intercepts) or differ across individuals, and their values are determined by the definition variables corresponding to the predictors in the mixed-effect framework.

Using the RAM notation introduced in Section 3.2, the model can be written as follows

$$\underbrace{\begin{pmatrix} M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Mi} \\ \pi_{MXi} \\ \pi_{0Yi} \\ \pi_{YXi} \\ \pi_{YMi} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \beta_0 \\ \beta_X \\ \gamma_0 \\ \gamma_X \\ \gamma_M \end{pmatrix}}_{\boldsymbol{\mu}_i} + \underbrace{\left(\begin{array}{cccccc|cc|ccc} 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & X_{i1} & 0 & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & X_{iT_i} & 0 & 0 & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & X_{i1} & M_{i1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & X_{iT_i} & M_{iT_i} \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}_i} \underbrace{\begin{pmatrix} M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Mi} \\ \pi_{MXi} \\ \pi_{0Yi} \\ \pi_{YXi} \\ \pi_{YMi} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \varepsilon_{Mi1} \\ \vdots \\ \varepsilon_{MiT_i} \\ \varepsilon_{Yi1} \\ \vdots \\ \varepsilon_{YiT_i} \\ b_{0i} \\ b_{Xi} \\ g_{0i} \\ g_{Xi} \\ g_{Mi} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where $\boldsymbol{\eta}_i$ contains the repeated measures of the mediator and the outcome for subject i, and the subject-specific random intercepts and slopes. The central top panel of matrix $\boldsymbol{\Gamma}_i$ reports the coefficients of the mediator random factors on the mediator, and the central panel on the left shows the outcome random factor loadings.

Using definition variables implies that, not only the factor loading, but also the

covariance matrix is subject-specific:

$$\boldsymbol{\Psi}_i = \begin{pmatrix} \boldsymbol{\Psi}_i^W & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_i^{\boldsymbol{B}} \end{pmatrix}.$$

$\boldsymbol{\Psi}_i$ is a block diagonal matrix, where the first block refers to the within (level-1) covariance matrix and the second to the between (level-2) covariance matrix. In design 1-1-1, $\boldsymbol{\Psi}_i^W$ can be further decomposed into two blocks

$$\begin{pmatrix} \boldsymbol{\Theta}_{\varepsilon_{Mi}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_{\varepsilon_{Yi}} \end{pmatrix},$$

where $\boldsymbol{\Theta}_{\varepsilon_{Mi}}$ and $\boldsymbol{\Theta}_{\varepsilon_{Yi}}$ are the mediator and outcome residual covariance matrices, respectively. They are assumed to be diagonal and the diagonal elements are constrained to be equal, within each matrix:

$$\boldsymbol{\Theta}_{\varepsilon_{Mi}} = \mathbf{I}_{T_i} \otimes \theta^2_{M} = \begin{pmatrix} \theta^2_{M} & & & \\ 0 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \theta^2_{M} \end{pmatrix} \qquad \boldsymbol{\Theta}_{\varepsilon_{Yi}} = \mathbf{I}_{T_i} \otimes \theta^2_{Y} = \begin{pmatrix} \theta^2_{Y} & & & \\ 0 & \ddots & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & \theta^2_{Y} \end{pmatrix}.$$

$\boldsymbol{\Psi}_{\boldsymbol{i}}^{\boldsymbol{B}}$ is the covariance matrix of the latent factor disturbances (level-2 variables) and, contrary to the residual matrices above, it can be non-diagonal or, in other words, disturbances of factors are free to covary:

$$\boldsymbol{\Psi}_{\boldsymbol{i}}^{\boldsymbol{B}} = \begin{pmatrix} \psi^2_{b_0} & & & & \\ \psi_{b_0 b_X} & \psi^2_{b_X} & & & \\ \psi_{b_0 g_0} & \psi_{b_X g_0} & \psi^2_{g_0} & & \\ \psi_{b_0 g_X} & \psi_{b_X g_X} & \psi_{g_0 g_X} & \psi^2_{g_X} & \\ \psi_{b_0 g_M} & \psi_{b_X g_M} & \psi_{g_0 g_M} & \psi_{g_X g_M} & \psi^2_{g_M} \end{pmatrix}.$$

An interesting aspect of this approach is that the variances and covariances of repeated measures within the same individual are functions of the predictors used as definition variables. Following Mehta and Neale (2005), it is easy to prove that, for each $i = 1, \ldots, n$ and $s, t = 1, \ldots, T_i$,

$$\begin{aligned} \mathrm{V}(M_{it}) &= \mathrm{V}(b_{0i}) + X^2_{it}\mathrm{V}(b_{Xi}) + 2X_{it}\mathrm{Cov}(b_{0i}, b_{Xi}) + \mathrm{V}(\varepsilon_{Mit}) \\ &= \psi^2_{b_0} + X^2_{it}\psi^2_{b_X} + 2X_{it}\psi_{b_0 b_X} + \theta^2_{M} \end{aligned} \tag{3.41}$$

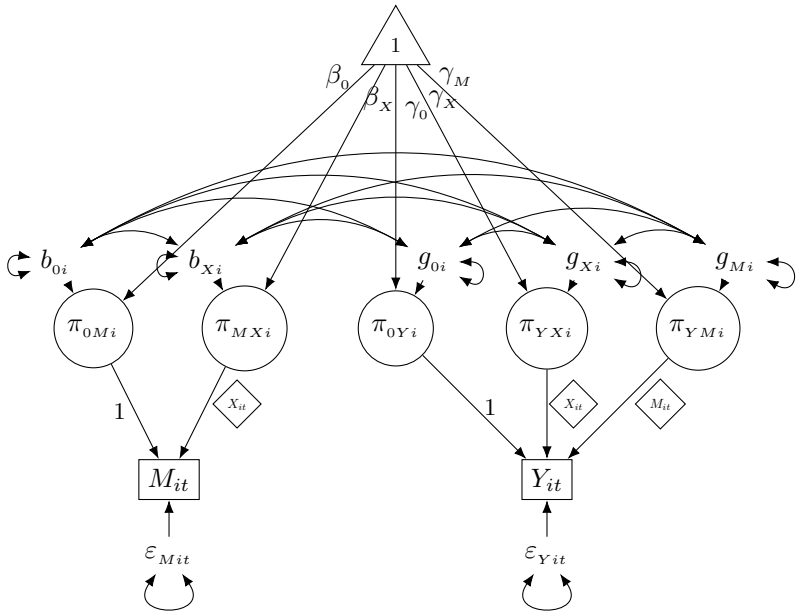

Figure 3.5: 1-1-1 design expressed through definition variables.

$$\begin{aligned}\mathrm{Cov}(M_{is}, M_{it}) &= \mathrm{V}(b_{0i}) + X_{is}X_{it}\mathrm{V}(b_{Xi}) + (X_{is} + X_{it})\mathrm{Cov}(b_{0i}, b_{Xi}) \\ &= \psi^2_{b_0} + X_{is}X_{it}\psi^2_{b_X} + (X_{is} + X_{it})\psi_{b_0 b_X}\end{aligned} \tag{3.42}$$

$$\begin{aligned}\mathrm{V}(Y_{it}) &= \mathrm{V}(g_{0i}) + X^2_{it}\mathrm{V}(g_{Xi}) + M^2_{it}\mathrm{V}(g_{Mi}) + 2X_{it}\mathrm{Cov}(g_{0i}, g_{Xi}) \\ &\quad + 2M_{it}\mathrm{Cov}(g_{0i}, g_{Mi}) + 2X_{it}M_{it}\mathrm{Cov}(g_{Xi}, g_{Mi}) + \mathrm{V}(\varepsilon_{Yit}) \\ &= \psi^2_{g_0} + X^2_{it}\psi^2_{g_X} + M^2_{it}\psi^2_{g_M} + 2X_{it}\psi_{g_0 g_X} \\ &\quad + 2M_{it}\psi_{g_0 g_M} + 2X_{it}M_{it}\psi_{g_X g_M} + \theta^2_Y\end{aligned} \tag{3.43}$$

$$\begin{aligned}\mathrm{Cov}(Y_{is}, Y_{it}) &= \mathrm{V}(g_{0i}) + X_{is}X_{it}\mathrm{V}(g_{Xi}) + M_{is}M_{it}\mathrm{V}(g_{Mi}) \\ &\quad + (X_{is} + X_{it})\mathrm{Cov}(g_{0i}, g_{Xi}) + (M_{is} + M_{it})\mathrm{Cov}(g_{0i}, g_{Mi}) \\ &\quad + (X_{it}M_{is} + X_{is}M_{it})\mathrm{Cov}(g_{Xi}, g_{Mi}) \\ &= \psi^2_{g_0} + X_{is}X_{it}\psi^2_{g_X} + M_{is}M_{it}\psi^2_{g_M} + (X_{is} + X_{it})\psi_{g_0 g_X} \\ &\quad + (M_{is} + M_{it})\psi_{g_0 g_M} + (X_{it}M_{is} + X_{is}M_{it})\psi_{g_X g_M}\end{aligned} \tag{3.44}$$

where it can be noticed the presence of X_{it} and M_{it}.

A graphical representation of the model is shown in Figure 3.5, where diamonds represent definition variables. It is worth mentioning that, in spite of this somewhat uncommon formulation of multilevel models as SEMs, the meaning of estimated coefficients is the same and the indirect effect can be obtained with the traditional formula $\beta_X\gamma_M + \psi_{b_X g_M}$, presented also in Bauer et al. (2006).

2-1-1 design

Let us move to a 2-1-1 design, where the exposure is measured at baseline while the mediator and the outcome are assessed over time. For instance, consider an experimental setting where a group of patients suffering from vascular inflammation are randomly assigned to a new therapy against inflammation or to the classical therapeutic regimen. The aim of the study is to understand if the new therapy is more effective than the old one in reducing the inflammation, measured as the level of C-reactive protein, a marker of blood vessels inflammation. The effect of treatment may be mediated by the levels of cholesterol, which, as the outcome, is measured repeatedly over the study follow-up period, for example every three months.

In the multilevel framework, the mediator and the outcome models can be written as:

$$M_{it} = (\beta_0 + b_{0i}) + \beta_X X_i + \varepsilon_{Mit} \tag{3.45}$$

$$Y_{it} = (\gamma_0 + g_{0i}) + \gamma_X X_i + (\gamma_M + g_{Mi}) M_{it} + \varepsilon_{Yit} \tag{3.46}$$

or, equivalently, as

$$M_{it} = \pi_{0Mi} + \varepsilon_{Mit} \tag{3.47}$$

$$\pi_{0Mi} = \beta_0 + \beta_X X_i + b_{0i} \tag{3.48}$$

$$Y_{it} = \pi_{0Yi} + \pi_{YMi} M_{it} + \varepsilon_{Yit} \tag{3.49}$$

$$\pi_{0Yi} = \gamma_0 + \gamma_X X_i + g_{0i} \tag{3.50}$$

$$\pi_{YMi} = \gamma_M + g_{Mi}. \tag{3.51}$$

The latter formulation shows that the exposure can be seen as a predictor of the latent intercepts π_{0Mi} and π_{0Yi}. It is then very easy to translate the model in SEM terms

$$\underbrace{\begin{pmatrix} X_i \\ M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Mi} \\ \pi_{0Yi} \\ \pi_{YMi} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} \mu_X \\ 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \beta_0 \\ \gamma_0 \\ \gamma_M \end{pmatrix}}_{\boldsymbol{\mu}_i} + \underbrace{\left(\begin{array}{cccccccccc} 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & M_{i1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & M_{iT_i} \\ \hline \beta_X & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ \gamma_X & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}_i} \underbrace{\begin{pmatrix} X_i \\ M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Mi} \\ \pi_{0Yi} \\ \pi_{YMi} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \varepsilon_{Xi} \\ \varepsilon_{Mi1} \\ \vdots \\ \varepsilon_{MiT_i} \\ \varepsilon_{Yi1} \\ \vdots \\ \varepsilon_{YiT_i} \\ b_{0i} \\ g_{0i} \\ g_{Mi} \end{pmatrix}}_{\boldsymbol{\zeta}_i}.$$

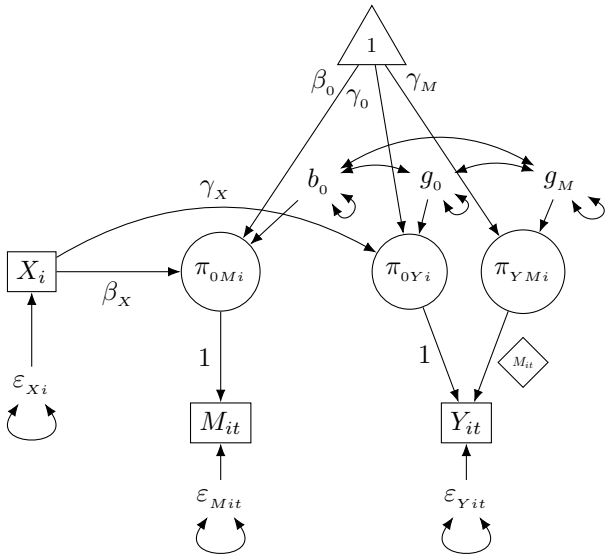

Figure 3.6: 2-1-1 design expressed through definition variables.

Note that in this case $\boldsymbol{\eta}_i$ contains the exposure, which affects both latent factor intercepts and that the only definition variable in $\boldsymbol{\Gamma}_i$ is M_{it}. The subject-specific covariance matrices have basically the same structure discussed for 1-1-1 design: $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon_{Mi}}}$ and $\boldsymbol{\Theta}_{\boldsymbol{\varepsilon_{Yi}}}$ do not vary, while $\boldsymbol{\Psi}_i^B$, for this design, includes also X and can be written as

$$\boldsymbol{\Psi}_{\boldsymbol{i}}^{\boldsymbol{B}} = \begin{pmatrix} \theta^2_{X} & & & \\ 0 & \psi^2_{b_0} & & \\ 0 & \psi_{b_0 g_0} & \psi^2_{g_0} & \\ 0 & \psi_{b_0 g_M} & \psi_{g_0 g_M} & \psi^2_{g_M} \end{pmatrix},$$

where θ^2_{X} is the variance of X.

Since there are only three latent factors, the expressions of the variances and covariances simplify to

$$\begin{aligned} \mathrm{V}(M_{it}) &= \mathrm{V}(b_{0i}) + \mathrm{V}(\varepsilon_{Mit}) = \psi^2_{b_0} + \theta^2_{M} \\ \mathrm{Cov}(M_{is}, M_{it}) &= \mathrm{V}(b_{0i}) = \psi^2_{b_0} \\ \mathrm{V}(Y_{it}) &= \mathrm{V}(g_{0i}) + M^2_{it}\mathrm{V}(g_{Mi}) + 2M_{it}\mathrm{Cov}(g_{0i}, g_{Mi}) + \mathrm{V}(\varepsilon_{Yit}) \\ &= \psi^2_{g_0} + M^2_{it}\psi^2_{g_M} + 2M_{it}\psi_{g_0 g_M} + \theta^2_{Y} \\ \mathrm{Cov}(Y_{is}, Y_{it}) &= \mathrm{V}(g_{0i}) + M_{is}M_{it}\mathrm{V}(g_{Mi}) + (M_{is} + M_{it})\mathrm{Cov}(g_{0i}, g_{Mi}) \\ &= \psi^2_{g_0} + M_{is}M_{it}\psi^2_{g_M} + (M_{is} + M_{it})\psi_{g_0 g_M} \end{aligned}$$

The model is represented in Figure 3.6. In this case the effect of π_{YMi} on the outcome is moderated by the time-varying values of M, or, using a more familiar multilevel terminology, the γ_M coefficient represents the within effect of the mediator

on the outcome. As a consequence, the indirect effect estimated simply as the product $\beta_X \gamma_M$ represents a within indirect effect. This contrasts what stated by Preacher et al. (2010), according to which in 2-1-1 designs the only indirect effect is at the cluster level. Pituch and Stapleton (2012) argue that in such a design the exposure can affect the level-1 outcome either through the individual level mediator and through a level-2 mediator construct. These two indirect effects have a different meaning since the mediator cluster construct may carry information on a different aspect of the phenomenon under investigation. Both a within and a contextual effect (i.e. the effect of a level-2 variable on a level-1 variable) are then plausible and, in fact, when no contextual effect is present, the estimation of only a cluster level indirect effect, as suggested by Preacher et al. (2010), can produce bias. The inclusion of a cluster level effect would imply the inclusion of a link between π_{0Mi} and π_{0Yi} in Figure 3.6, so that Equation (3.50) becomes

$$\pi_{0Yi} = \gamma_0 + \gamma_X X_i + \gamma_{\pi_{0M}} \pi_{0Mi} + g_{0i}.$$

The within indirect effect is always $\beta_X \gamma_M$, while the between or cluster level indirect effect is the product $\beta_X \gamma_{\pi_{0M}}$. An extensive discussion on the topic can be found in Pituch and Stapleton (2012).

2-2-1 design

Consider an observational study carried out for ten months on a group of depressed adolescents. Suppose that the outcome of interest is the number of self-destructive behaviours, assessed monthly, the exposure is parental support, measured at the beginning of the study and the mediator is sense of loneliness at baseline. This is an example of 2-2-1 design, where the exposure and the mediator are measured at the same level, then the mediator model is simply

$$M_i = \beta_0 + \beta_X X_i + \varepsilon_{Mi},$$

while the response is measured at level one, then

$$Y_{it} = (\gamma_0 + g_{0i}) + \gamma_X X_i + \gamma_M M_i + \varepsilon_{Yit}$$

or, alternatively,

$$\begin{aligned} Y_{it} &= \pi_{0Yi} + \varepsilon_{Yit} \\ \pi_{0Yi} &= \gamma_0 + \gamma_X X_i + \gamma_M M_i + g_{0i} \end{aligned}$$

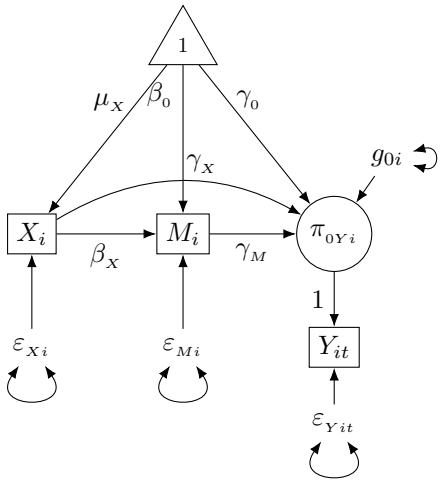

Figure 3.7: 2-2-1 design expressed in the MSEM framework.

from which it can be seen that both X_i and M_i affect the only latent factor in the model, i.e. the outcome random intercept.

The model can be expressed as

$$\underbrace{\begin{pmatrix} X_i \\ M_i \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Yi} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} \mu_X \\ \beta_0 \\ 0 \\ \vdots \\ 0 \\ \gamma_0 \end{pmatrix}}_{\boldsymbol{\mu}_i} + \underbrace{\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ \beta_X & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ \gamma_X & \gamma_M & 0 & \cdots & 0 & 0 \end{pmatrix}}_{\boldsymbol{\Gamma}_i} \underbrace{\begin{pmatrix} X_i \\ M_i \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Yi} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \varepsilon_{Xi} \\ \varepsilon_{Mi} \\ \varepsilon_{Yi1} \\ \vdots \\ \varepsilon_{YiT_i} \\ g_{0i} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where definition variables are not present. In addition, $\boldsymbol{\Psi}_i^W$ reduces to $\boldsymbol{\Theta}_{Yi} = \mathbf{I}_{T_i} \otimes \theta_Y^2$ and $\boldsymbol{\Psi}^B = \begin{pmatrix} \theta_X^2 & & \\ 0 & \theta_M^2 & \\ 0 & 0 & \psi_{g0}^2 \end{pmatrix}$. The outcome variance is simply $\mathrm{V}(Y_{it}) = \mathrm{V}(g_{0i}) + \mathrm{V}(\varepsilon_{Yit}) = \psi_{g0}^2 + \theta_Y^2$, while the covariance between any two time occasions $s, t = 1, \ldots, T_i$ reduces to $\mathrm{Cov}(Y_{is}, Y_{it}) = \mathrm{V}(g_0) = \psi_{g0}^2$.

The model is represented in Figure 3.7, from which it can also be derived the expression for the indirect effect, again simply the product $\beta_X \gamma_M$.

1-2-1 design

The designs described in this and in the next section, although plausible in a longitudinal setting, have never been addressed in the standard multilevel framework, since it is impossible to overcome the dimension mismatch between the exposure and the

mediator, in the case of 1-2-1 design, and between the outcome and the exposure and the mediator in the 1-1-2 design. This issue can be solved in the MSEM framework, as we will show in the following.

In the 1-2-1 design, the exposure and the outcome are repeatedly measured over time and the mediator is time-invariant. For example, consider an experimental study on a group of overweight subjects. They are randomly assigned to two different dietary regimens, the treatment group to an initial 2000-calories-per-day regime, subsequently reduced to 1850, 1600 and 1450 calories, with a balanced intake of proteins, fats and carbs; the control group is a 1800-calories-per-day regime, subsequently reduced to 1700, 1600 and 1500, where subjects are free to decide what to eat provided that they do not exceed the prescribed caloric intake. The aim of the study is to understand which dietary regimen is more effective for weight loss, the outcome of interest, measured bimonthly as the difference between the current weight and the pre-treatment weight. The effect of the treatment may be mediated by the difference between the patient's glycemic index assessed at half of follow-up period and the pre-intervention level. The mediator is then assessed only one time and it should be a proxy of a metabolic change due to diet.

The outcome model can be easily written as

$$Y_{it} = (\gamma_0 + g_{0i}) + (\gamma_X + g_{Xi})X_{it} + \gamma_M M_i + \varepsilon_{Yit} \tag{3.52}$$

or equivalently as

$$Y_{it} = \pi_{0Yi} + \pi_{YXi}X_{it} + \varepsilon_{Yit} \tag{3.53}$$

$$\pi_{0Yi} = \gamma_0 + \gamma_M M_i + g_{0i} \tag{3.54}$$

$$\pi_{YXi} = \gamma_X + g_{Xi}. \tag{3.55}$$

The mediator model is less immediate to write, since it involves a level-2 variable of size n affected by a level-1 variable of size $\sum_{i=1}^{n} T_i$. It can then be assumed that M is influenced only by the between component of the exposure (the η_{Xj} factor in Preacher et al. (2010) notation), which, for notational consistency with the other equations, will be denoted by π_{0Xi}. The mediator model equations are then

$$M_i = \beta_0 + \beta_{\pi_{0X}}\pi_{0Xi} + \varepsilon_{Mi} \tag{3.56}$$

$$\pi_{0Xi} = \alpha_0 + a_{0i}, \tag{3.57}$$

and the mediation model can be written as

$$\underbrace{\begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT_i} \\ M_i \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Xi} \\ \pi_{0Yi} \\ \pi_{YXi} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \beta_0 \\ 0 \\ \vdots \\ 0 \\ \alpha_0 \\ \gamma_0 \\ \gamma_X \end{pmatrix}}_{\boldsymbol{\mu}_i} + \underbrace{\left(\begin{array}{ccccccc|c|cc} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \beta_{\pi_{0X}} & 0 & 0 \\ \hline 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & X_{i1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & X_{iT_i} \\ \hline 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & \gamma_M & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}_i} \underbrace{\begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT_i} \\ M_i \\ Y_{i1} \\ \vdots \\ Y_{iT_i} \\ \pi_{0Xi} \\ \pi_{0Yi} \\ \pi_{YXi} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \varepsilon_{Xi1} \\ \vdots \\ \varepsilon_{XiT_i} \\ \varepsilon_{Mi} \\ \varepsilon_{Yi1} \\ \vdots \\ \varepsilon_{YiT_i} \\ a_{0i} \\ g_{0i} \\ g_{Xi} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

where the only definition variable is X. The within covariance matrix is the following block matrix

$$\boldsymbol{\Psi}_i^W = \begin{pmatrix} \boldsymbol{\Theta}_{\varepsilon_{Xi}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta}_{\varepsilon_{Yi}} \end{pmatrix},$$

where $\boldsymbol{\Theta}_{\varepsilon_{Xi}}$ and $\boldsymbol{\Theta}_{\varepsilon_{Yi}}$ are diagonal $T_i \times T_i$ matrices with constant diagonal elements, θ^2_{X} and θ^2_{Y} respectively, representing the variance of the exposure and the outcome residuals. $\boldsymbol{\Psi}_{\boldsymbol{i}}^{\boldsymbol{B}}$ is instead

$$\boldsymbol{\Psi}_{\boldsymbol{i}}^{\boldsymbol{B}} = \begin{pmatrix} \theta^2_{M} & & & \\ 0 & \psi^2_{a_0} & & \\ 0 & \psi_{a_0 g_0} & \psi^2_{g_0} & \\ 0 & \psi_{a_0 g_X} & \psi_{g_0 g_X} & \psi^2_{g_X} \end{pmatrix},$$

where θ^2_{M} is the variance of the mediator. The variance and the covariance of Y within the same subject are

$$\begin{aligned} \mathrm{V}(Y_{it}) &= \mathrm{V}(g_{0i}) + X_{it}^2 \mathrm{V}(g_{Xi}) + 2X_{it}\mathrm{Cov}(g_{0i}, g_{Xi}) + \mathrm{V}(\varepsilon_{Yit}) \\ &= \psi^2_{g_0} + X_{it}^2 \psi^2_{g_X} + 2X_{it}\psi_{g_0 g_X} + \theta^2_{Y} \\ \mathrm{Cov}(Y_{is}, Y_{it}) &= \mathrm{V}(g_{0i}) + X_{is}X_{it}\mathrm{V}(g_{Xi}) + (X_{is} + X_{it})\mathrm{Cov}(g_{0i}, g_{Xi}) \\ &= \psi^2_{g_0} + X_{is}X_{it}\psi^2_{g_X} + (X_{is} + X_{it})\psi_{g_0 g_X} \end{aligned}$$

where the equations have the same structure as those derived for design 2-1-1. It is worth remarking that, although X is affected by its between latent component, it does not play any role in the last two formulas: since X is regarded as a definition variable, its values are treated as fixed coefficients, not affected by any random component. The model is represented in Figure 3.8, and the indirect effect is the product $\beta_{\pi_{0X}}\gamma_M$.

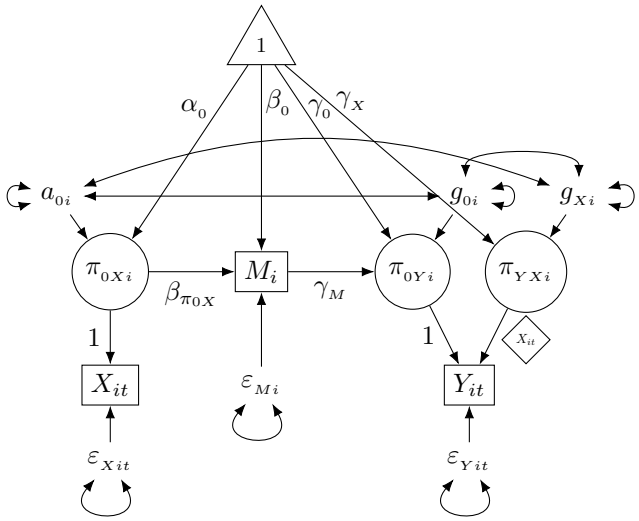

Figure 3.8: 1-2-1 design expressed through definition variables.

1-1-2 design

This design is typical of observational studies where the outcome is measured at the end of follow-up, like, for example, that presented in Lin, Young, Logan, Tchetgen Tchetgen and VanderWeele (2017). They analysed data from the Framingham Heart Study: a cohort of 5,209 participants aged from 30 to 62 years old underwent clinical examination every two years. At each occasion, several variables were assessed, including cardiovascular risk factors, demographics, physical data and many others. The authors focused on five time occasions (i.e. ten years) and their outcome of interest was systolic blood pressure measured at the fifth visit. The exposure is smoking status, measured as the self-reported average number of cigarettes smoked per day, and the mediator is BMI. Both the exposure and the mediator were assessed at each visit, then they are time-varying.

In this design, the mismatch of dimensions discussed in the last section involves both the exposure and the mediator, which are of size $\sum_{i=1}^{n} T_i$, while the outcome has size n. As before, it can be assumed that the between component of X has a direct effect on Y, but what about the mediator? If the mediator model is the same as that in Equation (3.39), the mediator is influenced by two latent factors, an intercept and a slope, so one may ask which of them affects the outcome variable. A plausible solution may be to assume the following model specification

$$M_{it} = \pi_{0Mi} + \pi_{MXi} X_{it} + \varepsilon_{Mit} \tag{3.58}$$

$$\pi_{0Mi} = \beta_0 + \beta_{\pi_{0X}} \pi_{0X} + b_{0i} \tag{3.59}$$

$$\pi_{MXi} = \beta_X + b_{Xi} \tag{3.60}$$

$$Y_i = \gamma_0 + \gamma_{\pi_{0X}} \pi_{0Xi} + \gamma_{\pi_{0M}} \pi_{0Mi} + \varepsilon_{Yi}, \tag{3.61}$$

where the response variable depends on the latent intercept of the mediator. A graphical representation of the model is given in Figure 3.9 and, in the unified notation, it can be written as

$$\underbrace{\begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT_i} \\ M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_i \\ \pi_{0Xi} \\ \pi_{0Mi} \\ \pi_{MXi} \end{pmatrix}}_{\boldsymbol{\eta}_i} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \\ \gamma_0 \\ \alpha_0 \\ \beta_0 \\ \beta_X \end{pmatrix}}_{\boldsymbol{\mu}_i} + \underbrace{\left(\begin{array}{ccccccc|c|cc} 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & X_{i1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 1 & X_{iT_i} \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \gamma_{\pi_{0X}} & \gamma_{\pi_{0M}} & 0 \\ \hline 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \beta_{\pi_{0X}} & 0 & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \end{array}\right)}_{\boldsymbol{\Gamma}_i} \underbrace{\begin{pmatrix} X_{i1} \\ \vdots \\ X_{iT_i} \\ M_{i1} \\ \vdots \\ M_{iT_i} \\ Y_i \\ \pi_{0Xi} \\ \pi_{0Mi} \\ \pi_{MXi} \end{pmatrix}}_{\boldsymbol{\eta}_i} + \underbrace{\begin{pmatrix} \varepsilon_{Xi1} \\ \vdots \\ \varepsilon_{XiT_i} \\ \varepsilon_{Mi1} \\ \vdots \\ \varepsilon_{MiT_i} \\ \varepsilon_{Yi} \\ a_{0i} \\ b_{0i} \\ b_{Xi} \end{pmatrix}}_{\boldsymbol{\zeta}_i},$$

and again the only definition variable is X. The within covariance matrix $\boldsymbol{\Psi}_i^W$ is a block matrix as that showed for the previous design, but $\boldsymbol{\Theta_{Yi}}$ is replaced by $\boldsymbol{\Theta_{Mi}} = \mathbf{I}_{Ti} \otimes \theta^2_{M}$ and the between covariance matrix is

$$\boldsymbol{\Psi_i^B} = \begin{pmatrix} \theta^2_{Y} & & & \\ 0 & \psi^2_{a_0} & & \\ 0 & \psi_{a_0 b_0} & \psi^2_{b_0} & \\ 0 & \psi_{a_0 b_X} & \psi_{b_0 b_X} & \psi^2_{b_X} \end{pmatrix},$$

where θ^2_{Y} is the variance of the outcome. The variance and the covariance of M within the same subject are as in Equations (3.41)-(3.42). The indirect effect is a level-2 effect and can be estimated as the product $\beta_{\pi_{0X}} \gamma_{\pi_{0M}}$.

3.4.3 Other multilevel mediation designs

The designs discussed so far are the most frequently encountered in longitudinal settings. Nonetheless, there are other possible multilevel designs which can be addressed in the MSEM framework. They will be discussed in the following and, since they are quite unusual for longitudinal data, we will use the standard notation adopted in the multilevel literature, denoting by i level-1 units and by j level-2 units. The number of groups is J and n_j is the number of individuals in each group.

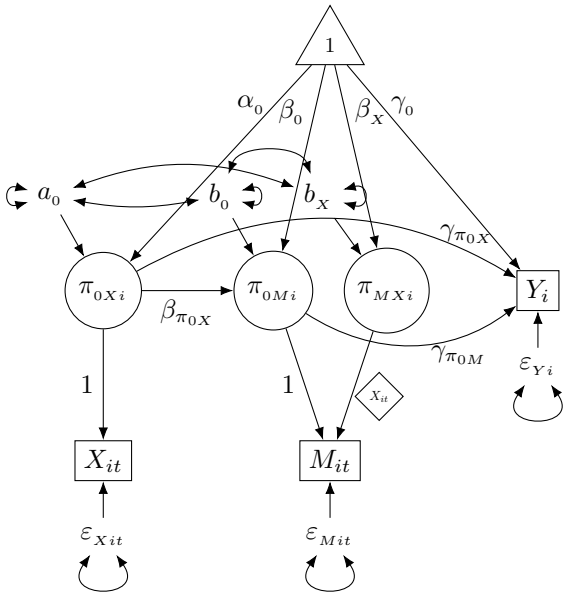

Figure 3.9: 1-1-2 design expressed through definition variables.

2-1-2 design

Consider a study involving different schools in a region or a county, and let X be the amount of state resources assigned to each school for extra activities, M the self-reported motivation of teachers and Y an aggregated score measuring the proficiency of students at the school level. It is plausible that a higher amount of funds may increase the teachers' motivation level and this can have a positive effects on students' learning outcome, measured at the school level.

In this design the exposure and the outcome are measured at the cluster level, while the mediator at the subject level, as a consequence it is impossible to use definition variables. As already explained, to overcome the problem due to the different dimensions of the mediator and the outcome, it is necessary to consider the between component of the mediator. The models are then

$$M_{ij} = \pi_{0Mj} + \varepsilon_{Mij} \tag{3.62}$$

$$\pi_{0Mj} = \beta_0 + \beta_X X_j + b_{0j} \tag{3.63}$$

$$Y_j = \gamma_0 + \gamma_X X_j + \gamma_{\pi 0M} \pi_{0Mj} + \varepsilon_{Yj}, \tag{3.64}$$

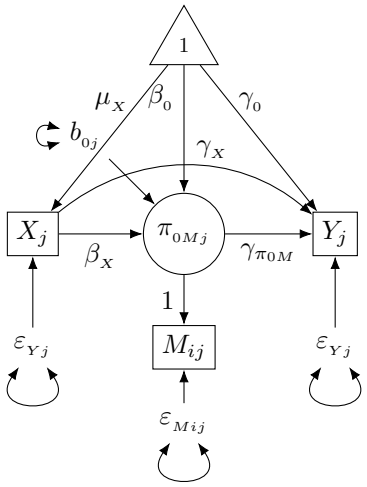

Figure 3.10: 2-1-2 design expressed in the MSEM framework.

which in matrix form can be written as

$$\underbrace{\begin{pmatrix} X_j \\ M_{1j} \\ \vdots \\ M_{n_j j} \\ Y_j \\ \pi_{0Mj} \end{pmatrix}}_{\boldsymbol{\eta}_j} = \underbrace{\begin{pmatrix} \mu_X \\ 0 \\ \vdots \\ 0 \\ \gamma_0 \\ \beta_0 \end{pmatrix}}_{\boldsymbol{\mu}_j} + \underbrace{\begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 1 \\ \gamma_X & 0 & \cdots & 0 & 0 & \gamma_{\pi_{0M}} \\ \beta_X & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{\Gamma}_j} \underbrace{\begin{pmatrix} X_j \\ M_{1j} \\ \vdots \\ M_{n_j j} \\ Y_j \\ \pi_{0Mj} \end{pmatrix}}_{\boldsymbol{\eta}_j} + \underbrace{\begin{pmatrix} \varepsilon_{Xj} \\ \varepsilon_{M1j} \\ \vdots \\ \varepsilon_{Mn_j j} \\ \varepsilon_{Yj} \\ b_{0j} \end{pmatrix}}_{\boldsymbol{\zeta}_j}.$$

The cluster-specific covariance matrix is always a block matrix where the within-block $\boldsymbol{\Psi}_j^W$ equals $\boldsymbol{\Theta}_{Mj} = \mathbf{I}_{n_j} \otimes \theta_M^2$ and the between covariance matrix is a diagonal matrix containing the variances of the level-2 variables

$$\boldsymbol{\Psi}_j^B = \begin{pmatrix} \theta_X^2 & & \\ 0 & \theta_Y^2 & \\ 0 & 0 & \psi_{b_0}^2 \end{pmatrix}.$$

The model is represented in Figure 3.10 , from which it is very easy to understand that the indirect effect can be estimated as the product $\beta_X \gamma_{\pi_{0M}}$.

1-2-2 design

Finally, in the last design only the exposure is measured at level 1, and again no definition variables are needed to express this model in the MSEM framework. As an example, consider a firm with several departments: the individual-level satisfaction of employees may affect the department social climate, which in turn can have an

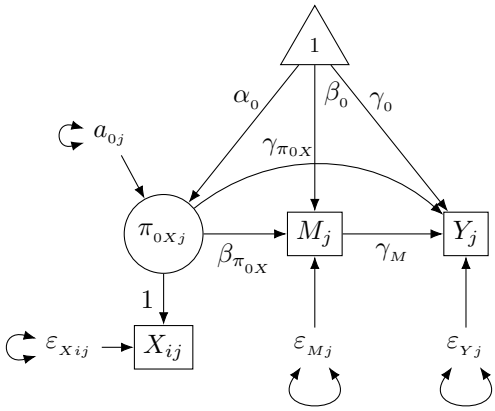

Figure 3.11: 1-2-2 design expressed in the MSEM framework.

effect on the department productivity.

The model equations are simply

$$M_j = \beta_0 + \beta_{\pi_{0X}}\pi_{0Xj} + \varepsilon_{Mj} \tag{3.65}$$

$$Y_j = \gamma_0 + \gamma_{\pi_{0X}}\pi_{0Xj} + \gamma_M M_j + \varepsilon_{Yj} \tag{3.66}$$

$$\pi_{0Xj} = \alpha_0 + a_{0j} \tag{3.67}$$

which can be rewritten as follows

$$\underbrace{\begin{pmatrix} X_{1j} \\ \vdots \\ X_{n_jj} \\ M_j \\ Y_j \\ \pi_{0Xj} \end{pmatrix}}_{\boldsymbol{\eta}_j} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \beta_0 \\ \gamma_0 \\ \alpha_0 \end{pmatrix}}_{\boldsymbol{\mu}_j} + \underbrace{\begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & 1 \\ \vdots & & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & 0 & \beta_{\pi_{0X}} \\ 0 & \cdots & 0 & \gamma_M & 0 & \gamma_{\pi_{0X}} \\ 0 & \cdots & 0 & 0 & 0 & 0 \end{pmatrix}}_{\boldsymbol{\Gamma}_j} \underbrace{\begin{pmatrix} X_{1j} \\ \vdots \\ X_{n_jj} \\ M_j \\ Y_j \\ \pi_{0Xj} \end{pmatrix}}_{\boldsymbol{\eta}_j} + \underbrace{\begin{pmatrix} \varepsilon_{X1j} \\ \vdots \\ \varepsilon_{Xn_jj} \\ \varepsilon_{Mj} \\ \varepsilon_{Yj} \\ a_{0j} \end{pmatrix}}_{\boldsymbol{\zeta}_j}.$$

with covariance matrix having the same structure as design 2-1-2 where $\boldsymbol{\Psi}_j^W = \boldsymbol{\Theta}_{Xj} = \mathbf{I}_{n_j} \otimes \theta_X^2$ and the between covariance matrix is a diagonal matrix containing the variances of the level-2 variables

$$\boldsymbol{\Psi}_j^B = \begin{pmatrix} \theta_M^2 & & \\ 0 & \theta_Y^2 & \\ 0 & 0 & \psi_{a0}^2 \end{pmatrix}.$$

The model is represented in Figure 3.11, where the indirect effect is again $\beta_{\pi_{0X}}\gamma_M$.

3.5 Inference and software implementation

In the last sections, we showed that the definition variable approach entails the use of cluster-specific factor loading and covariance matrices. This implies that traditional maximum-likelihood estimation methods are not appropriate, since they rely on the minimisation of the discrepancy between the sample covariance matrix and the model-implied covariance matrix. For unbalanced designs and multilevel models with continuous predictors, there is not a unique covariance matrix. In order to exploit all the information present in the data, a widely used approach is that based on *full information maximum likelihood* (FIML), which, for each row in the data set, filters out the missing values and uses only the observed data. Let us consider a scenario as that described in the previous sections, where n subjects were randomly sampled from a population of interest and for each of them a set of variables were measured repeatedly over time. Some subjects may have not been assessed at every time occasion, so that each subject i presents T_i measurements and p_i observed variables.

The individual likelihood is given by

$$-2\log\mathcal{L}_i = p_i\log(2\pi) + \log|\boldsymbol{\Sigma}_i| + (\boldsymbol{R}_i - \boldsymbol{\mu}_i^o)\boldsymbol{\Sigma}_i^{-1}(\boldsymbol{R}_i - \boldsymbol{\mu}_i^o)^T$$

where $\boldsymbol{\Sigma}_i$ is the $p_i \times p_i$ filtered model-implied covariance matrix for observed variables, $\boldsymbol{\mu}_i^o$ is the filtered model-implied mean vector for observed variables (Equation 3.17), and $\boldsymbol{R}_i$ is the row corresponding to subject i in the data set. The log-likelihood for the entire data set is obtained by summing the individual contributions

$$F_{ML} = -2\log\mathcal{L} = -2\sum_{i=1}^{n}\log\mathcal{L}_i.$$

The parameters of interest, i.e. $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$ can be estimated by solving $\partial F_{ML}/\partial\boldsymbol{\mu} = 0$, $\partial F_{ML}/\partial\boldsymbol{\Gamma} = 0$ and $\partial F_{ML}/\partial\boldsymbol{\Psi} = 0$. Generally speaking, the solutions of these equations cannot be expressed in closed form, then iterative algorithms, such as Newton-Raphson or EM, are required.

This approach can find application not only in the MSEM framework with definition variables, but it can also be used for the estimation of the other three models discussed. For CLPM and LDS models, FIML is useful in the presence of missing data, which make the design unbalanced. As regards LGMs, this estimation method is required not only in the case of balanced design (at the sampling level) with missing data, but also when the measurements occasions differ among subjects, a setting analysed by Mehta and West (2000). In this latter case, the λ coefficients coding time are no more common to all subjects and, as a consequence, it is not possible

to use a unique Γ matrix for parameterising the model. Then, the definition variable approach allows to fit a different factor loading matrix for each subject, where the $Time$ column contains the times corresponding to the individual measurement occasions (Mehta and West 2000).

Once the model has been fitted and the parameters have been estimated, it would be of considerable interest to measure the model goodness-of-fit and to have a criterion for model selection. For traditional, single-level SEMs there are several alternatives, such as the likelihood ratio test (LRT), the comparative fit index (CFI), the root mean square error of approximation (RMSEA) and the AIC and BIC for comparing non-nested models. These measures cannot be applied straightforwardly to a multilevel setting, since ML estimation in SEM requires the independence of statistical units, a condition which in multilevel models is clearly not satisfied. Yuan and Bentler (2007) point out three limitations of the standard approach when applied to multilevel models: first, the standard approach may not be sensitive to the goodness (or lack) of fit at the group level, second, when the standard approach indicates lack of fit it is not clear at what level it is detected, and finally, model specification at one level can negatively affect the estimates at the other level.

To overcome these issues, both Yuan and Bentler (2007) and Ryu and West (2009) propose to evaluate the model fit at each level. The former approach relies on two steps, first they estimate a saturated covariance matrix for each level, second this is used as input for a covariance structure analysis at that level. The latter approach instead uses partially saturated models to assess level-specific model fit indexes. In particular, denoting by $\mathbf{\Sigma}_k$, $k \in \{B, W\}$ the implied covariance matrix at the k-th level and by $\mathbf{\Sigma}_k^S$, $k \in \{B, W\}$ the covariance matrix corresponding to the saturated model at k-th level, to assess the within (level 1) model fit $F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W]$ is compared to $F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W^S]$ (i.e. the saturated model). Analogously, to assess the between-level goodness of fit $F_{ML}[\mathbf{\Sigma}_B, \mathbf{\Sigma}_W^S]$ is compared to $F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W^S]$.

Assuming a balanced design where each group contains n individuals and the number of observed variables is p, the LRT and the RMSEA can be written as

$$\chi^2_W = F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W] - F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W^S] \qquad RMSEA_W = \sqrt{\frac{\chi^2_W - df_W}{df_W \times n}} \tag{3.68}$$

where $df_W = np_W^S - np_W$ and n is the total sample size. Analogously, for the between level,

$$\chi^2_B = F_{ML}[\mathbf{\Sigma}_B, \mathbf{\Sigma}_W^S] - F_{ML}[\mathbf{\Sigma}_B^S, \mathbf{\Sigma}_W^S] \qquad RMSEA_B = \sqrt{\frac{\chi^2_B - df_B}{df_B \times J}}, \tag{3.69}$$

where $df_B = d_B^S - d_B$ (d is the number of parameters). Rappaport et al. (2020) proposes also a level-specific AIC, defined as

$$\text{AIC}_k = \chi_k^2 + 2d_k, \quad k \in \{B, W\}, \tag{3.70}$$

and implemented an algorithm in the `OpenMx R` package to estimate these indexes. They carried out a simulation study for evaluating its abilities to detect level-specific misspecification.

It is worth mentioning another fit index proposed by Lee and Song (2001). They propose a Bayes factor to compare two non-nested models M1 and M2. Let D denote the data, from the Bayes theorem the posterior probability of each model Mk, k $= 1, 2$, is

$$P(\text{Mk}|D) = \frac{P(D|\text{Mk})P(\text{Mk})}{P(D|\text{M1})P(\text{M1}) + P(D|\text{M2})P(\text{M2})}, \qquad \text{k} = 1, 2$$

where $P(\text{Mk})$ is the prior probability for model k. It follows that

$$\frac{P(\text{M1}|D)}{P(\text{M2}|D)} = \frac{P(D|\text{M1})P(\text{M1})}{P(D|\text{M2})P(\text{M2})}$$

where the first factor is known as the Bayes factor and can be denoted by B_{12}.

Obtaining B_{12} analytically would generally require the integration of $P(D|\text{Mk}), \text{k} = 1, 2$ over the parameter space, a task often very difficult. For this reason, Lee and Song (2001) propose an approximation capitalizing on the decomposition of total variance in its between and within components typical of MSEM, but we do not report the details here.

As regards the implementation of these methods, although there exist many SEM specialised software and FIML is now implemented in most of them, the same does not hold true for definition variables, which, to the best of our knowledge, are implemented only in Mplus and Mx (and its `R` version `OpenMx`). The former makes use of computationally intensive numerical integration procedures, the latter employs constrained optimisation methods based on Sequential Quadratic Programming (SQP).

The main advantage of `OpenMx` over Mplus is that it is a freely downloadable package within the open source `R` software, while Mplus has a cost of about 700$. In addition, `OpenMx` has a very user-friendly syntax which allows the user to choose between a path-specification approach, where the user has to write all paths present in the model, both directed and bidirected to represent variances/covariances, and a matrix-specification approach, where the user specifies the three matrices characterising the RAM approach ($\mathbf{\Gamma}, \mathbf{\Psi}$ and $\mathbf{F}$) plus the mean vector. The latter approach seems more appropriate when the number of variables is small. For single-level data,

the dataset is written in the wide format, i.e. a subject for each row and a variable for each column, while for multilevel data it is easier to use the long format, i.e. a level-1 unit for each row and a variable used to associate each unit to its cluster. In the multilevel case two models are fitted, one for level-2 variables and the other for level-1 variables. Preacher (2011) states that multilevel models fitted as SEMs through the definition variable approach can be estimated using a single-level data structure, i.e. a wide format. However, as the number of level-1 or level-2 units increases, the data managing problem becomes remarkable, since a wide format requires to have a column for each observed variable and in real-world cases the data table is expected to have hundreds of columns. For this reason, we suggest to use the multilevel approach based on a long data format and fit a model for each level.

3.6 Discussion

In this chapter we showed how SEMs and mixed-effect models can be integrated in a unique framework through the use of the RAM notation and the possibility to include definition variables. Although other unifying approaches have been proposed in the past (Skrondal and Rabe-Hesketh 2004, Niku et al. 2017, Usami et al. 2019, Heck and Thomas 2020), so far, to the best of our knowledge, none has specifically addressed mediation settings. Our approach is flexible enough to encompass several kinds of SEM, not only the three on which we focused, and the possibility to include definition variables makes unbalanced designs easy to handle.

Section 3.4 was entirely devoted to multilevel models via definition variables, an approach almost unexplored so far, since multilevel mediation has traditionally been restricted to cases where an upper level variable influence variables only at the same or at a lower level (Krull and MacKinnon 2001, Pituch and Stapleton 2012, Bauer et al. 2006) or has been addressed within the classical MSEM framework involving the decomposition of variables in their between and within components (Preacher et al. 2010, 2011). Preacher et al. claim the superiority of their MSEM approach over the traditional multilevel ones, since it can address any kind of mediation design, including those involving bottom-up effects (i.e. designs involving 1-2 components, in the standard notation to define designs), and it is able to decompose indirect effects in their within and between components, while traditional methods provide only *conflated estimates*, using their terminology.

However, as pointed out by Pituch and Stapleton (2012), the term conflated may generate confusion, since

> *casual readers may get the impression that the use of the statistical models*

> *associated with this approach* (the standard approach) *is always wrong or somehow harmful*

while, in fact, in their simulation study, Pituch and Stapleton prove that Preacher et al. (2010) approach may be biased. In addition, for scholars more familiar with the multilevel framework, the mediation setting proposed by Preacher et al. may appear very different from it, while the definition variable approach we introduced is closer in spirit to multilevel models and tries to follow the same structure from a different perspective.

We do not claim that our approach is superior to that of Preacher et al., suggesting instead that they are simply alternative methods to address the same issue. Both can deal with any kind of multilevel mediation setting, but the differences between the estimates of mediational effects in the two approaches are yet to be investigated. Researchers are then encouraged to try both types of modeling approach and compare them using appropriate indexes, like AIC or BIC.

Clearly, the proposed approach is not free of drawbacks. We made the traditional assumption of multivariate Normal distribution for all variables, both observed and latent, but in many settings this assumption is quite unrealistic. FIML proved to be robust to non-Normality for continuous variables, but the presence of discrete outcomes (e.g. Binomial or Poisson distributed) requires to fit nonlinear relationships (use link functions different from identity) which standard SEM software do not allow.

Extensions of SEMs to generalized linear models have been proposed by Skrondal and Rabe-Hesketh (2004) and Niku et al. (2017, 2019), but, at the moment, they seem not to be well known among scholars. As regards the Skrondal and Rabe-Hesketh (2004) approach, this may be due to the challenging theoretical framework and the difficult notation adopted by the authors, which may discourage researchers from studying and using GLLAMM. In addition, they implemented their approach only in Stata, which of course prevents the approach from spreading among non-Stata users. As regards the approach by Niku et al. (2017, 2019), it is relatively new and has been implemented in `R` only recently, so it has yet to be known among applied researchers. Moreover, although it allows to fit GLMs, the types of admitted latent variables are limited and latent factors cannot be function of other predictors.

Integrating our approach with GLM theory and investigating the associated inferential issues are tasks for future work.

Chapter 4

Causal multilevel and latent growth models: a separable effects approach

The focus of the last chapter was on associational models for longitudinal mediation analysis: we showed how they can be embedded in a unified framework and the advantages it presents. In contrast, this chapter addresses longitudinal mediation from a causal perspective. As discussed in Chapter 2, the causal literature has generally focused on longitudinal mediation models with survival outcomes, which are quite common in epidemiology and play a relevant role from a causal perspective. In Chapter 2, we also highlighted that mediation models including latent variables are not very common in the causal inference literature, and when they are taken into account it is generally to address unmeasured confounding.

In this chapter, we want to focus on outcomes other than survival by using models which are widely employed for longitudinal data and address latent variables as structural components. In particular, we propose a causal interpretation of generalised mixed-effect models and latent growth models in terms of separable effects. This enables us to provide explicit assumptions for endowing mediational effects with such an interpretation. Furthermore, these assumptions can be easily interpreted from an interventional point of view (Robins and Richardson 2011, Robins et al. 2020). We also derive formulas for these effects and discuss their differences.

The remainder of the chapter is organised as follows. Section 1 provides a short review of articles using mixed-effect or latent growth models in a causal framework. In Section 2 we introduce the separable effects approach and discuss the assumptions on which it relies. In Sections 3 and 4 we discuss mixed-effect models and latent growth models, respectively. We show how to derive the separable mediational effects in

both cases, which assumptions are needed, and compare the two models. Section 5 is devoted to a simulation study, where we want to assess how sensitive the estimates obtained through the g-formula are to misspecification. In Section 6 we discuss some possible extensions of the proposed approach and a discussion follows.

4.1 Background

In this section, we provide a brief overview on works analysing mixed-effect or latent growth models from a causal perspective, not necessarily involving mediation.

Theoretical foundations of causal multilevel models, not restricted to longitudinal settings, are discussed in Feller and Gelman (2015), Hill (2013), and Gitelman (2005), Raudenbush and Schwartz (2020), the latter focusing on educational contexts. To the best of our knowledge, there are not many examples of causal mixed models for longitudinal data. Hong and Raudenbush (2006) and Graham et al. (2014) use a mixed model propensity score to estimate the effect of being retained in kindergarten versus being promoted to the first grade on the academic learning of retainees, and the effect of road network capacity expansions on traffic volume and density, respectively. Another example of application of propensity scores is provided by Eckardt (2012), who investigates the causal effect of healthy eating habits on students' BMI within supportive school environments (SSEs) and non-SSEs. A very recent review on the different weighting strategies for propensity scores in multilevel settings can be found in Fuentes et al. (2021).

Shardell and Ferrucci (2018) address the issue of estimating the causal effect of a time-varying exposure on a time-varying outcome with observed baseline confounders. Their approach combines parametric joint mixed-effect models and g-computation.

As regards mediation more specifically, we have already introduced the work by Bind et al. (2016), who consider a mediation setting including time-varying exposure, mediators and outcome and propose a generalised mixed-effect modeling approach. They assume that the exposure at time $t-1$ has an autoregressive effect on itself at time t and an effect on the mediators and the outcome at the same time, while adjacent measurement of the mediators and the outcome are linked only through their common random effects.

Similar models, including both random effects and autoregressive components, have already been proposed in the literature. Schuurman et al. (2016) are interested in modeling how two variables affect each other over time, assuming a Granger-causal perspective. They develop a multilevel autoregressive model where the score of a variable for each subject at each time measurement is the sum of an individual mean,

stable over time, and an individual time-varying deviation from the mean, which may depend both on its own previous deviations and previous deviations of the other variable under analysis. All parameters in the model are thought of as drawn from a multivariate distribution with a fixed mean vector and a non-diagonal covariance matrix.

As regards LGMs, there are not many examples of their application within a causal framework. Pakpahan et al. (2017) compare four different structural equation models, among which LGM, and interpret them in a causal way, although no counterfactuals are introduced. In contrast, Jo et al. (2009) make explicit use of the counterfactual framework. They propose a two-step procedure for identifying the causal effect of a treatment on the response when growth trajectories include latent classes. Focusing on mediation, Cheong and MacKinnon (2012) discuss the decomposability of the total effect into direct and indirect effects using Bollen's theory and show how to define mediational effects and compute standard errors and confidence intervals. They also briefly address causal inference and some extensions of mediational LGM, as multilevel models. In many papers, authors select LGMs to model the variables trajectories and infer causation analysing data from RCTs, although they do not formally develop a causal theory for LGM, see Wehmeyer et al. (2013), Zhu et al. (2021). Finally, Tofighi et al. (2019) provide assumptions to interpret LGMs causally and they propose a technique for performing sensitivity analyses in latent growth mediation models.

Finally, it is relevant to mention the work by De Stavola et al. (2014) that, although does not specifically address LGMs or multilevel models, provides interesting insights into the relationship between natural mediational effects and those obtained in an SEM framework following a path-analytic approach. The authors discuss parametric assumptions that make mediational effects identifiable in both settings and prove that the estimands obtained in an SEM framework coincide with natural effects if such assumptions are satisfied. This equivalence holds even in the presence of an intermediate confounder.

4.2 Separable mediational effects

We discussed separable effects in Sections 1.2.5 and 2.3.5. Here we deepen some concepts which will be then applied to mixed-effect models and LGMs.

Recall that the basic idea of separable effects is to extend the model by including two additional variables, X^M and X^Y, which can be thought of as two separate components of the exposure, the former influencing directly only the mediator, the latter only the outcome. This is graphically depicted in Figure 1.14, since only an

arrow emanates from X^M and it goes into M, likewise there is only an arrow from X^Y to Y. Although these variables are not observed, since we observe only the value of X, and in the observational regime it holds that $X \equiv X^Y \equiv X^M$, it is possible to conceive a (future) four-arm trial in which these components are randomised independently. They turn out to be very useful to give insights into the mechanism linking X to Y. First, they allow us to disentangle the pathways through which the exposure effects propagate. Second, the two components of X are intervening variables and, as we shall detail below, they allow us to define mediational effects without devising an intervention on the mediator, as required, instead, by natural effects.

We will consider a setting with a baseline binary exposure X, and a mediator M and a response Y measured over time for T different time occasions, where M is measured before Y. As an example, consider an intervention on a group of overweight subjects to help them lose weight. They are randomised into two groups, the treatment group, where subjects follow lectures on how to eat healthy and participate in meetings to strengthen their motivation and self-esteem for one month, and the control group, where subjects just receive some brochures on healthy eating. Patients are then followed up for six months after the treatment was administered, and at each time occasion their weight and their waist are recorded, as well as some psychological indicators, such as the sense of self-esteem, self-efficacy and motivation.

It is expected that those in the treatment group are more likely to lose weight than those in the control group. If Y_t denotes the difference between the subject's weight at time t and at baseline, and M_t is the sense of self-efficacy at time t, the treatment can favour weight loss directly, but also indirectly through increasing subjects' sense of self-efficacy. Then, it may be plausible to decompose the treatment into two components: a dietary component, which impacts eating habits and then triggers weight loss, and a psychological component, which has a beneficial effect on subjects' self-conception, for example increasing their self-confidence, which in turn positively affects weight loss. In a future randomised trial, each of these two components can be randomised, for example the X^M treatment component could be the participation in psychological meetings to develop self-esteem ($X^M = 1$) or not ($X^M = 0$), and the X^Y component the assignment to attending lectures on healthy eating ($X^Y = 1$) or simply to reading a pamphlet ($X^Y = 0$).

As estimand of interest for each $t = 1, \ldots, T$, we will focus on $\mathbb{E}[Y_t(X^M = x^*, X^Y = x)]$, that is, the expectation of the response at time t under a hypothetical intervention setting the components of X to two different values x and x^*. The mediational effects are indeed defined in terms of quantities of this kind, namely the *longitudinal separable effects*, for each $t = 1, \ldots, T$, can be defined as the differences

(cfr. Equations (1.29)-(1.30))

$$SDE = \mathbb{E}[Y_t(X^M = x^*, X^Y = x)] - \mathbb{E}[Y_t(X^M = x^*, X^Y = x^*)]$$
$$SIE = \mathbb{E}[Y_t(X^M = x, X^Y = x)] - \mathbb{E}[Y_t(X^M = x^*, X^Y = x)].$$

As said, in contrast to the observational regime, we can conceive an *interventional regime*, where X^M and X^Y are different treatments randomised separately. However, since we do not observe X^M and X^Y, our aim is to identify the separable effects only from data on X, M and Y. For the moment, we assume that X is randomised, and this implies that

$$M_t(X) \perp\!\!\!\perp X, \quad Y_t(X) \perp\!\!\!\perp X \quad \text{for each } t = 1, \ldots, T$$

i.e. exchangeability holds. We assume also consistency

$$M_t(x) = M_t, \;\; Y_t(x) = Y_t \;\; \text{when } X = x, \;\; \text{for each } t = 1, \ldots, T$$

and property P1 in Didelez (2019*b*)

$$P(W_t(X^M = x, X^Y = x)) = P(W_t(X = x)) \quad x \in \{0, 1\}, \;\; W \in \{M, Y\}. \tag{4.1}$$

As we pointed out in Section 2.3.5, the separability of the components requires, under the interventional regime, the conditional independence of each variable from the component of which it is not a child, so M should be independent from X^Y and vice versa Y from X^M.

Recalling that, for any variable W, we denote the history of W_t, i.e. the set of variables W_s for $s \leq t$, by $\overline{W}_t$, the separability assumptions in a longitudinal setting correspond to

A1 For each time t, the mediator M_t is independent from the value of X^Y conditional on its observed past, previous values of Y and X^M,

$$M_t \perp\!\!\!\perp X^Y \,|\, \overline{M}_{t-1}, \overline{Y}_{t-1}, X^M$$

A2 For each time t, the response Y_t is independent from the value of X^M conditional on its observed past, previous values of M and X^Y,

$$Y_t \perp\!\!\!\perp X^M \,|\, \overline{M}_t, \overline{Y}_{t-1}, X^Y.$$

If these assumptions are satisfied[1] then, the quantity $\mathbb{E}[Y_t(X^M = x^*, X^Y = x)]$,

[1]This can be checked directly from modified versions of expanded DAGs corresponding to a four-arm trial, where

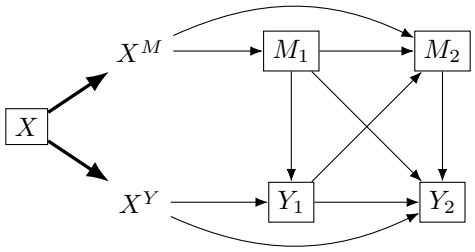

Figure 4.1: Longitudinal separable effect model with two time points.

where the components of X are set to two different values, is non-parametrically identifiable through the mediational g-formula (Didelez 2019*b*)

$$\begin{aligned}\mathbb{E}[Y_t(X^M = x^*, X^Y = x)] = \sum_{\overline{m}_t, \overline{y}_{t-1}} \mathbb{E}[Y_t \,|X = x, \overline{M}_t = \overline{m}_t, \overline{Y}_{t-1} = \overline{y}_{t-1}]\times \\ \prod_{k=1}^{t} P(\overline{M}_k = \overline{m}_k|X = x^*, \overline{M}_{k-1} = \overline{m}_{k-1}, \overline{Y}_{k-1} = \overline{y}_{k-1})\times \\ P(\overline{Y}_{k-1} = \overline{y}_{k-1}|X = x, \overline{M}_{k-1} = \overline{m}_{k-1}, \overline{Y}_{k-2} = \overline{y}_{k-2}).\end{aligned} \tag{4.2}$$

Assume that the causal structure is as in Figure 4.1. At each time, the mediators have a cross-sectional effect on the response, M_1 affects Y_2 and Y_1 affects M_2. Moreover, both the mediator and the outcome have autoregressive effects. It can be noticed that assumptions **A1**-**A2** are satisfied. Then, applying Equation (4.2), $\mathbb{E}[Y_2(X^M = x^*, X^Y = x)]$ results identified by

$$\sum_{m_1, m_2, y_1} \mathbb{E}[Y_2 \,|\, m_1, m_2, y_1, x]P(m_2|m_1, y_1, x^*)P(y_1 \,|\, m_1, x)P(m_1|x^*).$$

In the next sections we apply this approach to mixed effects and latent growth models and derive analytical expressions for the mediational effects.

4.3 Mixed-effect models

By using Laird and Ware (1982) notation, a mixed-effect model can be specified as

$$Y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \mathbf{z}'_{ij}\mathbf{u}_j + \varepsilon_{ij} \tag{4.3}$$

where i and j denote the subject and the cluster, respectively, Y_{ij} is the response variable for subject i in cluster j, $\mathbf{x}_{ij}$ and $\mathbf{z}_{ij}$ are $p \times 1$ and $q \times 1$ vectors of known covariates, $\boldsymbol{\beta}$ and $\mathbf{u}$ are $p \times 1$ and $q \times 1$ vectors of fixed and random coefficients,

X^M and X^Y are regarded as distinguished variables and X is absent.

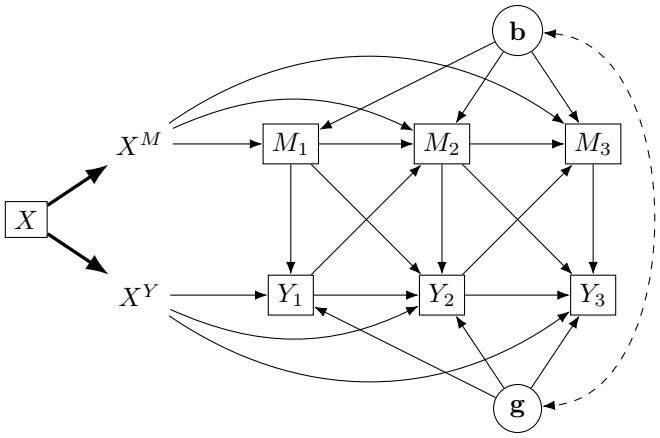

Figure 4.2: Three-wave mixed effect model with separable components of X. $\mathbf{b}$ and $\mathbf{g}$ are random effects.

respectively, with $\mathbb{E}[\mathbf{u}] = 0$ and ε_{ij} an error term with null expectation. The variables in $\mathbf{z}_{ij}$ are generally a subset of those in $\mathbf{x}_{ij}$.

In a longitudinal mediation setting, mixed-effect models can be fitted for both the mediator and the outcome, so that for each subject i

$$M_{it} = \mathbf{x}'_{M_{it}} \boldsymbol{\beta} + \mathbf{z}'_{M_{it}} \mathbf{b}_i + \boldsymbol{\varepsilon}_{M_{it}} \tag{4.4}$$

$$Y_{it} = \mathbf{x}'_{Y_{it}} \boldsymbol{\gamma} + \mathbf{z}'_{Y_{it}} \mathbf{g}_i + \boldsymbol{\varepsilon}_{Y_{it}}. \tag{4.5}$$

In a basic 1-1-1 setting, $\mathbf{x}_{M_{it}}$ includes only the intercept and the exposure and the same for $\mathbf{x}_{Y_{it}}$ with the addition of the mediator at time t. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are fixed effects common to all subjects, while $\mathbf{b}$ and $\mathbf{g}$ are subject-specific random effects.

Among articles dealing with longitudinal mediation using causal mixed-effect models, we have already mentioned that of Bind et al. (2016), who propose assumptions for the identifiability of natural mediational effects. We consider a causal structure similar to theirs, but slightly modified by considering a baseline exposure X and a mediator and a response measured at different time occasions. Unlike Bind et al. (2016), we allow subsequent measures of the mediators and the outcome to be directly linked, not just through the random coefficients in the models. In addition, cross-lagged effects are allowed. Figure 4.2 shows the data structure for three waves.

The graph encodes some dependencies among variables: first, notice that the only children of X^M are the mediators, and the only children of X^Y are the outcome measurements over time. This implies that X^M is independent of the outcome conditional on the mediators, and X^Y is independent of the mediators conditional on the outcomes and previous mediators. Second, notice that the bidirected arrow connecting random effects is dashed, meaning that they can be marginally independent or

correlated. Whether this arrow is present or not has dramatic impacts on the identifiability of separable effects. We will address the case of uncorrelated and correlated random effects in turn.

4.3.1 Uncorrelated random effects

Let us start from the easiest case, where the two sets of random effects are uncorrelated, i.e. there is no dashed link in Figure 4.2. In such a model, assumptions **A1** and **A2** ensure identifiability of the separable mediational effects. To show this, we start by considering the interventional expectation of the outcome under an intervention setting $X^M = x^*$ and $X^Y = x$, x, with $x^* \in \{0, 1\}$, $x \neq x^*$. Applying the law of iterated expectation:

$$
\begin{aligned}
&\mathbb{E}[Y_t(X^M = x^*, X^Y = x)] = \\
&\sum_{\overline{m}_t, \overline{y}_{t-1}} \mathbb{E}[Y_t(X^M = x^*, X^Y = x) \,|\, \overline{M}_t(X^M = x^*, X^Y = x) = \overline{m}_t, \\
&\qquad\qquad \overline{Y}_{t-1}(X^M = x^*, X^Y = x) = \overline{y}_{t-1}] \times \\
&\prod_{k=1}^{t} P(M_k(X^M = x^*, X^Y = x) = m_k \,|\, \overline{M}_{k-1}(X^M = x^*, X^Y = x) = \overline{m}_{k-1}, \qquad (4.6) \\
&\qquad\qquad \overline{Y}_{k-1}(X^M = x^*, X^Y = x) = \overline{y}_{k-1}) \times \\
&P(Y_{k-1}(X^M = x^*, X^Y = x) = y_{k-1} \,|\, \overline{M}_{k-1}(X^M = x^*, X^Y = x) = \overline{m}_{k-1}, \\
&\qquad\qquad \overline{Y}_{k-2}(X^M = x^*, X^Y = x) = \overline{y}_{k-2})
\end{aligned}
$$

with the assumption that variables with zero or negative subscripts are not present.

From assumption **A1** and (4.1) it follows that, for each k,

$$
\begin{aligned}
&P(M_k(X^M = x^*,\, X^Y = x) = m_k \,|\, \overline{M}_{k-1}(X^M = x^*,\, X^Y = x), \\
&\qquad\qquad \overline{Y}_{k-1}(X^M = x^*,\, X^Y = x)) \\
&= P(M_k(X^M = x^*, X^Y = x^*) = m_k \,|\, \overline{M}_{k-1}(X^M = x^*,\, X^Y = x^*), \\
&\qquad\qquad \overline{Y}_{k-1}(X^M = x^*,\, X^Y = x^*)) \\
&= P(M_k(X = x^*) = m_k \,|\, \overline{M}_{k-1}(X = x^*), \overline{Y}_{k-1}(X = x^*))
\end{aligned}
$$

and, since the treatment is randomised, this equals

$$
P(M_k = m_k \,|\, X = x^*, \overline{M}_{k-1}, \overline{Y}_{k-1}).
$$

The same holds for Y_k by using **A2** instead of **A1**, i.e.

$$\begin{aligned} P(Y_k(X^M = x^*,\, X^Y = x) &= y_k \,|\, \overline{M}_k(X^M = x^*, X^Y = x), \overline{Y}_{k-1}(X^M = x^*, X^Y = x)) \\ &= P(Y_k(X^M = x, X^Y = x) = y_k \,|\, \overline{M}_k(X^M = x, X^Y = x), \\ &\qquad\qquad \overline{Y}_{k-1}(X^M = x, X^Y = x)) \\ &= P(Y_k(X = x) = y_k \,|\, \overline{M}_k(X = x), \overline{Y}_{k-1}(X = x)) \\ &= P(Y_k = y_k \,|\, X = x, \overline{M}_k, \overline{Y}_{k-1}). \end{aligned}$$

It then follows that

$$\begin{aligned} \mathbb{E}[Y_t(X^M = x^*, &X^Y = x)] = \\ &\sum_{\overline{m}_t, \overline{y}_{t-1}} \mathbb{E}[Y_t \,|\, X = x,\, \overline{M}_t = \overline{m}_t,\, \overline{Y}_{t-1} = \overline{y}_{t-1}] \times \\ &\prod_{k=1}^{t} P(M_k = m_k \,|\, X = x^*,\, \overline{M}_{k-1} = \overline{m}_{k-1},\, \overline{Y}_{k-1} = \overline{y}_{k-1}) \times \\ &P(Y_{k-1} = y_{k-1} \,|\, X = x,\, \overline{M}_{k-1} = \overline{m}_{k-1},\, \overline{Y}_{k-2} = \overline{y}_{k-2}). \end{aligned} \tag{4.7}$$

A detailed proof is given in Appendix B.

Specifying parametric models for the mediator and the outcome allows us to derive separable direct and indirect effects in terms of regression coefficients. For example, assume a structure as shown in Figure 4.2. For each subject $i = 1, \ldots, n$ and time occasion $t = 1, \ldots, T$, if the mediator and the outcome are assumed to be Normally distributed and their expectations to be linear in the direct causes, possible models for their expectations can be

$$\mathbb{E}[M_{it} \,|\, X_i,\, \overline{M}_{it-1},\, \overline{Y}_{it-1}, \mathbf{b}_i] = (\beta_0 + b_{0i}) + \beta_{X} X_i + \beta_{\ell_1(M)} M_{it-1} + \beta_{\ell_1(Y)} Y_{it-1} \tag{4.8}$$

$$\begin{aligned} &\mathbb{E}[Y_{it} \,|\, X_i,\, \overline{M}_{it},\, \overline{Y}_{it-1}, \mathbf{g}_i] = \\ &(\gamma_0 + g_{0i}) + \gamma_{X} X_i + (\gamma_{M_t} + g_{Mi}) M_{it} + (\gamma_{\ell_1(M)} + g_{\ell_1(M)i}) M_{it-1} + \gamma_{\ell_1(Y)} Y_{it-1}. \end{aligned} \tag{4.9}$$

where the subscripts $\ell_1(M)$ and $\ell_1(Y)$ denote the coefficients referring to M_{t-1} and Y_{t-1} respectively (ℓ_1 stands for the lag operator of order 1), and we are assuming that $\mathbf{u}_i = (b_{0i},\, g_{0i},\, g_{Mi},\, g_{\ell_1(M)i})' \sim \text{MVN}(\mathbf{0}, \mathbf{\Phi})$, with $\mathbf{\Phi}$ diagonal.

Going back to the initial example of the randomised treatment for weight loss, these models imply that the treatment has an effect on both self-efficacy and weight loss over time. Subjects can show heterogeneity in the extent to which the mediator affects the response, so that the effect of M_t and M_{t-1} on weight loss at time t may vary across subjects. The random effects g_{Mi} and $g_{\ell_1(M)i}$ are included in the model to capture such heterogeneity. It is also plausible that the mediator and the

outcome have autoregressive as well as cross-lagged effect, since, for example, a higher weight loss at time t may induce an increased sense of self-efficacy at the subsequent measurement.

To give an example, suppose that one is interested in the separable effects of the exposure on the outcome at time $t = 2$. Considering the difference as contrast and two different values of X, x and x^*, it can easily be proved that, applying the g-formula in (4.7), the separable effects, conditional on random effects, take the form

$$SDE_{|\mathbf{b},\mathbf{g}} = \gamma_X \big[\, 1 + \beta_{\ell_1(M)}(\gamma_M + g_{Mi}) + \gamma_{\ell_1(Y)} \,\big](x - x^*) \tag{4.10}$$

and

$$\begin{aligned} SIE_{|\mathbf{b},\mathbf{g}} = \beta_X \big[\, &(\gamma_M + g_{Mi}) + \beta_{\ell_1(M)}(\gamma_M + g_{Mi}) + (\gamma_M + g_{Mi})\beta_{\ell_1(Y)}(\gamma_M + g_{Mi}) \\ &+ (\gamma_{\ell_1(M)} + g_{\ell_1(M)i}) + (\gamma_M + g_{Mi})\gamma_{\ell_1(Y)} \,\big](x - x^*). \end{aligned} \tag{4.11}$$

Since one is usually interested in the average causal effects, random effects in the formulas above can be integrated out. In addition, since the random effects are assumed to be uncorrelated, the resulting effects are obtained by simply deleting the random coefficients, i.e.

$$SDE = \gamma_X \big[\, 1 + \beta_{\ell_1(M)}\gamma_M + \gamma_{\ell_1(Y)} \,\big](x - x^*) \tag{4.12}$$

$$SIE = \beta_X \big[\, \gamma_M + \beta_{\ell_1(M)}\gamma_M + \beta_{\ell_1(Y)}\gamma_M^2 + \gamma_{\ell_1(M)} + \gamma_M\gamma_{\ell_1(Y)} \,\big](x - x^*). \tag{4.13}$$

It is interesting, but not entirely surprising, to notice that each term of these effects refers to a path contributing to the effect under examination: the SDE includes products of coefficients along all the paths connecting X^Y to Y_2, that is $X^Y \to Y_2$, $X^Y \to Y_1 \to M_2 \to Y_2$ and $X^Y \to Y_1 \to Y_2$, while SIE includes all path coefficients between X^M and Y_2, for instance the first two terms represent the paths $X^M \to M_2 \to Y_2$ and $X^M \to M_1 \to M_2 \to Y_2$. This means that the separable effects are functions of time, since the more time is elapsed between the baseline measurement and that of interest, the more paths are involved.

It can also be noticed that the more complex the model is, the more complex the expressions for separable effects become. However, this special case of mixed-effect models is *recursive*, i.e. they do not present loops and correlated error terms. As a consequence, exploiting the theory developed by Bollen (1987), if the models are linear and their coefficients are arranged in a matrix $\mathbf{B}$, the SDE and SIE can easily be retrieved by inspecting $\underline{\mathbf{B}} = \sum_{k=1}^{s} \mathbf{B}^k$, where s is the number of mediators and outcomes. For example, consider models (4.10)-(4.11) for three time occasions. They

can be rewritten as follows

$$\begin{pmatrix} X^M \\ X^Y \\ M_1 \\ M_2 \\ M_3 \\ Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{X} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \beta_{X} & 0 & \beta_{\ell_1(M)} & 0 & 0 & \beta_{\ell_1(Y)} & 0 & 0 \\ \beta_{X} & 0 & 0 & \beta_{\ell_1(M)} & 0 & 0 & \beta_{\ell_1(Y)} & 0 \\ 0 & \gamma_{X} & \gamma_{M} & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma_{X} & \gamma_{\ell_1(M)} & \gamma_{M} & 0 & \gamma_{Y} & 0 & 0 \\ 0 & \gamma_{X} & 0 & \gamma_{\ell_1(M)} & \gamma_{M} & 0 & \gamma_{Y} & 0 \end{pmatrix}}_{\mathbf{B}} \begin{pmatrix} X^M \\ X^Y \\ M_1 \\ M_2 \\ M_3 \\ Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \varepsilon_{M1} \\ \varepsilon_{M2} \\ \varepsilon_{M3} \\ \varepsilon_{Y1} \\ \varepsilon_{Y2} \\ \varepsilon_{Y3} \end{pmatrix}.$$

The separable direct and indirect effects of X on Y at each time can be found in the submatrix of $\underline{\mathbf{B}}$ made up of the first two columns and the last three rows. This allows to express the separable mediational effects in closed form, at least when the models are linear (identity link functions and no interactions).

4.3.2 Correlated random effects

The case of correlated random effects can be further divided into two sub-cases: the non-null correlation concerns random effects related to the same variable, i.e. $\mathbf{\Phi}$ is block diagonal, $\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_b & \mathbf{0} \\ \mathbf{0} & \mathbf{\Phi}_g \end{pmatrix}$, or random effects are free to covary with any element, so that $\mathbf{\Phi}$ is a full, non-diagonal matrix.

The former case is not conceptually different from that with uncorrelated effects, and assumptions **A1-A2** are still valid. The only difference is that integrating out random effects is less straightforward. To show how the correlation of random effects impacts the formulas of separable effects, let us consider again Normally distributed mediator and outcome, and models (4.8)-(4.9). Suppose that g_{Mi} is correlated with $g_{\ell_1(M)i}$ in model (4.9), and the other random coefficients are uncorrelated.

In formulas (4.10)-(4.11) there are not paths involving both random terms, since the time elapsed is too short. But consider the separable indirect effect of X on Y_3: among the different paths contributing to this effect, there is $X^M \to M_1 \to Y_2 \to M_3 \to Y_3$, which is analytically expressed by the product $\beta_X(\gamma_{\ell_1(M)}+g_{\ell_1(M)i})\beta_{\ell_1(Y)}(\gamma_M + g_{Mi})$. Given the correlation between g_M and $g_{\ell_1(M)}$, random effects cannot simply be deleted as in the previous case.

To obtain an expression free of random terms, it is necessary to solve the integral

$$\iint (\gamma_{\ell_1(M)} + g_{\ell_1(M)})(\gamma_M + g_M)\, f(g_{\ell_1(M)},\, g_M)\, dg_{\ell_1(M)}\, dg_M$$

where f is the joint density of the two random factors. Since f is a bivariate Normal with zero mean and non-diagonal covariance matrix, it can be proved that the integral above reduces to $\gamma_{\ell_1(M)}\gamma_M + \phi_{g\ell_1(M),gM}$, with $\phi_{g\ell_1(M),gM} = \text{Cov}(g_{\ell_1(M)}, g_M)$. Thus, the g-formula in (4.7) is not non-parametrically identified, since to derive the previous formula we assumed the Normality of random effects.

As the number of paths increases, the expressions for the separable effects become increasingly complex and deriving their closed form is not trivial. For this reason, if random effects are believed to be correlated and/or the mediator and the outcome models to be non linear, one of the solutions is to implement a code for the g-formula, without trying to solve it analytically.

Let us move to the case of non-diagonal $\mathbf{\Phi}$, so there is at least an element of $\mathbf{b}$ correlated to an element of $\mathbf{g}$. In Figure 4.2 the dashed bidirected arrow is then present. This simple modification makes separable effects unidentifiable, since the mediators and the outcomes are now part of the same unique district, which is recanting, since both treatment components affect nodes in the district. In addition, assumptions **A1** and **A2** fail, since M is no more conditionally independent from X^Y and Y is not conditionally independent from X^M. These considerations shed light into the nature of districts characterising mixed-effect models. Indeed, if random effects are uncorrelated, or if they are correlated only within their 'block', the mediators and the outcomes belong to two separate districts $\{M_1, \ldots, M_T\}$ and $\{Y_1, \ldots, Y_T\}$, which are not recanting, since nodes in a district are affected by only one of the components of X. The link between separable components and recanting districts has already been noted by Didelez (2019*b*).

4.4 Latent growth models

Considering a mediational LGM as that in Equations (2.8), it is possible to notice the differences with the mixed-effects models in Equations (4.4)-(4.5): in LGMs there is not a direct relationship between the observed variables, that is, the repeated measurements of mediator and outcome, instead they are indirectly connected through their latent factors. Moreover, latent factors in LGMs determine the trajectories of observed variables over time, while, in mixed-effect models, random effects explain heterogeneity among subjects and can be viewed as deviations from a common mean. From another point of view, mixed-effects models assume that the phenomenon happens at the level of repeated measures. In contrast, in LGMs, where a latent structure underlying the object of investigation is assumed, observed measurements are just indicators of this structure, since relationships of association and dependence involve

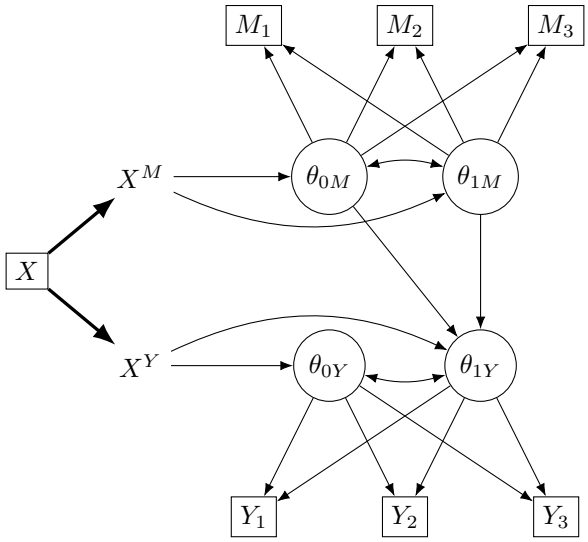

Figure 4.3: Three-wave latent growth model with separable components of X.

random factors, not measurements.

The different specification impacts also the definition of intervention and the corresponding causal interpretation of effects. In mixed-effect models we assume that intervening on X^M produces a change on the mediator and, likewise, an intervention on X^Y modifies the outcome, at each time. In contrast, in LGMs, intervening on X^M possibly leads to a change in the latent intercept θ_{0M} and the latent slope θ_{1M} of the mediator model. The same holds for X^Y and the latent factors in the outcome model. Then, intervening on X^M and X^Y affects indirectly the measurements of the mediator and the outcome, respectively. Once again, the effects of the intervention work at a latent level and have an indirect impact on the observed variables.

Considering again the running example, the intervention for favouring weight loss affects the mediator intercept, i.e. the average sense of self-efficacy at the beginning of the study, and its slope, that is, self-efficacy change rate, as well as the average weight loss at baseline and its change rate. In turn, the latent factors of self-efficacy may have an effect on those of weight loss. This mechanism is different from that described for the mixed-effect model, since treatment assignment does not modify self-efficacy and weight directly, but it affects their latent determinants. So, in a sense, the causal mechanism acts at a different, underlying level.

In contrast to what happens for mixed models with uncorrelated random effects, the separable effects in a latent growth model are never non-parametrically identified. Nonetheless, a modified version of assumptions **A1** -**A2** allows us to express the effects in terms of model parameters. Let $\boldsymbol{\Theta}$ denote the set of latent factors and $\boldsymbol{\Theta}^M$, $\boldsymbol{\Theta}^Y$ the subsets of factors in the mediator and the outcome model, respectively. We need

to extend consistency and exchangeability also to latent factors, i.e.:

- For each $\theta \in \mathbf{\Theta}$, $\theta(x) = \theta$ when $X = x$;
- For each $\theta \in \mathbf{\Theta}$, $\theta(x) \perp\!\!\!\perp X$.

In addition we assume that

B1.1 Each random factor in the mediator process is independent of X^Y conditional on the value of X^M,

$$\forall\ \theta_M \in \mathbf{\Theta}^M,\ \theta_M \perp\!\!\!\perp X^Y \,|\, X^M$$

B1.2 Each random factor in the outcome process is independent of X^M conditional on its predictors and the value of X^Y,

$$\forall\ \theta_Y \in \mathbf{\Theta}^Y,\ \theta_Y \perp\!\!\!\perp X^M \,|\, \mathbf{\Theta}^M,\ X^Y$$

B2 For each time t the response Y_t is independent of the value of X^M conditional on $\mathbf{\Theta}^M$ and X^Y,

$$Y_t \perp\!\!\!\perp X^M \,|\, \mathbf{\Theta}^M,\ X^Y.$$

Applying again the law of iterated expectation one can derive expressions for the mediational effects as follows:

$$\begin{aligned}
&\mathbb{E}[Y_t(X^M = x^*, X^Y = x)] = \\
&\sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t(X^M = x^*, X^Y = x) \,|\, \theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}, \\
&\qquad \theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M}, \theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y}, \\
&\qquad \theta_{1Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1Y}] \times \\
&P(\theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M})\ P(\theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M}) \times \\
&P(\theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y}) \times \\
&P(\theta_{1Y}(X^M = x^*, X^Y = x) \,|\, \theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}, \\
&\qquad \theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M},\ \theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y})
\end{aligned} \tag{4.14}$$

Making use of (4.1) and assumptions **B1.1**, **B1.2** and **B2** yields

$$\begin{aligned}
&\mathbb{E}[Y_t(X^M = x^*, X^Y = x)] = \\
&\sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t \,|\, X = x,\ \theta_{0M} = \tilde{\theta}_{0M},\ \theta_{1M} = \tilde{\theta}_{1M},\ \theta_{0Y} = \tilde{\theta}_{0Y}] \times \\
&P(\theta_{0M} = \tilde{\theta}_{0M} \,|\, X = x^*)\ P(\theta_{1M} = \tilde{\theta}_{1M} \,|\, X = x^*)\ P(\theta_{0Y} = \tilde{\theta}_{0Y} \,|\, X = x) \times \\
&P(\theta_{1Y} = \tilde{\theta}_{1Y} \,|\, X = x,\ \theta_{0M} = \tilde{\theta}_{0M},\ \theta_{1M} = \tilde{\theta}_{1M},\ \theta_{0Y} = \tilde{\theta}_{0Y})
\end{aligned} \tag{4.15}$$

A detailed proof is provided in Appendix B.

Assumptions **B1.1**, **B1.2** and **B2** do not lead to non-parametric identification. This is evident looking at Equation (4.15), because it depends on both observed and unobserved variables and involves quantities non-identifiable in absence of other parametric assumptions. Then, assumptions **B1.1**, **B1.2** and **B2** are useful for expressing the interventional expectation $\mathbb{E}[Y_t(X^M = x^*, X^Y = x)]$ as in (4.15), which is a 'pseudo' g-formula, since it involves unobservable quantities. Parametric identification is achieved by exploiting the parametric assumptions encoded by LGMs.

Notice that in Figure 4.3, **B1.1**, **B1.2** and **B2** hold. In particular, assumption **B1.2** holds for θ_{0Y} without conditioning on $\boldsymbol{\Theta}^M$, since X^Y suffices to ensure conditional independence from X^M, while $\theta_{1Y} \perp\!\!\!\perp X^M | X^Y, \theta_{0M}, \theta_{1M}$. It is important to remark that these assumptions hold in a model like that in Figure 4.3 either if the latent factors are uncorrelated or if the factors referring to the same process are correlated, for example θ_{0M} with θ_{1M} and θ_{0Y} with θ_{1Y}. However, if there are reasons to believe that all factors are correlated, this leads to a problem of identification of separable effects, since all factors belong to the same recanting district, similarly to what happens for mixed-effect models with correlated **b** and **g**. Assumptions **B1.1**, **B1.2** and **B2** are not satisfied and $\mathbb{E}[Y_t(X^M = x^*, X^Y = x)]$ cannot be expressed as in (4.15).

For a linear LGM with corresponding graph as that in Figure 4.3

$$\begin{aligned}
M_{it} &= \theta_{0Mi} + \theta_{1Mi}\lambda_t + \varepsilon_{Mit} \\
Y_{it} &= \theta_{0Yi} + \theta_{1Yi}\lambda_t + \varepsilon_{Yit} \\
\theta_{0Mi} &= \beta_0^{\theta_0} + \beta_X^{\theta_0} X_i + \zeta_{\theta_0 Mi} \\
\theta_{1Mi} &= \beta_0^{\theta_1} + \beta_X^{\theta_1} X_i + \zeta_{\theta_1 Mi} \\
\theta_{0Yi} &= \gamma_0^{\theta_0} + \gamma_X^{\theta_0} X_i + \zeta_{\theta_0 Yi} \\
\theta_{1Yi} &= \gamma_0^{\theta_1} + \gamma_X^{\theta_1} X_i + \gamma_{\theta_{0M}}^{\theta_1} \theta_{0Mi} + \gamma_{\theta_{1M}}^{\theta_1} \theta_{1Mi} + \zeta_{\theta_1 Yi}.
\end{aligned} \tag{4.16}$$

the separable direct effect on the difference scale is

$$SDE = (\gamma_X^{\theta_0} + \gamma_X^{\theta_1}\lambda_t)(x - x^*) \tag{4.17}$$

and the separable indirect effect

$$SIE = \lambda_t(\beta_X^{\theta_0}\gamma_{\theta_{0M}}^{\theta_1} + \beta_X^{\theta_1}\gamma_{\theta_{1M}}^{\theta_1})(x - x^*). \tag{4.18}$$

It is easy to notice that these expressions are the same one would obtain by means

of path analysis, since the effects are defined as sums of path-specific effects, which are in turn obtained as products of coefficients lying on the path. For example the direct effect is obtained by summing the effect through the path $X^Y \to \theta_{0Y} \to Y_t$ and that through $X^Y \to \theta_{1Y} \to Y_t$ for each t. In addition, these expressions do not depend on the correlation between random factors.

As remarked for mixed effect models, also in this case the mediational effects are time-varying. Notice, however, that equations (4.12)-(4.13) are very different from (4.17)-(4.18). The former show time dependence in the fact that, at each t, the number of paths connecting variables, and thus the number of terms in the formulas, increases. The latter encode time dependence only via λ_t, $t = 1, \ldots, T$.

4.5 Simulation study

In principle, the g-formula does not require any parametric assumption. However, if the number of variables is large, it can be difficult to apply it without recurring to (semi-)parametric models. The parametric g-formula relies on the correct specification of such models. In general, researchers are concerned about ignoring relevant variables, i.e. unobserved confounders. In longitudinal settings, there is also the issue of modeling the dynamic aspect of the phenomenon, which can be done in a wide variety of ways, involving latent structures or not. When a latent structure is taken into account, it is however difficult to select the most appropriate one.

In this section, we conduct a simulation study to assess how model misspecification and the choice of an incorrect latent structure, either in the mediator or the outcome model, affect the estimation of separable effects via the g-formula. We consider a simple scenario consisting of a binary exposure and Normally-distributed mediator and outcome for $n = 1,000$ subjects and $T = 5$ measurement occasions. Data were simulated from two different models: a linear mixed model as specified in Equations (4.8)-(4.9) with uncorrelated random effects, drawn from a multivariate Normal distribution with zero mean vector and identity covariance matrix, and a latent growth model as in Equations (4.16), where the ζ terms are from a standard Normal distribution. The model coefficients are reported in Table 4.1.

To analyse the extent to which estimates are affected by the use of wrong models, we considered two degrees of misspecification: moderate misspecification, where the model is very similar to the true one, except for a term, which is missing, and severe misspecification, where the models are completely wrong. Specifically, for the mixed-effect model we considered a moderate misspecification in the mediator model, which is assumed to be as in Equation (4.8) but with $\beta_{\ell_1(Y)} = 0$; for the LGM, we did not

Table 4.1: Coefficients of the mixed-effect and latent growth models used to generate data.

Mixed-effect model		**LGM**	
Coefficient	Value	Coefficient	Value
β_0	1.3	$\beta_0^{\theta_0}$	0.21
β_X	0.5	$\beta_X^{\theta_0}$	0.16
$\beta_{\ell_1(M)}$	0.27	$\beta_0^{\theta_1}$	0.7
$\beta_{\ell_1(Y)}$	0.11	$\beta_X^{\theta_1}$	0.47
γ_0	0.45	$\gamma_0^{\theta_0}$	0.3
γ_X	0.7	$\gamma_X^{\theta_0}$	0.14
γ_{Mt}	0.2	$\gamma_0^{\theta_1}$	0.59
$\gamma_{\ell_1(M)}$	0.08	$\gamma_X^{\theta_1}$	0.27
$\gamma_{\ell_1(Y)}$	0.34	$\gamma_{\theta_{0M}}^{\theta_1}$	0.44
		$\gamma_{\theta_{1M}}^{\theta_1}$	0.19

include θ_{0M} in the model for θ_{1Y}. As regards severe misspecification, the mixed-effect model was addressed as an LGM and, vice versa, the LGM as a mixed-effect model. This mirrors the case in which a researcher is completely agnostic about the true nature of the phenomenon under study.

For each generating mechanism we simulated $K = 500$ datasets on which we fitted each misspecified model and estimated the separable direct and indirect effects on the difference scale through the mediational g-formula. True values of the parameters were obtained straightforwardly for LGMs by applying formulas (4.17)-(4.18), while they were estimated asymptotically for mixed effect models, since the analytical form of effects is more complex. We evaluated the relative bias defined as

$$\text{Relative bias} = \frac{\sum_{k=1}^{K}(\hat{\theta}_{kt} - \theta_t)}{K\theta_t},$$

where $\hat{\theta}_{kt}$ is the estimate of the SDE or the SIE at time t obtained in the k-th simulation; the root mean square error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K}(\hat{\theta}_{kt} - \theta_t)^2}{K}},$$

and the coverage rate of 95% confidence intervals, obtained through $B = 500$ bootstrap samples.

The g-formula algorithm can be divided into three steps:

1. For each $t = 1, \ldots, T$, fit parametric models for the mediator and the out-

come conditional on the treatment and their histories, i.e. estimate densities $f_M(M_t|X, \overline{M}_{t-1}, \overline{Y}_{t-1})$ and $f_Y(Y_t|X, \overline{M}_t, \overline{Y}_{t-1})$.

2. Select $S \geq 10,000$.

   - Specify an intervention or a set of interventions to compare, creating two variables X^M and X^Y and setting them to the values of interest.
   - For each $s = 1, \ldots, S$ and $t = 1, \ldots, T$, draw a value $\tilde{m}_{st}$ for the mediator from $f_M(M_t|x^M, \overline{\tilde{m}}_{st-1}, \overline{\tilde{y}}_{st-1})$ estimated in step 1, conditional on x^M and its (simulated) history. Do the same for Y, conditional on x^Y and its history, i.e. draw $\tilde{y}_{st}$ from $f_Y(Y_t|x^Y, \overline{\tilde{m}}_{st}, \overline{\tilde{y}}_{st-1})$. For continuous distributions whose variance is not a function of the mean, the variance is estimated through the model residual mean squared error (Lin et al. 2020).

3. Compute the intervention mean estimate at each time $t =, \ldots, T$, by averaging the outcome expectation over simulated subjects:

$$\mathbb{E}[Y_t(X^M = x^M, X^Y = x^Y)] = \frac{1}{S}\sum_{s=1}^{S}\mathbb{E}[\tilde{y}_{st}] \tag{4.19}$$

Standard errors and confidence intervals for the average intervention effect can be estimated through non-parametric bootstrap, by repeating steps 2 and 3 B times, where B is the number of bootstrap samples.

The separable direct and indirect effects can easily be obtained by comparing expressions of the form (4.19), appropriately selecting x^M and x^Y. For example, if one wants to estimate the SDE and X is binary, one should compare $X^M = 0$, $X^Y = 1$ with $X^M = 0$, $X^Y = 0$.

4.5.1 Results

Simulations were conducted in the statistical software R. Results are shown in Tables 4.2 and 4.3. As expected, for data generated from a mixed-effect model, moderate misspecification produced an understimation of both SDE and SIE. This is consistent with the fact that the term expressing the lagged influence of the outcome was removed from the mediator model. When the misspecification is severe, the SDE is underestimated, while the SIE is overestimated, and the estimates are progressively farther from the true values as the amount of time elapsed increases. Both relative bias and RMSE are smaller (in absolute values) for effects estimated through the moderately misspecified model than for effects estimated using the severely misspecified model, and those of direct effects are generally lower than those of the indirect

Table 4.2: Results of simulations for data generated from a mixed model as in Figure 4.2 with uncorrelated random effects. For every model, each row refers to a different time $t = 1, \ldots, 5$.

Misspecification	**True**		**Estimates**		**Rel. bias**		**RMSE**		**Coverage rate**	
	SDE	SIE	SDE	SIE	SDE	SIE	SDE	SIE	SDE	SIE
	0.700	0.100	0.699	0.114	-0.001	0.140	0.100	0.029	0.866	0.832
	0.956	0.257	0.902	0.245	-0.056	-0.047	0.140	0.052	0.862	0.854
Moderate	1.064	0.384	0.961	0.316	-0.097	-0.178	0.173	0.096	0.784	0.668
	1.123	0.527	0.978	0.348	-0.129	-0.340	0.202	0.194	0.710	0.222
	1.162	0.660	0.983	0.361	-0.154	-0.453	0.229	0.309	0.622	0.320
	-	-	0.880	0.000	0.257	-0.999	0.251	0.100	0.814	0.000
	-	-	0.796	0.355	-0.167	0.383	0.261	0.137	0.886	0.868
Severe	-	-	0.712	0.711	-0.331	0.849	0.448	0.378	0.738	0.632
	-	-	0.628	1.066	-0.440	1.023	0.613	0.610	0.690	0.542
	-	-	0.544	1.422	-0.532	1.154	0.769	0.851	0.696	0.486

Table 4.3: Results of simulations for data generated from a latent growth model as in Figure 2.3. For every model, each row refers to a different time $t = 1, \ldots, 5$.

Misspecification	**True**		**Estimates**		**Rel. bias**		**RMSE**		**Coverage rate**	
	SDE	SIE	SDE	SIE	SDE	SIE	SDE	SIE	SDE	SIE
	0.140	0.000	0.138	0.000	-0.017	0.000	0.078	0.000	0.966	1.000
	0.410	0.160	0.479	0.089	0.169	-0.445	0.124	0.074	0.878	0.308
Moderate	0.680	0.319	0.821	0.177	0.207	-0.445	0.212	0.149	0.852	0.230
	0.950	0.479	1.162	0.266	0.223	-0.445	0.309	0.222	0.838	0.206
	1.220	0.639	1.504	0.354	0.233	-0.445	0.407	0.297	0.832	0.210
	-	-	0.252	0.038	0.802	0.038	0.130	0.039	0.460	1.000
	-	-	0.456	0.112	0.031	-0.030	0.127	0.029	0.874	0.892
Severe	-	-	0.625	0.329	-0.081	0.031	0.172	0.056	0.868	0.894
	-	-	0.770	0.546	-0.190	0.131	0.271	0.110	0.734	0.802
	-	-	0.897	0.794	-0.265	0.242	0.402	0.199	0.586	0.646

effects. In addition, they show an increasing trend over time. Coverage rates are lower than nominal level and they decrease as time elapses, so they are higher for the effect of X on Y at the first time occasions and tend to become smaller at subsequent times. The SDE coverage rates are very similar in both cases of misspecification, while for SIE they appear higher in the severe case, except for time 1.

Results for data generated from an LGM are less clear. The indirect effects estimated through the moderately misspecified model are smaller than the true ones, as expected, since the term $\beta_X^{\theta_0}\gamma_{\theta_{0M}}^{\theta_1}$ in Equation (4.17) is missing. Setting $\gamma_{\theta_{0M}}^{\theta_1} = 0$ influences also the estimate of SDE, which, on the contrary, is overestimated. In contrast, in the situation with severe misspecification, the SDE is underestimated, while the SIE is overestimated. Compared to mixed-effect model, it seems that misspecification affects the estimates in a more severe way, even when misspecification is only modest.

Relative bias and RMSE do not show clear patterns, and, in fact, in most cases they are smaller in the severe misspecification condition. Coverage rates for SDE are higher in the case of moderate misspecification, although below the nominal level. For SIE, coverage rates are better in the severe case and this may be due to the fact that estimates are heterogeneous enough to include the true effects, while in the moderate misspecification case the estimates are too far from the true values and the confidence intervals are not wide enough to include them.

These results, although referring to very simple models, suggest that even a moderate misspecification can have a remarkable impact on the estimates of separable effects, as well as on their confidence intervals. Then, researchers should carefully think about the most appropriate latent structure and model specification, especially in a longitudinal setting, where the time dynamics can be difficult to catch and many interactions among variables can be present.

4.6 Extensions

In sections 4.3 and 4.4, we have considered very basic, and unrealistic, scenarios for illustrating the separable effect approach to mixed-effect and LGMs. However, real-world settings are much more complex and may generally present additional issues. In this section, we are going to address some possible complications of the basic settings introduced previously.

4.6.1 Non-randomised treatment

In observational studies, the exposure cannot be assumed to be randomised, then $P(Y(X = x)) \neq P(Y \mid X = x)$. However, it is possible to achieve identification if conditional exchangeability holds, i.e. if it is possible to found sufficient sets of adjusting variables so that the exposure-mediator and exposure-outcome relationships are unconfounded. Classical adjustment methods can be applied, like the adjustment criterion discussed in Chapter 2 and the recanting witness/district criterion.

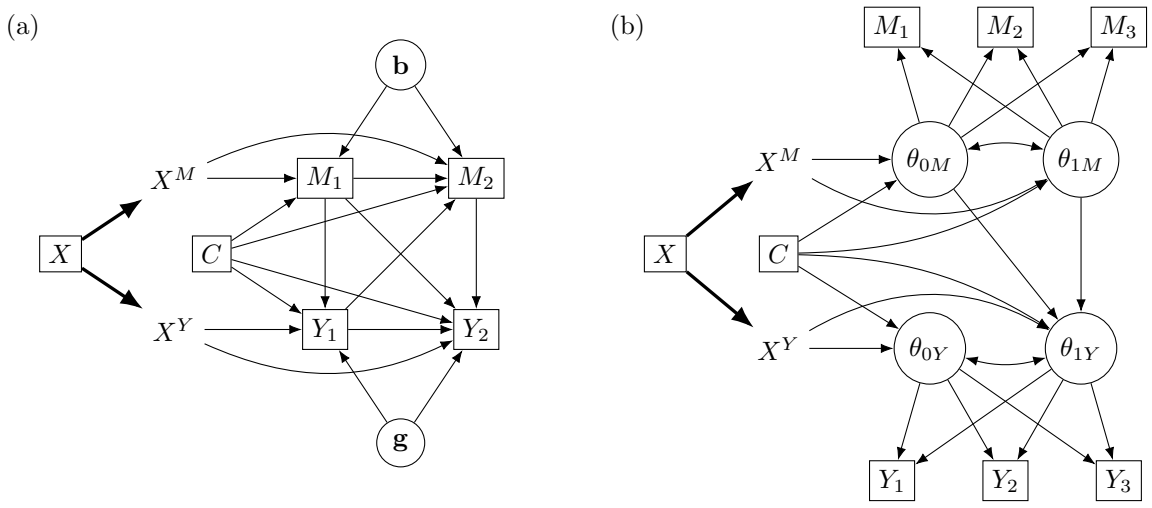

Figure 4.4: A separable mixed-effect model (a) and LGM (b) including a set of observed covariates C.

4.6.2 Inclusion of covariates

There are different types of covariates which may complicate the study setting and deserve discussion.

Let us start from the typical covariates included in any study, baseline covariates such as sex and age. Figure 4.4 shows a mixed-effect model (a) and an LGM (b) with a vector of observed baseline covariates C. In this case, the separability of the exposure components can be obtained by simply modifying assumptions **A1**-**A2** and **B1.1**, **B1.2** and **B2**, so that C is included in the conditioning set. Then, for example, assumption **A1** becomes $M_t \perp\!\!\!\perp X^Y \mid \overline{M}_{t-1}, \overline{Y}_{t-1}, C, X^M$ and the others can be rewritten analogously. As remarked by Didelez (2019*b*), variables in C confound the relationship between the mediator and the outcome, but, unlike the traditional natural effect approach to mediation analysis, conditioning on them is not necessary to make the counterfactuals $Y(x, m)$ and $M(x)$ independent, since the target estimand does not involve any intervention on the mediator. In contrast, conditioning on C ensures the conditional independence of each variable from the opposite component of X.

Another kind of covariates which is often present in longitudinal studies are post-treatment confounders, which, as discussed in Section 1.2.5, can be problematic for the estimation of some mediational effects. In a separable effect framework, the presence of a variable confounding the mediator-outcome relationship and affected by the treatment gives rise to different scenarios.

Let us start from the mixed-effect model in Figure 4.5 (a): X affects the observed variable (or set of variables) Z, which is affected by the mediator and the outcome at

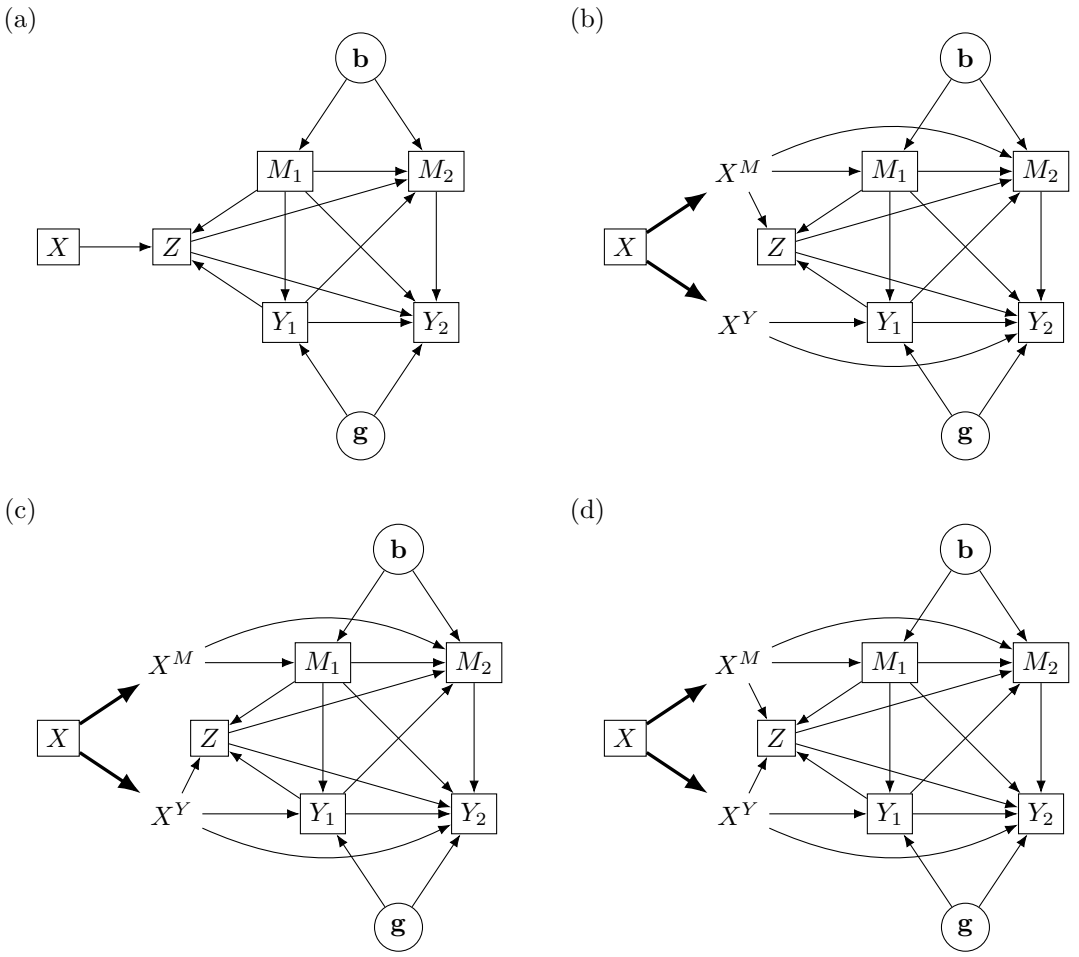

Figure 4.5: A mixed-effect model with post-treatment confounding (a), and the possible scenarios in a separable effects framework: (b) and (c) preserve separability, while it fails in scenario (d).

time 1, but affects them at the subsequent time. As already discussed in Section 1.2.5, this setting can correspond to three different scenarios, (b) and (c) where just one of the treatment components affects the covariate, and (d) where both the components have an effect on Z.

In cases (b) and (c) the separability of components can be retained by using the following assumptions, which are an expanded version of **A1** and **A2**:

A1′ $M_t \perp\!\!\!\perp X^Y \,|\, \overline{M}_{t-1}, \overline{Y}_{t-1}, Z, X^M$

A2′ $Y_t \perp\!\!\!\perp X^M \,|\, \overline{M}_t, \overline{Y}_{t-1}, Z, X^Y$

A3.1 $Z_{X^M} \perp\!\!\!\perp X^Y \,|\, \overline{M}_{t-1}, \overline{Y}_{t-1}, Z_{X^Y}, X^M$

A3.2 $Z_{X^Y} \perp\!\!\!\perp X^M \,|\, \overline{M}_t, \overline{Y}_{t-1}, Z_{X^M}, X^Y$

where $Z \equiv (Z_{X^M}, Z_{X^Y})$, Z_{X^M} and Z_{X^Y} being subsets of Z such that, for each $t = 1, \ldots, T$, X^M affects Y_t only through M or variables in Z_{X^M}, and X^Y affects M_t only

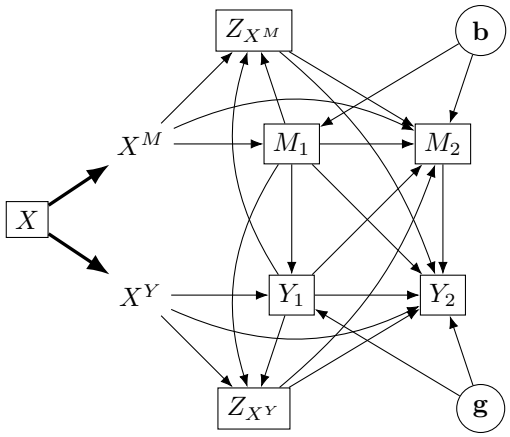

Figure 4.6: Mixed-effect model with Z partition.

through Y or variables in Z_{X^Y}. If such subsets exist, they are called a Z *partition* (Stensrud et al. 2021).

In Figure 4.5 (b) $Z \equiv Z_{X^M}, Z_{X^Y} = \emptyset$, vice versa in (c) $Z \equiv Z_{X^Y}, Z_{X^M} = \emptyset$. In both cases, the assumptions introduced below are satisfied, then, if exchangeability and consistency hold, the separable effects are identifiable.

In contrast, separability cannot be achieved in scenario (d). Indeed, if we assume $Z \equiv Z_{X^M}, Z_{X^Y} = \emptyset$, assumption **A3.1** fails, since $Z \equiv Z_{X^M}$ is not independent of X^Y; similarly, if we assume $Z \equiv Z_{X^Y}, Z_{X^M} = \emptyset$, assumption **A3.2** fails, since $Z \equiv Z_{X^Y}$ is not independent of X^M. Figure 4.6 shows a scenario in which Z is properly partitioned into nonempty subsets, with $Z_{X^M} \cap Z_{X^Y} = \emptyset$. It is easy to see that **A1′**, **A2′** and both assumptions in **A3** are satisfied.

Now, let us move to LGMs. There are two possible ways in which Z can be a post-treatment confounder, as shown in Figure 4.7: by affecting the latent factors of both the mediator and the outcome (a), or affecting the mediator and the outcome directly (b). As shown for mixed-effect models, these settings can correspond to different scenarios in the separable effects framework. If Z can be partitioned, separability of the X components is preserved, while, if Z is affected by both components X^M and X^Y, this makes identification of separable effects impossible.

In setting (a), assumptions **B1.1, B1.2** and **B2** need to be modified as follows

B1.1′ $\forall\ \theta_M \in \mathbf{\Theta}^M,\ \theta_M \perp\!\!\!\perp X^Y \,|\, X^M, Z_{X^Y}$

B1.2′ $\forall\ \theta_Y \in \mathbf{\Theta}^Y,\ \theta_Y \perp\!\!\!\perp X^M \,|\, X^Y, \mathbf{\Theta}^M, Z$

B2′ $Y_t \perp\!\!\!\perp X^M \,|\, X^Y, \mathbf{\Theta}^M, Z$

B3.1 $Z_{X^M} \perp\!\!\!\perp X^Y \,|\, X^M$

B3.2 $Z_{X^Y} \perp\!\!\!\perp X^M \,|\, X^Y$

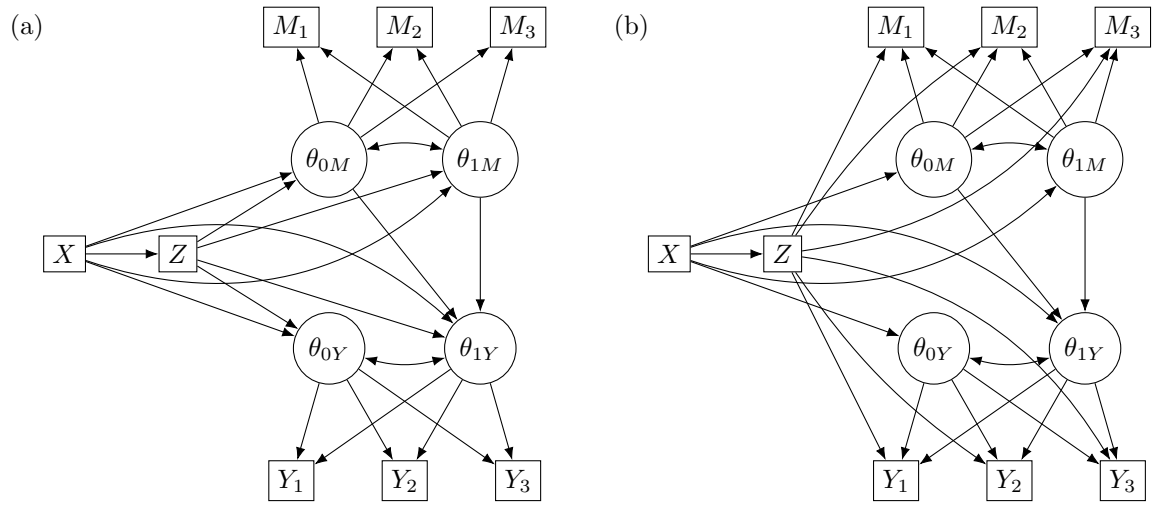

Figure 4.7: A latent growth model with a post-treatment confounder Z, where it affects the mediator and the outcome latent factors (a) or their observed indicators (b).

These assumptions hold in Figures 4.8 (a)-(c), but fail in (d), since if we assume $Z \equiv Z_{X^M}, Z_{X^Y} = \emptyset$, then Z_{X^M} is not independent of X^Y and similarly, if $Z \equiv Z_{X^Y}, Z_{X^M} = \emptyset$, then Z_{X^Y} is not independent of X^M.

As regards the setting depicted in Figure 4.7 (b), assumptions **B1.1** and **B1.2** stay the same, **B2** needs to be modified as follows

B2$''$ $Y_t \perp\!\!\!\perp X^M \,|\, X^Y, \boldsymbol{\Theta}^M, Z_{X^M}$

and **B3.1**, **B3.2** are added. Also in this case, if both the treatment components affect Z, then the assumptions are violated.

It is worth remarking that, when the partition of Z is improper, in the sense that one of its two subsets is the empty set, separability of effects does not hold completely. Consider Figure 4.5 (b): in this case the X^M separable effect encompasses both indirect effects of the treatment on Y through the mediator, e.g. the path $X^M \rightarrow Z \rightarrow M_2 \rightarrow Y_2$, and also direct effects not through M, e.g. the path $X^M \rightarrow Z \rightarrow Y_2$. In contrast, the X^Y effect still quantifies the direct effect of the treatment on Y, although it does not capture all direct effects, since some paths start from X^M. An analogous interpretation can be given for (c), where X^M captures only the indirect effects of X on the outcome, but not all, since part of these indirect effects originate from X^Y.

An alternative way to address post-treatment confounders when they cannot be partitioned is to decompose X into three components X^M, X^Y and X^Z, the last being the component of the treatment other than X^M and X^Y. For example, the graphs in Figures 4.5 (a) and 4.7 (a) can be alternatively represented as in Figure 4.9. In such settings, assumptions need to be redefined so that each component results

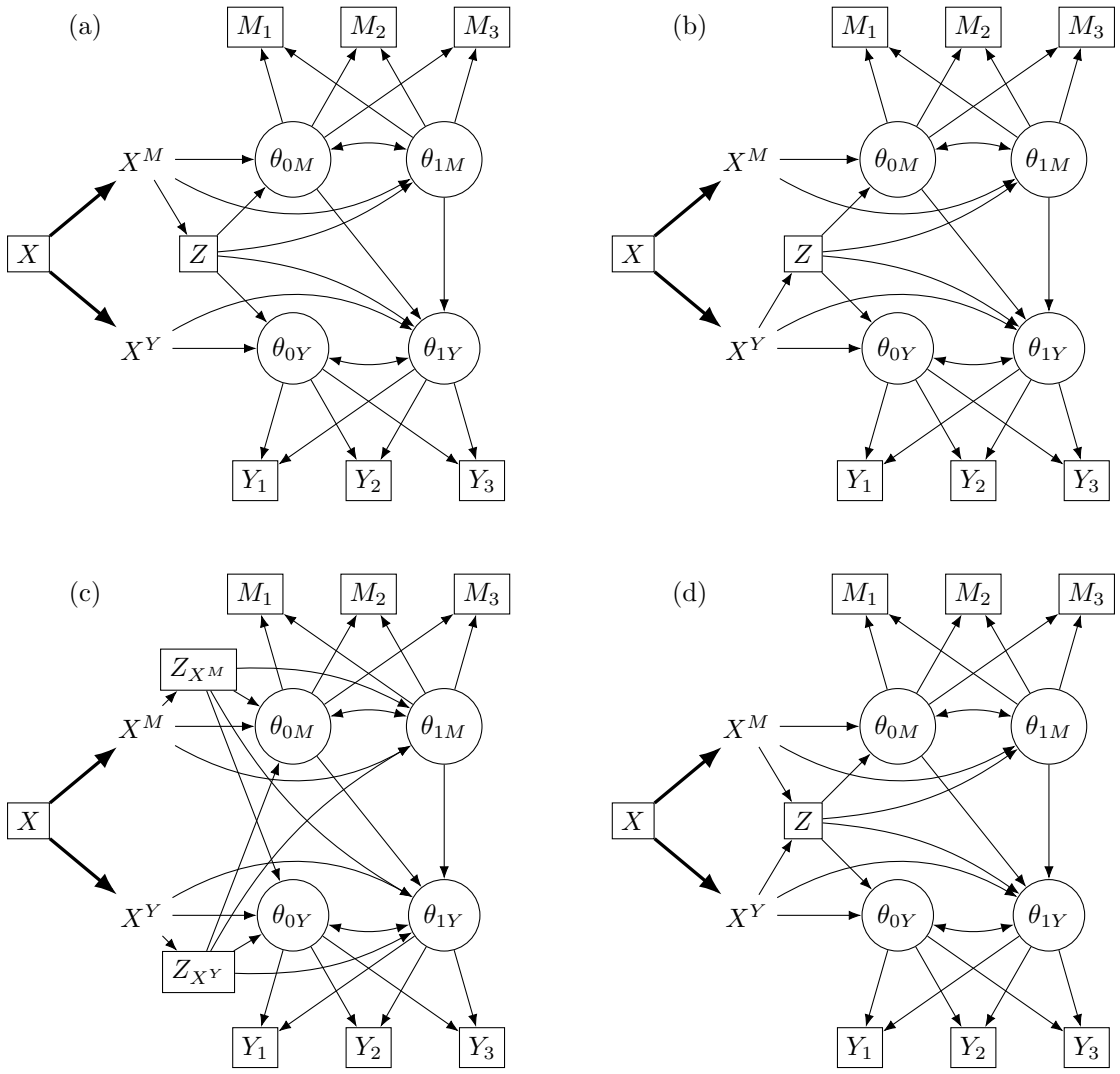

Figure 4.8: Expanded graphs showing different scenarios related to Figure 4.7 (a).

independent of the other two conditioning on an appropriate set of variables. We will not deepen this aspect. For an extensive discussion on post-treatment confounders in a separable effect survival context, Z partitions and the way on which they affect the identifiability and the meaning of separable effects see Stensrud et al. (2021).

If Z were time-varying, this could be easily addressed in a mixed-effect model, and the assumptions need only moderate modifications, like including a temporal subscript for Z and conditioning also on its history. However, the inclusion of a time-varying confounder in an LGM is not so straightforward. Indeed, Z can affect the observed indicators or it can be modeled as having its own underlying trajectory, so that its latent factors have an effect on those of the mediator and the outcome. These two cases are represented in Figure 4.10 (a) and (b), respectively. The setting in (a) is not conceptually different from that in Figure 4.7 (b): assumption **B2″** should just

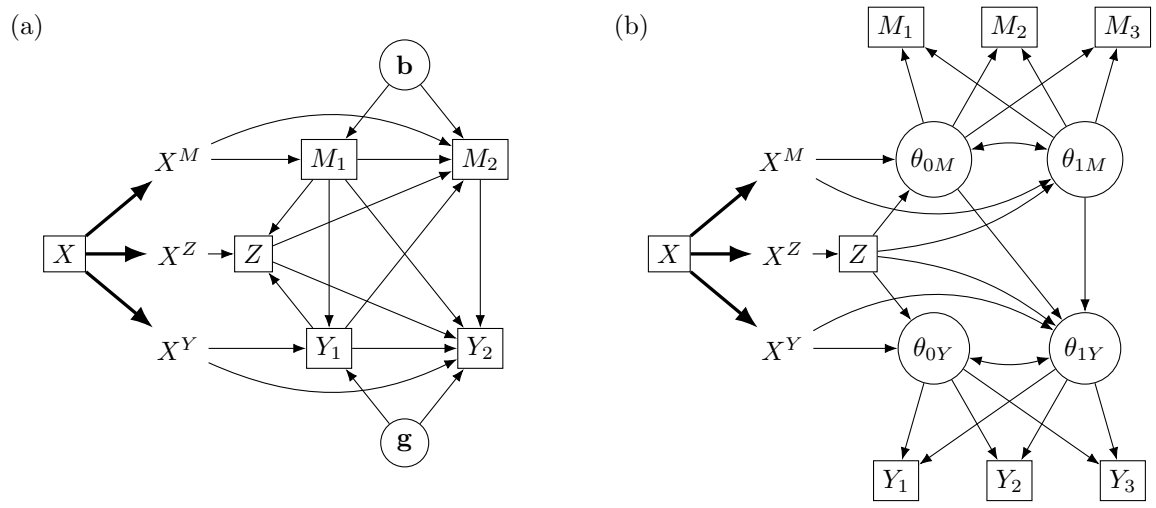

Figure 4.9: A separable mixed-effect model (a) and LGM (b) with a three-way decomposition of treatment.

take into account the time-varying nature of Z, i.e. $Y_t \perp\!\!\!\perp X^M \mid X^Y, \mathbf{\Theta}^M, Z_{t,X^M}$, $t = 1, \ldots, T$, and the other assumptions do not vary.

In contrast, the situation depicted in (b) is more complex. The post-treatment confounders in this case are the latent factors characterising the trajectory of Z. The existence of a proper partition would imply the existence of four latent factors, two intercepts and two slopes, where each intercept-slope pair is affected by only one between X^M and X^Y. Alternatively, one should assume that each component of the treatment exerts its effect only on one of the two latent factors, for example X^M on the intercept θ_{0Z} and X^Y on θ_{1Z}. The former hypothesis is meaningless, the latter is quite unrealistic, and it is difficult to think of a real-world context in which this kind of partition can happen. It is much more plausible to assume that both Z latent factors are influenced by only one of the treatment components. Assumptions are as in **B1′**, **B2′** and **B3**, where Z is substituted by its latent factors.

The same line of reasoning applies to time-varying covariates not affected by the exposure and analogous assumptions can be derived.

4.6.3 Time-varying treatment

So far, we have addressed settings where the treatment is measured at baseline and does not vary over time. However, in several cases estimating the effect of a time-varying treatment may be of interest. To the best of our knowledge, this issue has not yet been addressed in a separable effect framework, so in this section we just discuss the possible implications of including a time-varying treatment in mixed-effect or latent growth models.

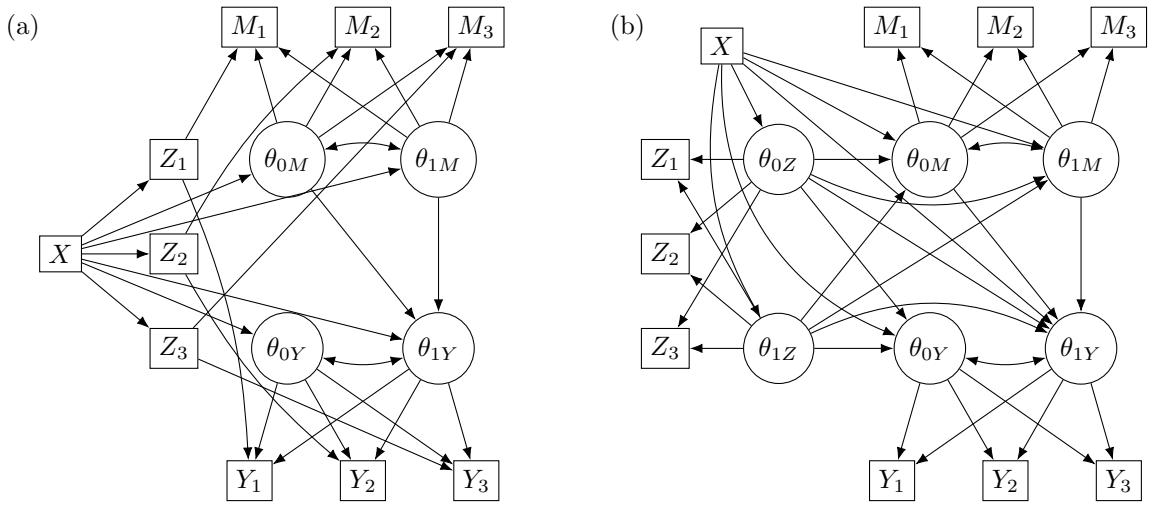

Figure 4.10: Latent growth model with a time-varying covariate Z. In (a) Z affects the observed variables, in (b) two latent factors shape its trajectory and affect the latent factors of the mediator and the outcome.

Robins et al. (2020) mention the possibility to extend the separable effects approach to multiple treatments, although they do not develop this idea. A setting with multiple treatments is very similar to a setting with a unique treatment measured over time, for $t = 1, \dots, T$. At each time t, the treatment can be decomposed into three components, X_t^M and X_t^Y as before and X_t^X, which affects the treatment at the subsequent time. For example, a mixed-effect model with time-varying treatment can be represented as in Figure 4.11.

In contrast, a time-varying treatment in an LGM can be addressed as in Figure 3.2, i.e. the treatment is assumed to have its own trajectory characterised by some latent factors. These factors affect those of the mediator and the outcome. As a consequence, the separability should involve the treatment latent factors, but this would imply that each factor should be split into different components, each one affecting only the factors of M or Y. We believe that this decomposition is very counter-intuitive and it would require an intervention on latent variables, unfeasible in practice. Thus, the separable effect approach seems not to be appropriate to address LGMs from a causal perspective when the treatment is time-varying.

It is worth remarking that the presence of a time-varying treatment presents more challenges than a baseline treatment, see the last chapters of Hernán and Robins (2020). In particular, when the estimation of causal effects is done through the g-formula as we proposed, Robins and Wasserman (1997) show that the g-formula is always subject to misspecification, a phenomenon known as *g-null paradox*. To be more precise, the authors show that if time-varying confounders are affected by the treatment and the *sharp g-null hypothesis* hold, i.e. if the treatment has no causal

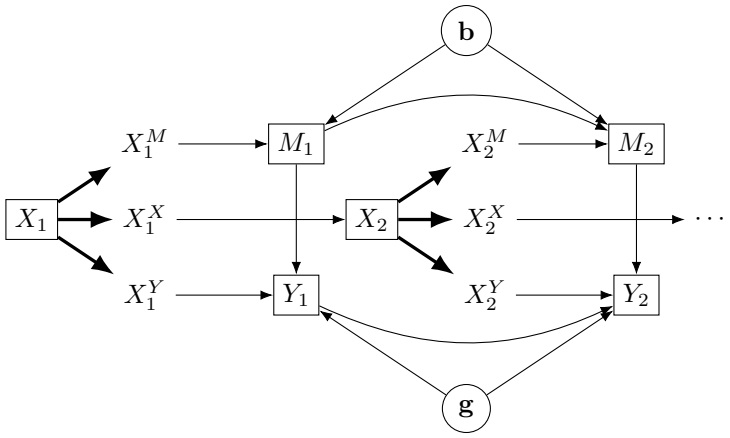

Figure 4.11: Mixed-effect model with time-varying treatment in a separable effects framework. Cross-lagged effects were not included to avoid clutter.

effect on the outcome at any time, then the parametric g-formula will falsely reject the hypothesis of null causal effect with probability approaching one, even if all the identifiability assumptions are satisfied.

The models used by Robins and Wasserman (1997) to introduce the g-null paradox did not involve random effects or latent trajectory factors, so a first issue to investigate should be if the g-null paradox occurs even when using these kinds of models. If so, this would call for the employment of different estimation methods such as inverse probability weighting (IPW) or g-estimation. How these methods could be applied in a separable effects framework is still to be investigated, although IPW has already been proposed in a survival setting by Stensrud et al. (2020).

4.7 Discussion

In this chapter, we have applied the separable effects approach for mediation analysis proposed by Robins and Richardson (2011) to mixed-effect and latent growth models. For each of them, we proposed a set of assumptions which suffice for the identification of separable effects and derived formulas to estimate them using the g-formula.

When the relationships are linear, the separable effects can be expressed in closed forms that have a direct correspondence with the graphs representing the models. As we saw, the separable direct effect is the sum of effects along all paths starting from X^Y, while the separable indirect effect is obtained as the sum of all paths having X^M as starting node. This is a connection with path analysis. Another advantage of the separable effects approach is that the estimands do not require any intervention on the mediator: this feature is particularly useful in LGMs, where it would be difficult to conceive an intervention on the latent factors characterising the mediator trajectory.

In our discussion, we limited ourselves to linear models, although mixed-effect models have been extended to the natural exponential family. We have not addressed this issue, but, in principle, the g-formula in (4.7) can accommodate any variables whose distribution belongs to this class. However, the complexity of the models, in terms of non-linearity of link functions, interaction terms and order of lagged effects, impacts the estimation procedure in two ways: first, finding closed forms for mediational effects becomes unpractical, so they have to be obtained through simulation, as described in Section 4.5; second, the more complex the model, the higher the computational intensity of the algorithm. This is one of the drawbacks of the estimation via g-formula, mainly due to the need to use bootstrap for estimating the effects' confidence intervals, a highly time-consuming task.

This issue can be exacerbated when the design is unbalanced. We have not addressed this complication, and we think that it could be an interesting extension of the present work. Unbalanced designs are the rule, not the exception in real-world analyses, as the one discussed in the next chapter shows. We have considered balanced designs to make the presentation of the approach clearer and easier to follow, but we acknowledge that this is a limitation. A special case of unbalanced design is given by censoring in survival analyses, which has been addressed in a separable effect framework by Stensrud et al. (2021) and Stensrud et al. (2020). How design unbalance should be addressed in other model settings is yet to be investigated, although it should not add conceptual difficulties. Combining separable effects and definition variables discussed in Chapter 3 could be a possible promising direction.

Finally, although this limitation has already been highlighted in the previous chapters, we want to remark again that conceiving different and separate components of X, i.e. components each having an effect only on the mediator or the outcome and which can in principle be randomised separately, is not always feasible. When researchers believe that the outcome component exerts its effect also on the mediators or vice versa, the separable effects approach should be avoided.

Chapter 5

An application: the COVCO-Basel study

In this chapter, we analyse a data set from a cohort study carried out in Switzerland, the *COVCO-Basel study* (Keidel et al. 2021). Conducted by the Swiss Tropical and Public Health Institute (TPH), this study provides rich information about the spread of Covid-19 in the adult population of Basel and changes in their mental health due to the epidemiological situation. The data set includes a high number of variables, but just a subset of them will be used in the statistical analyses carried out in this chapter.

In the first section we describe how data were collected and the variables included, in the second we discuss the aim of the study, in the third we carry out an exploratory analysis, while the fourth section is devoted to the application of causal mixed-effects models as described in Chapter 4. In the last section we draw some conclusions.

5.1 Data collection and description

The COVCO-Basel study is an extension of the Corona Immunitas study in Basel. Corona Immunitas is a study carried out in Switzerland at the cantonal and national level strarting from April 2020, which aimed to assess the spread of SARS-CoV-2 infection in the general population and subgroups of interest (West et al. 2020). The COVCO-Basel study complements this Seroprevalence cohort with another one called Digital. In both cohorts, subjects completed a baseline questionnaire and at least one online questionnaire per month; in addition, participants belonging to the former cohort were tested for antibodies and they were allowed to invite family members being at least seven years old. Assignment to either of the two cohorts is randomised.

The data collection from the Swiss TPH began in July 2020 and is still ongoing.

A sample of adult subjects resident in the cantons of Basel-Stadt (BS) and Basel-Landschaft (BL) was provided to Swiss TPH by the Swiss Federal Statistical Office. Each subject received an invitation letter by Swiss TPH, including login credential to access the study website. Those who gave their consent to participate received a link to fill the baseline and the follow-up questionnaires through the RedCap platform. Subjects in the Seroprevalence cohort were additionally asked if they wanted to participate in the study alone or with some family member, and they were invited to the study center to collect blood and saliva samples. Invitations were sent between July and December 2020, so each subject entered the study in a different date.

At the time of writing, data up to August 2021 on $n = 12,048$ subjects are available. The data set includes demographic and social variables such as sex, age, nationality, education and monthly income, which were measured at baseline, and study-specific variables. In the online questionnaires, subjects were asked to assess their levels of depression, stress and anxiety, whether they felt lonely or had worries concerning different aspects of their life, like health, economic situation and the impact of restrictions on their social life. These variables were assessed monthly, as well as the positivity to a Covid test of the interviewed subject or one of his/her family members (only for those in the Seroprevalence cohort).

Specifically, the variables *depression, anxiety* and *stress* were measured through the Depression and Anxiety Stress Scale (DASS-21), which includes twenty-one questions, seven for each item, whose answers are ordinal, ranging from 0 to 3. The scores are obtained as sums of the answers to each question, thus, each variable can range between 0 and 21, where higher scores represent more severe mental distress.

The variables related to worries involve different aspects of life. Subjects were asked to answer questions concerning their degree of concern about: their own and the national economic situation (*worries_econ*); their own and their family's health as well as the risk of catching/spreading Covid-19 (*worries_health*); the quality of their relationships with family and friends/colleagues (*worries_social*); and the possible restrictions to holidays/travel abroad and cultural life (*worries_cultural*). All variables are dichotomous and measured monthly.

The sample includes $6,531$ females, $5,509$ males and 8 subjects who do not recognise themselves in either category. The sample is rather balanced regarding the canton of residence, with 49.6% of subjects living in BS and 50.4% in BL. Approximately 70% of subjects is over 50 and most participants have a monthly household income ranging from 3,000 to 9,000 CHF (47%). As already mentioned, the study covers the period from July 2020 to August 2021, but in this chapter we will analyse just a subset of the data set, as detailed in the next section.

5.2 Objective of the study

The primary goal of this study is to investigate the evolution of depression in a period of high spread of Covid-19 in Switzerland, and its relationship with income, considered as a proxy of socio-economic status. Many studies have investigated the link between income or socio-economic status and mental health, highlighting a negative relationship, i.e. the lower the income the higher the risk of developing depression and related mental disorders. This pattern seems to be consistent across countries, see for example Freeman et al. (2016), Hoebel et al. (2017), Domènech-Abella et al. (2018), Hudson (2005), Kourouklis et al. (2020), Osafo Hounkpatin et al. (2015), and Patel et al. (2018) for a systematic review on the topic.

Some of the aforementioned studies (Hudson 2005, Domènech-Abella et al. 2018) considered also intermediate variables in the pathway connecting income or socio-economic status to mental disorders, such as life-style habits (smoke, unhealthy diet) and psycho-social factors (loneliness, social isolation). In our analyses, we use a similar approach: in particular, we aim to quantify the effect of income on mental health during the pandemic, and disentangle the direct effects from the indirect effects mediated by worries concerning different aspects of life. We will focus just on weeks between October and December 2020, which, as shown in Figures 5.1 and 5.2, correspond to the second wave of Covid-19 spread in Switzerland.

Because of the complex design of the study, the data structure is equally complex. Each subject has his/her own date of entry into the study and subsequent (unique) dates of follow-up. In particular, although follow-up questionnaires were sent approximately every 28 days, subjects often answered some days after, and these delays led to some subjects having up to three observations in the same month. As a consequence, one of the first problems is the identification of an appropriate time interval: weeks are too tight and would lead to an excessive number of occasions per subject, while months are too coarse and, as mentioned before, some subjects may have multiple measurements associated to a certain month. An adequate solution seemed then to use fortnights, or, in other words, to use intervals of 14 days.

The period of interest includes roughly seven fortnights, whose corresponding dates are shown in Table 5.1, and we selected subjects with at least three measurements in this period. This restricted the sample to 3,411 individuals. There are no subjects having variables collected at the penultimate fortnight, then our interest is actually on six time intervals. Among the numerous variables present in the data set, our focus is on the evolution of depression. In order to gain insights into its dynamics and investigate if it is associated or caused by other variables, such as worries, we

Figure 5.1: Daily new confirmed COVID-19 cases per million people in Switzerland, 7-day rolling average. Source: Our World in Data https://ourworldindata.org/covid-cases.

Figure 5.2: Share of daily positive Covid-19 tests in Switzerland, 7-day rolling average. Source: Our World in Data https://ourworldindata.org/covid-cases.

Table 5.1: Fortnights considered in the analysis and corresponding dates.

Fortnight	**Dates**
6	2-15 October 2020
7	16-29 October 2020
8	30 October - 12 November 2020
9	13-26 November 2020
10	27 November - 10 December 2020
11	11-24 December 2020
12	25 December - 7 January

will start by an exploratory analysis and proceed with a causal modeling approach.

5.3 Exploratory analysis

In this section, we show the results of some preliminary analyses carried out to explore the relationship between income, depression and the different types of worries.

Figure 5.3 shows how the percentage of worried people varies over the fortnights considered, i.e. those from the sixth to the twelfth, excluding the eleventh, according to the coding in the data set. It is interesting to notice that, for every kind of worry, the percentage of worried people is always lower than that of non-worried subjects, except for cultural worries, where, at each occasion, the proportion of worried subjects is approximately the same or even higher than the proportion of non-worried ones.

Next, we explore the relationship between worries and monthly income, using its 5-category version. As can be seen from Figure 5.4, the percentage of subjects having cultural worries, stratified by income, is approximately the same as that of non worried subjects. In contrast, as concerns the non-cultural worries, the majority of worried subjects has an income lower than 9,000 CHF. As expected, subjects belonging to the highest categories of income show less degree of worry. This seems to suggest a possible relationship between income and worries.

Focusing on depression, Figure 5.5 shows the distribution of depression scores over time stratified by sex (top panel), income (central panel) and age category (bottom panel). In general, female distribution of depression is more variable than the male one, and this holds true also for the distribution corresponding to more deprived and younger subjects. Not surprisingly, subjects in the highest category of income have median depression score null and the variability in their score is clearly smaller than that of the other groups. Similar observations can be made for age, where younger people show generally higher median and variability than older subjects.

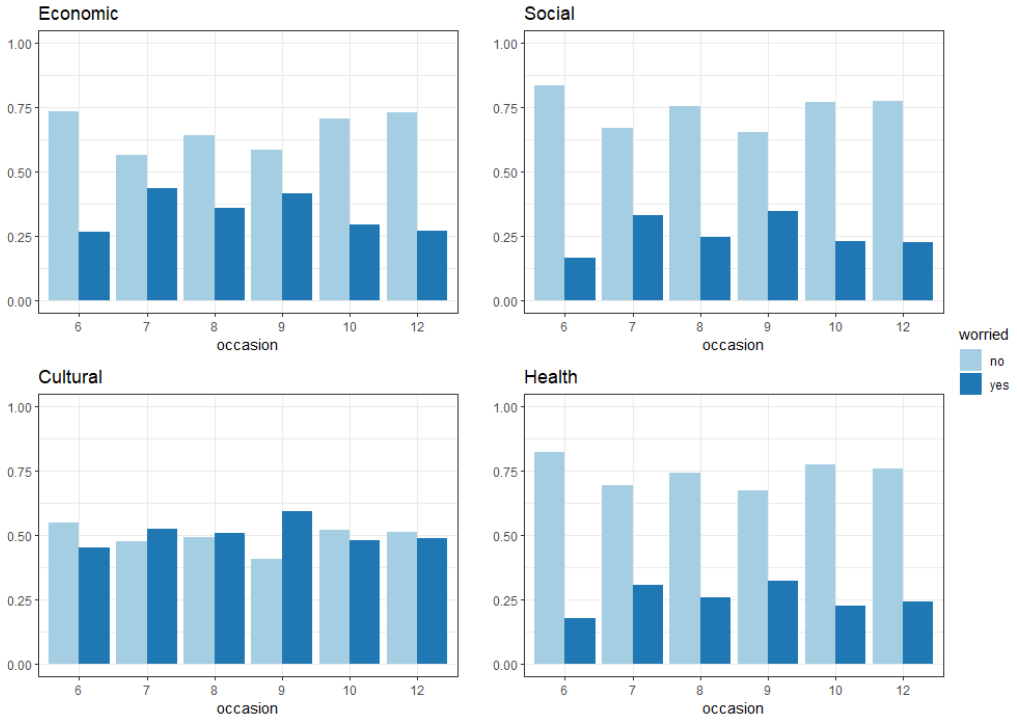

Figure 5.3: Proportion of people having economic, social, cultural and health worries over the period of interest.

Figure 5.6 shows the distribution of depression by worries. For any kind of worry, the distribution of depression scores of worried subjects has a higher variance and, in some cases, a highest median than that of non-worried subjects, as is visible for social worries. Cultural worries correspond to the least variable distributions, and the median depression score generally does not differ among worried and non worried people. These graphs suggest that investigating the relationship between income, worries and depression may be worthy, and this will be the object of the next section.

5.4 Inference

The data set includes several variables having non-Normal distributions: there are categorical variables, such as income and education, binary variables, such as sex or the different worries, and numeric asymmetric variables. Given this wide variety of distributions, the methodology proposed in Chapter 3 is not applicable straightforwardly to the present data. We can however use the approach discussed in Chapter 4, and use mixed-effect models within a separable effects framework.

As already mentioned, our exposure is monthly income and we are interested in estimating its direct effect on depression and the indirect effects through worries. Using income as exposure poses conceptual challenges concerning the causal meaning

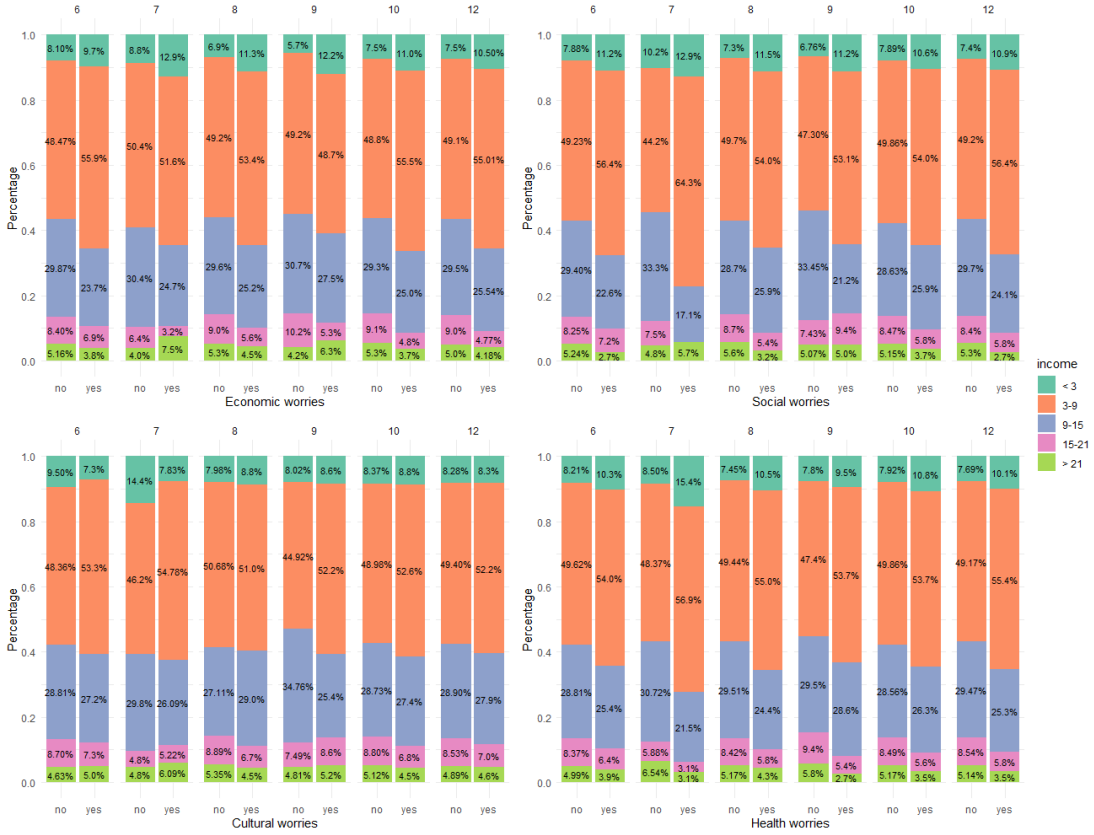

Figure 5.4: Proportion of worried subjects over the occasions of interest stratified by income, expressed in thousands of CHF.

of an intervention of such a variable, and issues related to the specific approach to be employed, i.e. what the separation of income implies and how it can be implemented. These points will be addressed in the next two sections and will be followed by the description of the analyses conducted and the results obtained.

5.4.1 A note on the meaning of interventions on income

In the causal inference literature, causal effects are generally defined in terms of an intervention on or a manipulation of the exposure of interest. For some variables, it is very easy to think to a possible intervention, while for others, like sex or race, this does not hold true. Income is in a grey area: if you asked to an economist and an epidemiologist whether income can be an exposure of interest for a causal analysis, the former would answer yes, the latter would probably say no. Although it is easier to conceive an intervention on income than one on sex or race, epidemiologists are used to consider income as a baseline covariate, not as a manipulable exposure. However, at least hypothetically, income could be increased through state subsidies, or decreased indirectly through taxation.

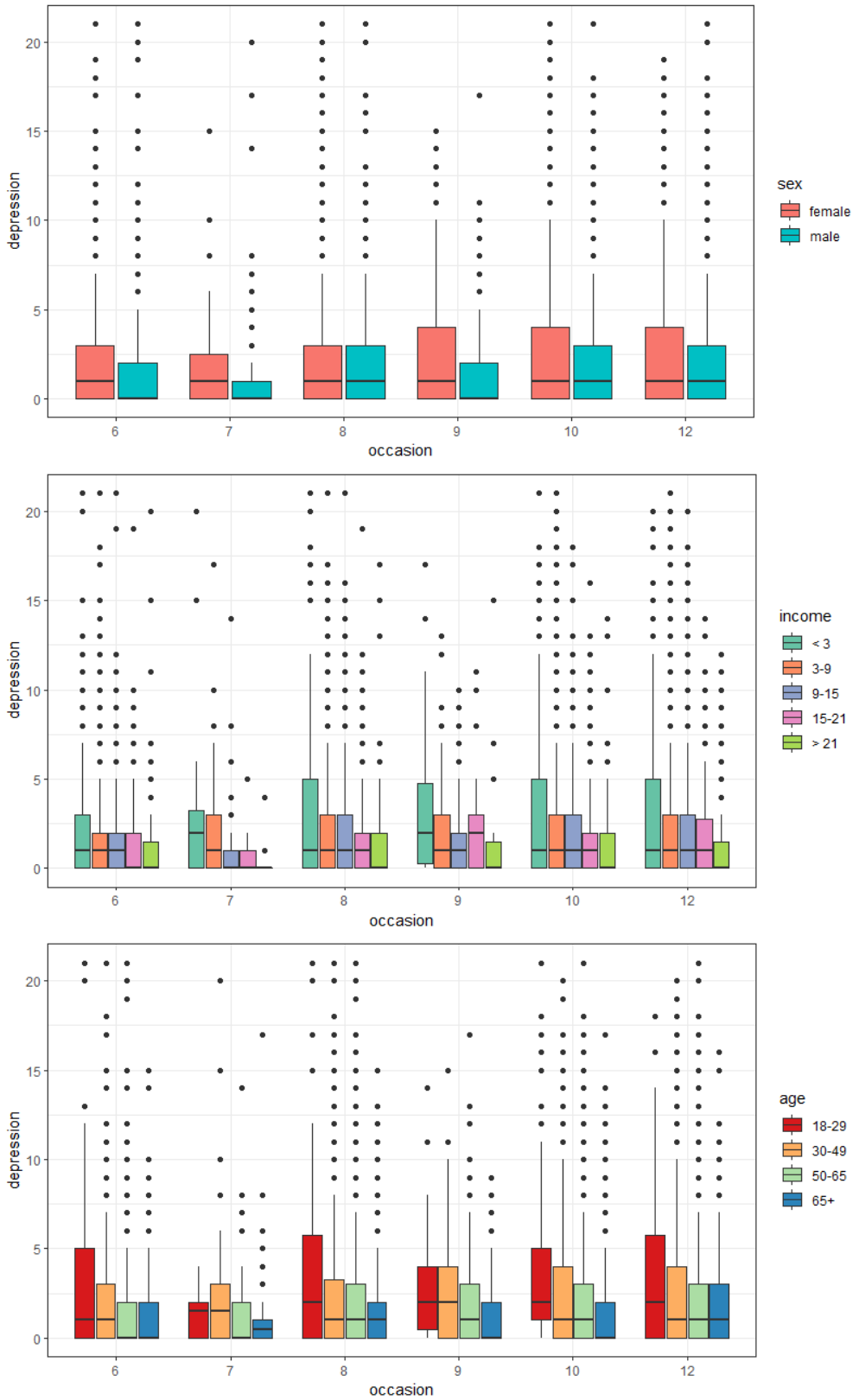

Figure 5.5: Boxplot of depression by sex (top panel), income (central) and age (bottom).

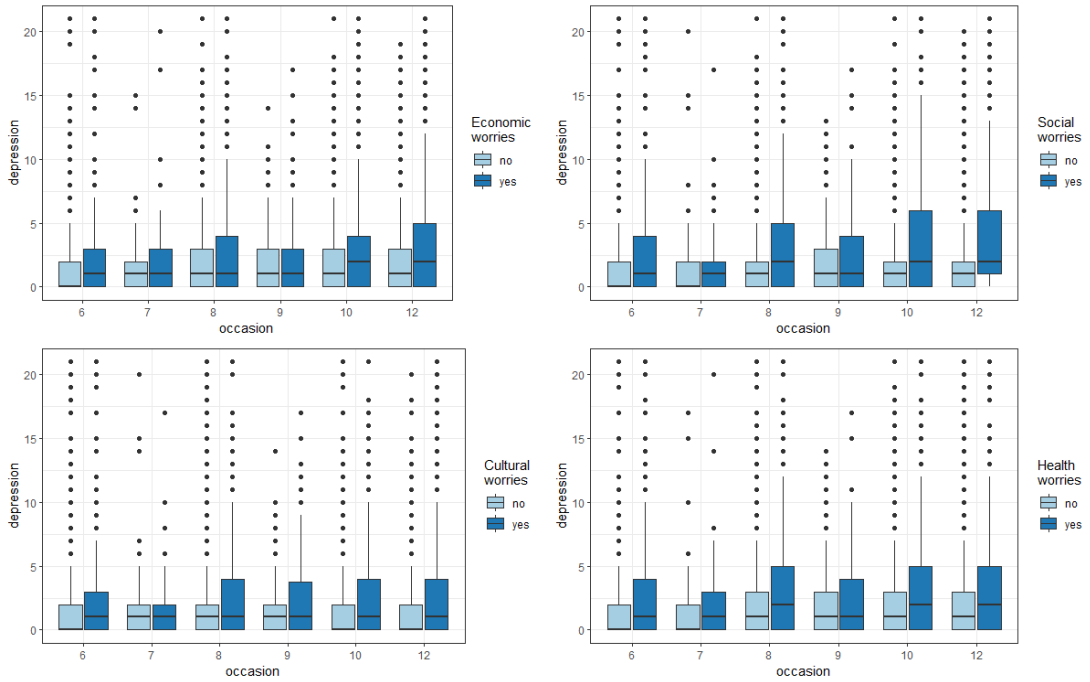

Figure 5.6: Depression scores by kind of worries.

VanderWeele and Hernán (2012) discuss how to interpret non-manipulable exposures from a physical point of view, without using the traditional notion of intervention. They exploit the fact that gender is randomised at birth, and show that causal effects can be defined in terms of physical laws. Income cannot be assumed to be randomised, then we should include in the models as many variables as possible that should make the exposure-mediator and exposure-outcome relationships unconfounded.

Income is very different from sex and race and, in particular, the pandemic produced a sort of natural intervention on it, since in many countries, such as Switzerland, several workers accepted a layoff scheme or lost their job, and this reduced their monthly income. On the other hand, the Swiss government implemented several measures to contain the economic impact of Covid-19. In particular, it provided subsidies to employees who lost their jobs due to the pandemic, giving them a monetary compensation of up to 196 CHF per day[1]. We then believe that income is an interesting exposure to analyse in causal terms.

[1] https://www2.deloitte.com/ch/en/pages/legal/articles/the-measures-being-used-to-mitigate-the-social-and-economic-impact-of-covid-19.html

5.4.2 The separability of income

Having shown that interventions on income are meaningful, we can now move to discuss the notion of separability for such a variable. In a setting as that hypothesised, where income affects depression directly and indirectly through worries, separating income into two components implies that one of them affects the degree of worry but not the depression, and the other one influences depression without affecting worry. If we were to carry out a randomised four-arm trial, we should then randomise independently these two components. Concretely, how can this be accomplished?

So far, all papers addressing the separable effects approach, independently from the interest in mediational questions, have considered binary treatments (Didelez 2019*b*, Aalen et al. 2020, Stensrud et al. 2020, 2021). This assumption makes plausible and intuitive the relationship

$$P(Y(X^M = x,\, X^Y = x)) = P(Y(X = x)), \qquad x \in \{0, 1\},$$

corresponding to property P1 in Didelez (2019*b*). This equation simply says that, when a subject receives neither of the two treatments this corresponds to the case of no treatment, while, when a subject is randomised to both treatments, this is analogous to take the unique treatment X. However, when the exposure or treatment is continuous, as in the case of income, or, for example, when the treatment is a drug whose dosage may vary, property P1 is not so trivial and needs careful consideration.

Assume we want to estimate the causal effect on depression of an intervention doubling the average Basel income, i.e. of moving from 5,000 to 10,000 CHF. Translating this intervention in terms of separable components is not straightforward. We may follow the established approach for binary treatments and assume that

$$\begin{aligned} P(Y(X^M = 5,000,\, X^Y = 5,000)) &= P(Y(X = 5,000)) \\ P(Y(X^M = 10,000,\, X^Y = 10,000)) &= P(Y(X = 10,000)). \end{aligned} \tag{5.1}$$

However, in terms of analogy with a four-arm randomised trial, it may seem more natural to interpret $P(X^M = x,\, X^Y = x)$ as if the subject got a certain amount of money x to deal with his/her degree of worry and the same amount to address the depression score. For example, consider economic worries as mediator. We could conceive a (very unrealistic) intervention where the government delivers a certain amount of money x to citizens for addressing their primary needs, such as buying food and paying the bills. This should affect economic worries. On the other hand, the same amount of money x could be provided to pay psychological counselling sessions, in order to prevent the insurgence of depression or reduce its level.

Then, in the observational regime, X does not equal x, but $2x$. In other words, the most natural conceptualisation of such an intervention is the additive one. Nonetheless, even under this perspective, there is no a unique way to decompose 5,000 CHF, which, for example, can be written as $X^M = 0$, $X^Y = 5,000$ or $X^M = 2,000, X^Y = 3,000$, among other ways. It is then the researcher that should choose plausible values to assign to the different components.

Coming back to the intervention of interest, 5,000 vs 10,000 CHF, an option could be to compare $X^M = 2,000, X^Y = 3,000$ and $X^M = 4,000, X^Y = 6,000$, so that the separable direct and indirect effect are given by the contrasts

$$\begin{aligned} &P(Y(X^M = 2,000,\, X^Y = 6,000)) \quad \text{vs.} \quad P(Y(X^M = 2,000,\, X^Y = 3,000)) \\ &P(Y(X^M = 4,000,\, X^Y = 6,000)) \quad \text{vs.} \quad P(Y(X^M = 2,000,\, X^Y = 6,000)), \end{aligned}$$

respectively. This can be interpreted as an intervention shifting a monthly income of 5,000 CHF, 2,000 of which are to address worries and 3,000 for depression, to a monthly income of 10,000 CHF, where the two components of the income are doubled.

Since, as already mentioned, the separable effects approach has not yet been extended to continuous treatments/exposures, we do not know which conceptualisation is the most appropriate, both in statistical and concrete terms. For this reason, we will perform both the interventions and will compare them.

5.4.3 Model selection and estimation of effects

As already mentioned, the present analysis focuses on the period from October to December 2020, which corresponds to the fortnights from the sixth to the twelfth in the data set, excluding the eleventh.

The variable *income* is present in the data set in two different versions, both categorical: one with eight categories, the other one with five. In order to make *income* continuous, we started from the variable with eight categories and take the mean of each class' lower and upper bounds. Thus, for example, the first category, corresponding to monthly incomes lower than 3,000 CHF, was linked to a monthly income of 1,500 CHF; the second category, incomes between 3,000 and 6,000 CHF, corresponded to an income of 4,500 CHF and so on. The highest category does not have an upper bound, since it is the class of incomes higher than 21,000 CHF. We then assumed 30,000 CHF as a plausible upper bound, and this category was then linked to an income of 25,500 CHF.

We addressed all types of worries in the data set (economic, health, social and cultural) as potential mediators. In fact, information about a fifth kind of worries,

i.e. financial, were collected as well, but they were discarded from the analysis since this variable was measured only at baseline. Our interest is on the effect of income on depression mediated by a single type of worry at a time.

To estimate the separable direct and indirect effects, we used the g-formula algorithm described in the previous chapter. The first step entailed fitting the mediator and the outcome models: since both kinds of variables are time-varying, we used a mixed-effects model for each. All worries are binary categorical variables, then we modeled the probability of being worried through a binomial generalised linear mixed model (GLMM). The variable *depression* is a non-negative score: after translating this score of 0.1, we modeled it as a Gamma GLMM. We will denote income in its continuous version by X, worries by M and depression by Y.

The analysis was carried out using the statistical software R, and the package chosen to fit mixed-effects models is `glmmTMB`, an improved and more general version of `lme4`. We fitted several models, starting from the most complex including several variables and interactions, proceeding by including and excluding variables via AIC. At the end of the process, we selected the following mixed models with a random intercept for $i = 1, \dots, 3,411$, $t \in \{6, 7, 8, 9, 10, 12\}$ and $w \in$ {economic, health, social, cultural}.

$$\begin{aligned} M_{it}^w &\sim \text{Bern}(\pi_{it}^w) \\ \text{logit}(\pi_{it}^w) &= \log\left(\frac{\pi_{it}^w}{1-\pi_{it}^w}\right) = \boldsymbol{x}_i^{w\prime}\boldsymbol{\beta}^w + b_i^w \end{aligned} \tag{5.2}$$

$$\begin{aligned} Y_{it}^w &\sim \text{Gamma}(\mu_{it}^w, \nu^w) \\ \log(\mu_{it}^w) &= \boldsymbol{x}_i^{dw\prime}\boldsymbol{\gamma}^w + g_i^w \end{aligned} \tag{5.3}$$

$\boldsymbol{x}_i^w$ and $\boldsymbol{x}_i^{wd}$ are $p_i^w \times 1$ and $p_i^{dw} \times 1$ vectors of predictors for worries and depression, respectively. $\boldsymbol{\beta}^w$ and $\boldsymbol{\gamma}^w$ are vectors of fixed effects and b_i^w, g_i^w are the random deviations from the mediator and the outcome model intercepts, respectively. π_{it}^w is the probability that subject i at occasion t has worries of type w. μ_{it}^w is the expectation of depression in the model including worry w as mediator and ν^w is the inverse of the dispersion parameter.

Table 5.2 shows the predictors in $\boldsymbol{x}_i^w$ and $\boldsymbol{x}_i^{dw}$ for each w. In addition to the continuous version of income, obtained as described before, the variables included in the analysis are

- *age*, divided into four classes: 18-29 years, 30-49, 50-65 and 65+, as in the last

Table 5.2: Variables in the linear predictors for each worry type. The first column refers to variables in the mediator models, the second to variables in the outcome model conditioning on a single type of worry (the mediator) at a time.

Worries	$\boldsymbol{x}_i^w$	$\boldsymbol{x}_i^{dw}$
Economic	*income, age, occasion* *depression_past*, *worries_econ_lag1*	*income, sex worries_econ,* *worries_econ_lag1, canton* *age*depression_past, occasion*
Health	*income, age, sex*, *susc_covid*occasion*, *worries_health_lag1*	*income, sex, worries_health*, *worries_health_lag1, canton* *age*depression_past, occasion*
Social	*income, sex* *occasion* *worries_social_lag1*	*income, sex worries_social*, *worries_social_lag1, canton* *age*depression_past, occasion*
Cultural	*income, age* *occasion*worries_cultural_lag1*, *canton*	*income, sex worries_cultural*, *worries_cultural_lag1, occasion*, *age, depression_past, canton, susc_covid*

panel of Figure 5.5;

- *sex*, a factor with three categories, male, female and other. Since there are only few subjects in the last category, we limited ourselves to consider the first two. Female is the baseline category.
- *canton* is the canton of residence, BS or BL, the reference category is BL;
- *susc_covid* is a binary variable, it is 'yes' when the subject has comorbidities making him/her more susceptible to Covid-19 (diabetes, cancer, hypertension, obesity...), 'no' otherwise;
- *depression_past* is a baseline variable and is the DASS-21 depression score referring to the previous year;
- all variables ending in 'lag1' are worries lagged of one fortnight, i.e. worries at the previous time occasion.

The selected variables constitute almost all of the variables collected from the surveys. We excluded only variables that did not result significant, like, for example, subjects' education level, their nationality, and their living status (alone or with some relatives).

Table 5.3: Results of the mediator and the outcome models for each type of worry. Stars refer to p-values of coefficients: *** 0, ** 0.01, * 0.05, $^{(.)}$ 0.1.

	Economic		**Health**		**Social**		**Cultural**	
	M	Y	M	Y	M	Y	M	Y
income	-0.03***	-0.01*	-0.02***	-0.01**	-0.03***	-0.01*	-0.01*	-0.01**
worry	–	0.23***	–	0.31***	–	0.42***	–	0.25***
worry_lag1	2.58***	0.12***	2.83***	0.07*	2.66***	0.18***	1.38***	0.17***
sex	–	-0.18***	-0.14*	-0.18***	-0.12*	-0.18***	–	-0.18***
age 30-49	0.13	-0.59***	-0.02	-0.57***	–	-0.57***	0.04	-0.26*
age 50-65	0.24$^{(.)}$	-0.79***	-0.31*	-0.74***	–	-0.73***	0.15	-0.39***
age 65+	0.28*	-0.96***	-0.18	-0.91***	–	-0.90***	0.21$^{(.)}$	-0.49***
canton	–	0.15**	–	0.14**	–	0.14**	0.23***	0.13**
susc_covid	–	–	0.51***	–	–	–	–	0.08
occasion	-0.02$^{(.)}$	0.02***	0.06***	0.02***	0.04*	0.02***	-0.04*	0.02***
depression_past	0.03**	0.11***	-	0.10***	–	0.11***	–	0.26***
age 30-49:depression_past	–	0.12***	–	0.12***	–	0.12***	–	–
age 50-65:depression_past	–	0.18***	–	0.19***	–	0.17***	–	–
age 65+:depression_past	–	0.28***	–	0.28***	–	0.28***	–	–
worry_lag1:susc_covid	–	–	-0.32**	–	–	–	–	–
worry_lag1:occasion	–	–	–	–	–	–	0.11***	–
Intercept	-1.71**	0.18	-2.49***	0.22	-2.23***	0.16	-1.10***	-0.28*
SD random effects	0.001	1.24	0.002	1.24	0.14	1.21	0.01	1.24

5.4.4 Results

As mentioned earlier, our aim is to estimate the separable direct and indirect effects of an intervention changing monthly income from 5,000 to 10,000 CHF, and another intervention changing income from 0 to 5,000 CHF. We first fitted the models in Equations (5.2)-(5.3) with predictors specified as in Table 5.2. The parameter estimates obtained from these models are used to simulate $S = 10,000$ pseudo-subjects to estimate the separable mediational effects through the g-formula (see the algorithm in Section 4.5). Standard errors and confidence intervals are obtained non-parametrically via 100 bootstrap iterations.

In Chapter 4, we have discussed the assumptions for making separable effects identifiable when applied to a mixed-effect model. We made the untestable assumption that the mediator and the outcome random intercepts are uncorrelated, in order to estimate the effects of interest. With this assumption, **A1** and **A2** are satisfied, since the graphs corresponding to the selected models are basically a simplified version of Figure 4.2, with the inclusion of baseline covariates.

The results of regression models are shown in Table 5.3. Income has a significant negative effect on each type of worry and on depression, then having a higher monthly income reduces the probability of having any kind of worries and the average score

of depression. All worries show autoregressive effects, then being worried at occasion t increases the probability of being worried at the subsequent occasion; as regards depression, the effect of worries is both instantaneous and lagged, positive in either case. Being male seems associated to lower depression scores, and men are less likely to be worried about health and social life. The older the subjects, the lower the depression score (in comparison to younger individuals), but middle-aged and old people seem more likely to have economic and health worries than youngsters, while age influences only slightly the probability of having cultural worries. The canton of residence seems not to have any effect on worries, but people living in Basel city have higher depression scores. The passage of time is associated to a slight increase in depression scores and, as expected, the depression score in the previous year is positively associated to the current score. It is interesting to notice that there is an interaction between age and the past depression score, whose effect is positive: this means that, although the elderly have lower depression scores than young individuals, among subjects used to have higher depression levels in the past, the older ones show scores higher than young subjects. Finally, having other comorbidities and having health worries at the previous occasion seems (surprisingly) to reduce the probability of having health worries compared to those not having comorbidities. In the depression model considering cultural worries as mediator, subjects being worried at the previous time have higher depression scores over time.

Let us move to the estimation of effects. We have already discussed the controversies linked to a continuous exposure and the different ways to decompose income. In this chapter, we will discuss the effects of two interventions using the decomposition closer to the original spirit of separable effects: an intervention doubling the average monthly income of a Basel citizen, from 5,000 CHF to 10,000 CHF, and an intervention changing a null income to an average monthly income, i.e. moving it from 0 to 5,000 CHF. In the former case we estimated

$$SDE = \mathbb{E}[Y_t(X^M = 5,000, X^Y = 10,000)] - \mathbb{E}[Y_t(X^M = 5,000, X^Y = 5,000)]$$
$$SIE = \mathbb{E}[Y_t(X^M = 10,000, X^Y = 10,000)] - \mathbb{E}[Y_t(X^M = 5,000, X^Y = 10,000)],$$

in the latter

$$SDE = \mathbb{E}[Y_t(X^M = 0, X^Y = 5,000)] - \mathbb{E}[Y_t(X^M = 0, X^Y = 0)]$$
$$SIE = \mathbb{E}[Y_t(X^M = 5,000, X^Y = 5,000)] - \mathbb{E}[Y_t(X^M = 0, X^Y = 5,000)],$$

for any occasion/fortnight t, i.e. we assume that (5.1) holds. Nonetheless, we considered also the alternative decomposition discussed in Section 5.4.2 and the results are

Table 5.4: Point estimates, standard errors and confidence intervals of separable direct and indirect effects of income on depression mediated by one kind of worry at a time, under an intervention changing income from 5,000 CHF to 10,000 CHF.

Worries	**Effect**	**Fortnight**		
		8	10	12
Economic	SDE	**-0.105** (0.034)	**-0.110** (0.035)	**-0.113** (0.036)
		(-0.174, -0.045)	(-0.177, -0.045)	(-0.183, -0.047)
	SIE	**-0.022** (0.008)	**-0.023** (0.007)	**-0.024** (0.007)
		(-0.037, -0.008)	(-0.038, -0.011)	(-0.037, -0.012)
Health	SDE	**-0.104** (0.036)	**-0.109** (0.036)	**-0.112** (0.037)
		(-0.173, -0.043)	(-0.179, -0.041)	(-0.186, -0.039)
	SIE	-0.012 (0.008)	-0.012 (0.007)	-0.013 (0.008)
		(-0.027, 0.002)	(-0.027, 0.003)	(-0.030, 0.003)
Social	SDE	**-0.107** (0.035)	**-0.112** (0.035)	**-0.117** (0.037)
		(-0.181, -0.043)	(-0.189, -0.054)	(-0.200, -0.053)
	SIE	-0.018 (0.012)	**-0.023** (0.010)	**-0.026** (0.010)
		(-0.037, 0.003)	(-0.043, -0.005)	(-0.043, -0.004)
Cultural	SDE	**-0.122** (0.035)	**-0.124** (0.037)	**-0.129** (0.038)
		(-0.202, -0.064)	(-0.205, -0.063)	(-0.214, -0.065)
	SIE	-0.007 (0.008)	-0.007 (0.007)	-0.007 (0.005)
		(-0.023, 0.008)	(-0.021, 0.005)	(-0.016, 0.002)

reported in Appendix C.

Table 5.4 shows the separable effects of the first intervention, 5,000 vs 10,000 CHF. It is worth remarking that the only estimable effects refers to occasions 8, 10 and 12, since the other ones (6, 7 and 9) are baseline occasions. In other words, 6, 7 and 9 are occasions at which some subjects are assessed for the first time; for this reason, worries at the previous time are missing for these subjects and the g-formula algorithm cannot estimate the expected outcomes and mediators in these occasions. Notice that this issue arises because worries are binary variables: had they been continuous, it would have been sufficient to set the value of lagged worries to 0.

The direct effect is always significant and negative, then the average depression score if all subjects had a monthly income of 10,000 CHF is lower than the average score obtained if all subjects had a 5,000 CHF income. Only economic and social worries seem to have a mediating role, and their indirect effects are negative as well. This is consistent with the regression models results, since income has a negative

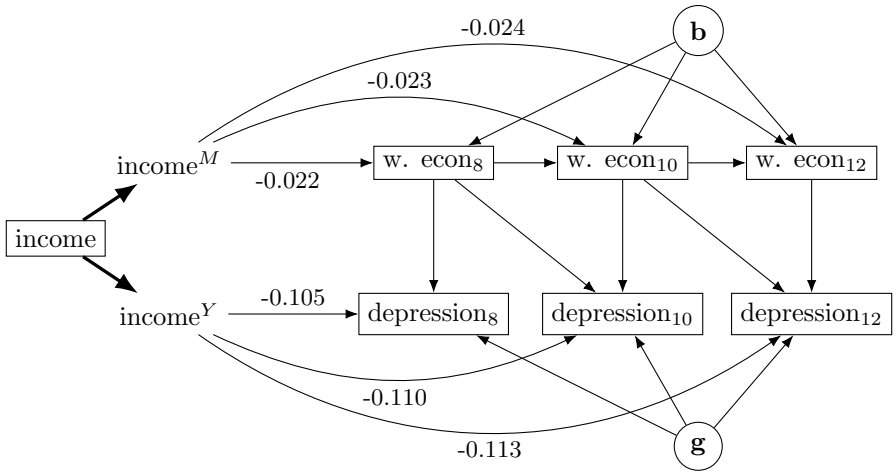

Figure 5.7: Graphical representation of direct and indirect effects in the model having economic worries as mediator considering the intervention changing income from 5,000 to 10,000 CHF. Covariates are not included to avoid clutter.

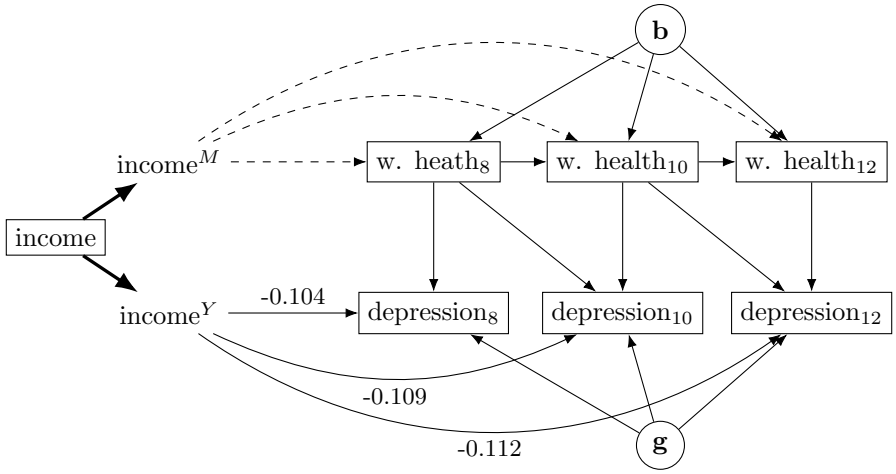

Figure 5.8: Graphical representation of direct and indirect effects in the model having health worries as mediator considering the intervention changing income from 5,000 to 10,000 CHF. Dashed lines represent non significant effects. Covariates are not included to avoid clutter.

effect on worries, which in turn affects depression positively.

Figures from 5.7 to 5.10 offer a schematic representation of these effects for each of the mediation models considered. Notice that effects appear near the arrows departing from the components of income just to make the representation as simple and clear as possible. As discussed in Chapter 4, the separable direct and indirect effects derive from all the paths starting from the Y and the M component of the exposure, respectively. Then, the indirect effect of income on depression at occasion 10 includes both the effects conveyed by worries at occasion 8 and worries at occasion 10. Analogously, the indirect effect on depression at the last occasion involves the effects transmitted by worries at each occasion.

Table 5.5 shows the effects of the second intervention, i.e. 0 vs 5,000 CHF. The

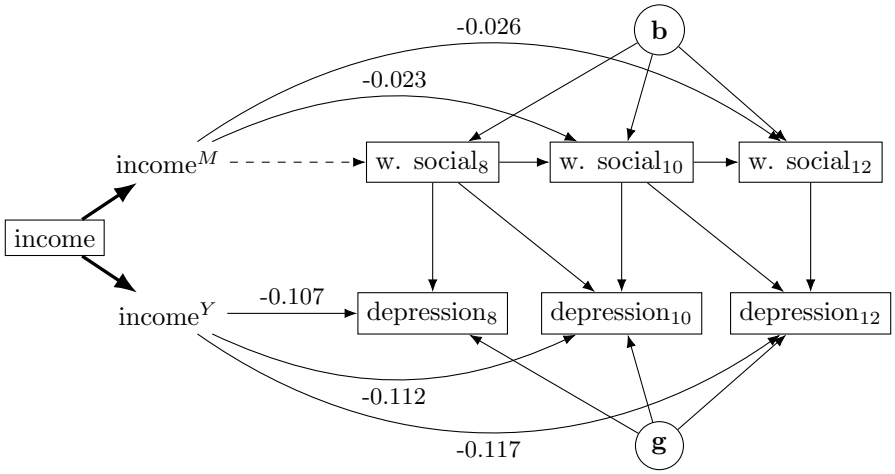

Figure 5.9: Graphical representation of direct and indirect effects in the model having social worries as mediator considering the intervention changing income from 5,000 to 10,000 CHF. Covariates are not included to avoid clutter.

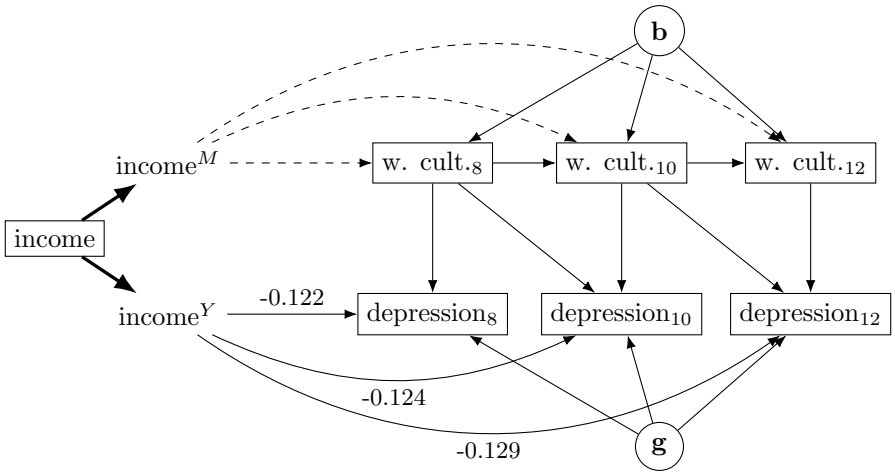

Figure 5.10: Graphical representation of direct and indirect effects in the model having cultural worries as mediator considering the intervention changing income from 5,000 to 10,000 CHF. Dashed lines represent non significant effects. Covariates are not included to avoid clutter.

Table 5.5: Point estimates, standard errors and confidence intervals of separable direct and indirect effects of income on depression mediated by one kind of worry at a time, under an intervention changing income from 0 to 5,000 CHF.

Worries	**Effect**	**Fortnight**		
		8	10	12
Economic	SDE	**-0.111** (0.041)	**-0.117** (0.043)	**-0.120** (0.044)
		(-0.210, -0.039)	(-0.203, -0.042)	(-0.215, -0.050)
	SIE	**-0.025** (0.009)	**-0.026** (0.008)	**-0.026** (0.008)
		(-0.046, -0.009)	(-0.040, -0.013)	(-0.042, -0.013)
Health	SDE	**-0.115** (0.045)	**-0.120** (0.047)	**-0.123** (0.049)
		(-0.201, -0.035)	(-0.204, -0.035)	(-0.213, -0.029)
	SIE	-0.013 (0.008)	**-0.014** (0.008)	**-0.015** (0.008)
		(-0.027, 0.002)	(-0.030, -0.002)	(-0.029, -0.002)
Social	SDE	**-0.111** (0.036)	**-0.117** (0.037)	**-0.123** (0.036)
		(-0.182, -0.045)	(-0.194, -0.043)	(-0.189, -0.055)
	SIE	-0.021 (0.013)	**-0.027** (0.011)	**-0.030** (0.011)
		(-0.045, 0.005)	(-0.047, -0.009)	(-0.053, -0.012)
Cultural	SDE	**-0.129** (0.043)	**-0.132** (0.044)	**-0.139** (0.046)
		(-0.252, -0.053)	(-0.241, -0.062)	(-0.257, -0.073)
	SIE	-0.008 (0.019)	-0.006 (0.018)	-0.007 (0.019)
		(-0.025, 0.008)	(-0.019, 0.008)	(-0.017, 0.006)

effects show approximately the same pattern as the previous ones, but with two differences: first, the magnitude of direct effects is bigger (in absolute terms), and this is particularly evident for cultural worries; second, health worries assume a mediating role at occasions 10 and 12. This is interesting, since it seems to suggest that what matters in an intervention on income is not just the magnitude, since in both interventions considered there is a 5,000 CHF increase, but also the 'position' of the intervention. Apparently, an increase of 5,000 CHF in monthly income in two different settings produces a difference in terms of the indirect effect of health worries. For both interventions, it is also possible to notice a slight increase, in absolute values, in the magnitude of both direct and indirect effects over time.

5.5 Conclusions

In this chapter, we analysed data from the COVCO-Basel study, investigating whether, during Covid-19, the relationship between income and depression in a group of Swiss subjects is only direct or it is also mediated by worries concerning different aspects of life.

The variables of interest cannot be assumed to follow a Normal distribution, therefore we could not use the structural and multilevel models discussed in Chapter 3. We used instead the separable effects approach applied to mixed-effect models, as described in Chapter 4. This analysis poses conceptual and practical challenges, due to the complex nature of the sample design and of the variables collected. Subjects entered the study at different time occasions, and it was impossible to estimate mediational effects at the baseline occasions due to the nature of the mediators. In addition, the continuous version of income, not present among the collected variables, was used as exposure and the conceptual difficulties connected to its 'separability' were discussed.

The results, relative to occasions 8, 10 and 12, i.e. to the period between the end of October and the end of December 2020, show that the direct effect of income on depression is always negative and significant, thus, if the monthly income of all participants increased of 5,000 CHF, they would have a lower depression score. Economic and social worries have a mediating role for both the interventions described, while health worries play the role of mediator only when monthly income moves from 0 to 5,000 CHF.

This analysis can offer interesting clues on the evolution of mental health during the Covid-19 pandemic and its tight relationship with socio-economic status. We focused only on depression, but the study of anxiety and stress could give us other insights on this issue. An analysis with multiple outcomes might show relationships which are ignored when the focus is on a single outcome, and this call for a development of SEMs to nonlinear cases.

Clearly, our approach is not free from limitations. First, we did not have information about the real income of each subject in a numeric form, and we had to derive such a variable from an 8-category version of income. This is an approximation which undoubtedly had an impact on the results of our analysis. Second, the analysis concerns only Basel citizens: an interesting direction for future research may be to include subjects from other Swiss cantons and analyse the relationship between mental health and the canton epidemiological situation. Third, it is worth remarking that the number of covariates included in the model is relatively small, and this

can cast some doubts on the plausibility of assumptions **A1**, **A2** and **A3**. This is a problem affecting any observational study and can be attenuated only conducting sensitivity analyses. Stensrud et al. (2021) proposed a type of sensitivity analysis for separable effects in a competing risk setting, and extending it to models with random coefficients is a challenging future direction. As already remarked, we included almost all the variables collected in the sample, but our analysis could definitely be improved by including more variables, especially time-varying covariates. Finally, we have assumed that the latent structure underlying the data involves just random intercepts; however, as we have seen so far, many other latent structures are possible, but unfortunately to date their employment seems limited to linear Normal settings. All these issues pave the way for future work.

Chapter 6

Conclusions and future directions

Time governs our life, and it is an essential aspect of any mediation analysis. However, integrating the passage of time into mediational models can be done in several ways, differing between associational models and causal approaches. In this dissertation, we addressed some of these differences and proposed unifying perspectives by mean of new methodological approaches.

In Chapter 2, we provided a comprehensive literature review on longitudinal mediation analysis, discussing the two main families of approaches proposed to address this topic. The passage of time is linked to the concept of change and, in SEM, the several dynamics of change are formalised through different latent structures, the most popular of which were discussed in Section 2.2. Mixed-effect models present a latent structure as well, but it does not entail an hypothesis about the type of change. It captures instead the heterogeneity among level-2 units, indeed the use of these models is not restricted to longitudinal settings. The main focus of causal approaches is instead on estimands, which are inextricably linked to causal models and to the concept of intervention underlying the causal effects.

These features may make all these models/approaches appear as opposite, and this is quite natural, since they stem from very different traditions and backgrounds. In this thesis, our aim was to combine some of these approaches, as shown in Chapters 3 and 4. The former gives two contributions: a unification of structural and multilevel mediational models and a new perspective on multilevel mediation models within this unified framework, which was demonstrated to overcome the main limitations of the traditional multilevel setting. The latter provides formal assumptions to endow mixed-effect and latent growth models with a causal interpretation, by means of the separable effects approach. We also derived formulas for identifying these mediational effects, proving that they coincide with those from path analysis in the case of linearity.

Finally, in the fifth chapter we analysed real data collected in Switzerland between July 2020 and August 2021. We investigated the direct effects of income on depression and its indirect effects through different type of worries, during the last three months of 2020 corresponding to a period of Covid-19 high spread. Although this analysis was mainly intended as an example showing a practical application of some of the theoretical approaches proposed in the thesis, nonetheless it posed conceptual and practical challenges very interesting to address.

This dissertation addresses longitudinal mediation analysis from both an associational and a causal perspective, overcoming the traditional dichotomy recognisable in the literature on the topic. In particular, Chapter 4 poses the first brick of a bridge connecting latent variable models with causal inference. As already discussed in Chapter 2, latent variables are generally a source of bias in the causal literature, and SEM are not the most widely used models. We hope that this work can be the beginning of a new stream of literature increasingly unifying SEMs and causality.

Despite the contributions of this thesis, it is worth acknowledging some limitations. The literature review, although rich in terms of models and approaches described, focuses on model specification and does not address inference. This issue would deserve an entire chapter, given the variety of approaches proposed to estimate latent variable models and causal effects, especially in the presence of time-varying confounders. In addition, we limited ourselves to considering static treatments: dynamic ones are much more complex and they still are the object of an active branch of research.

In Chapter 3, the unification proposed clearly holds only for linear models. This is a disadvantage for multilevel models, which have been extended to non-Normal outcomes and nonlinearities through the GLMM theory. It should also be noticed that the analogy between LGMs and mixed-effect models holds only for certain types of mixed-effect models. For example, expressing the one depicted in Figure 4.2 in Chapter 4 in SEM terms would be definitely not so straightforward, and the framework developed would probably need to be extended. We focused on the formalisation of the unified model and the different types of multilevel settings which can be embodied within this framework, but we neglected the aspects concerning identifiability (in the SEM terms) and inference, although the latter was briefly addressed at the end of the chapter. Finally, we ignored issues related to centering, which have long been discussed in the multilevel literature. This is an extremely complex topic, still debated among scholars, and we believe that a thorough discussion would have been beyond the scope of this dissertation.

We have thoroughly discussed difficulties related to the separable effects approach. We restricted our discussion to binary treatments, since, as demonstrated by the

empirical analysis, continuous treatments are definitely more challenging, from both a conceptual and a practical point of view. The main issue connected to separable effects applied to latent variable models is that the estimation formulas rely on a correct model specification. Since it is impossible to know which is the true latent structure underlying some data, separable mediational effects seem to be damned to be considered at least questionable. This call for the development of a sensitivity analysis, whose theoretical implementation is yet far from being completed.

The data analysis was made difficult from the complex sample design: subjects entered the study at different points in time and some relevant variables, such as a continuous version of income, were not collected. This analysis can be improved in several ways: using more complex models, including more detailed variables about subjects' income and the epidemiological situation in Switzerland, as well as worries. In particular, a continuous or at least numeric version of worries could allow us to estimate separable effects for each time occasion.

All the aforementioned issues give us many hints on possible future research directions. As regards the literature review, inferential methods to estimate SEM and multilevel model are well known, while a systematic discussion of causal estimation methods in the presence of mediation is yet to be provided. Some reviews of inferential approaches to deal with time-varying confounders were already published (Daniel et al. 2013, Vansteelandt and Joffe 2014, Clare et al. 2019), but they do not focus specifically on mediation. Other interesting topics to cover are dynamic treatments and continuous-time mediation analysis, both of which are receiving increasingly attention from scholars.

Moving to Chapter 3, there are numerous aspects to deepen or explore. We restricted our discussion to linear models, exploiting the traditional theory underlying SEM, i.e. that based on the multivariate Normal distribution of the variables. This makes the estimation of effects quite natural via path analysis, but restricts the types of variable that can be modelled, which have at least to be continuous. Although there exist approaches to deal with binary variables in the SEM framework and they are implemented in statistical software such as Mplus and `OpenMx`, they rely anyway on Normal approximations. In addition, nonlinear path analysis is not yet well developed, and therefore a formal definition of direct and indirect effects in such settings is missing. Concerning multilevel SEM, different aspects need further investigation, such as identification criteria, the development of goodness-of-fit measures and hypothesis testing.

The separable effects approach is quite recent, and it has been applied to longitudinal mediational settings even more recently. This approach is experiencing a rapid

development and has yet to reveal its full potential. In this thesis, it was applied to mixed-effect and latent growth models, but our conjecture is that it can be applied to other classes of latent variable models with appropriate assumptions. Adopting this perspective presents several advantages, but it is in any case subject to a sound substantive (and non testable) theory justifying the separability of exposure and the components' meaningfulness. Moreover, as proved by our empirical application, the extension of this approach to non-binary exposures is not straightforward and it is for sure an issue which deserves attention in the future.

Longitudinal mediation analysis has become a very active research area over the last decades. Its applications are almost unlimited, since most phenomena of interest in different fields evolve over time and longitudinal mediation analysis contributes to unveil how the change takes place. We believe that our thesis can suggest a direction toward a more and more tight connection of latent variable models and causal approaches, which may exploit the strengths of both, and lead to increasingly accurate insights into the real world surrounding us.

Appendix A

A note on causal interpretation of associational models

At the end of Chapter 2, we claimed that associational models can be endowed with a causal interpretation under some assumptions, which also ensure identifiability of mediational parameters.

For example, suppose to fit a CLPM with observed variables as in Equations (2.1)-(2.3). If one is interested in natural effects, comparing two static regimes, the assumptions for identifiability are

C1 *Consistency*: if the exposure history $\overline{X}$ takes value $\bar{x}$, then the potential outcome $M_t(\bar{x})$ equals the observed M_t. Analogously, if, in addition, mediator history is set to $\overline{m}$, then $Y_t(\bar{x}, \overline{m})$ equals the observed outcome Y_t.

C2 *Exchangeability*:

- C2.1 $Y_{t'}(\bar{x}, \overline{m}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1}, \overline{M}_{t-1}, \overline{Y}_{t-1}$ for each $t' > t$ (no exposure-outcome unobserved confounders)
- C2.2 $Y_{t'}(\bar{x}, \overline{m}) \perp\!\!\!\perp M_t \,|\, \overline{X}_{t-1}, \overline{M}_{t-1}, \overline{Y}_{t-1}$ for each $t' > t$ (no mediator-outcome unobserved confounders)
- C2.3 $M_{t'}(\bar{x}) \perp\!\!\!\perp X_t \,|\, \overline{X}_{t-1}, \overline{M}_{t-1}, \overline{Y}_{t-1}$ for each $t' > t$ (no exposure-mediator unobserved confounders)
- C2.4 $Y_{t'}(\bar{x}, \overline{m}) \perp\!\!\!\perp M_t(\bar{x}^*)$ for each $t' > t$ (no mediator-outcome observed and unobserved confounders affected by the exposure)

where we suppress subject subscript to make notation clearer and we assume that variables with a null or negative subscript are not defined. Under these assumptions, one can derive formulas to compute natural direct and indirect effects from

observational data. In particular, the natural direct effect is given by

$$\begin{aligned}\mathbb{E}[Y_t(\bar{x}, \overline{M}(\bar{x}^*)) - Y_t(\bar{x}^*, \overline{M}(\bar{x}^*))] = \\ \sum_{\overline{m}_{t-1}, \bar{y}_{t-1}} \left\{\mathbb{E}[Y_t \,|\, \bar{x}_{t-1}, \overline{m}_{t-1}, \bar{y}_{t-1}] - \mathbb{E}[Y_t \,|\, \bar{x}^*_{t-1}, \overline{m}_{t-1}, \bar{y}_{t-1}]\right\} \\ \prod_{k=1}^{t-1} \left\{P(Y_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}, \bar{y}_{k-1}) - P(Y_k \,|\, \bar{x}^*_{k-1}, \overline{m}_{k-1}, \bar{y}_{k-1})\right\} P(M_k \,|\, \bar{x}^*_{k-1}, \overline{m}_{k-1}),\end{aligned} \tag{A.1}$$

and the natural indirect effect by

$$\begin{aligned}\mathbb{E}[Y_t(\bar{x}, \overline{M}(\bar{x})) - Y_t(\bar{x}, \overline{M}(\bar{x}^*))] = \\ \sum_{\overline{m}_{t-1}, \bar{y}_{t-1}} \mathbb{E}[Y_t \,|\, \bar{x}_{t-1}, \overline{m}_{t-1}, \bar{y}_{t-1}] \prod_{k=1}^{t-1} P(Y_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}, \bar{y}_{k-1}) \times \\ \left\{P(M_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}) - P(M_k \,|\, \bar{x}^*_{k-1}, \overline{m}_{k-1})\right\}.\end{aligned} \tag{A.2}$$

To prove it, it is sufficient to estimate $\mathbb{E}[Y_t(\bar{x}, \overline{M}(x^*))]$, which, under assumptions C1-C2, is identified as

$$\begin{aligned}\mathbb{E}[Y_t(\bar{x}, \overline{M}(\bar{x}^*))] = \\ \sum_{\overline{m}_{t-1}, \bar{y}_{t-1}} \mathbb{E}[Y_t \,|\, \bar{x}_{t-1}, \overline{m}_{t-1}, \bar{y}_{t-1}] \prod_{k=1}^{t-1} P(Y_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}, \bar{y}_{k-1}) P(M_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}).\end{aligned} \tag{A.3}$$

Proof. To give the intuition behind the proof, let us start by considering three waves of data, $T = 3$. Then, we have

$$\begin{aligned}&\mathbb{E}[Y_3(\bar{x}_2, \overline{M}_2(x^*_1))] \\ &= \sum_{m_1} \mathbb{E}[Y_3(\bar{x}_2, m_1, M_2(x^*_1)) \,|\, M_1(x^*_1) = m_1]\, P(M_1(x^*_1) = m_1) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{by iterated expectations} \\ &= \sum_{m_1} \mathbb{E}[Y_3(\bar{x}_2, m_1, M_2(x^*_1))]\, P(M_1 = m_1) \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{by C2.4 and } M_1(x^*_1) = M_1\text{, since } M_1 \text{ is not affected by } X_1 \\ &= \sum_{m_1} \mathbb{E}[Y_3(\bar{x}_2, m_1, M_2(x^*_1)) \,|\, x_1]\, P(M_1 = m_1) \qquad\qquad \text{by C2.1}\end{aligned}$$

$$= \sum_{m_1} \mathbb{E}[Y_3(\bar{x}_2, m_1, M_2(x_1^*)) \,|\, x_1, m_1]\, P(M_1 = m_1) \qquad \text{by C2.2}$$

$$= \sum_{m_1, y_1} \mathbb{E}[Y_3(\bar{x}_2, m_1, M_2(x_1^*)) \,|\, x_1, m_1, y_1]\, P(Y_1 = y_1)\, P(M_1 = m_1)$$

by iterated expectations

$$= \sum_{m_1, m_2, y_1} \mathbb{E}[Y_3(\bar{x}_2, \overline{m}_2) \,|\, x_1, x_2, m_1, M_2(x_1^*) = m_2, y_1]\, P(M_2(x_1^*) = m_2 \mid x_1^*, m_1) \times$$

$$P(Y_1 = y_1)\, P(M_1 = m_1) \qquad \text{by iterated expectations, C2.1 and C2.3}$$

$$= \sum_{m_1, m_2, y_1} \mathbb{E}[Y_3(\bar{x}_2, \overline{m}_2) \,|\, x_1, x_2, m_1, m_2, y_1]\, P(M_2(x_1^*) = m_2 \,|\, x_1^*, m_1) \times$$

$$P(Y_1 = y_1)\, P(M_1 = m_1) \qquad \text{by C2.4 and C2.2}$$

$$= \sum_{m_1, m_2, y_1, y_2} \mathbb{E}[Y_3(\bar{x}_2, \overline{m}_2) \,|\, x_1, x_2, m_1, m_2, y_1, y_2]\, P(Y_2(x_1, m_1) = y_2 \,|\, x_1, m_1, y_1) \times$$

$$P(M_2(x_1^*) = m_2 \,|\, x_1^*, m_1)\, P(Y_1 = y_1)\, P(M_1 = m_1)$$

by iterated expectations, C2.1 and consistency

$$= \sum_{\overline{m}_2, \bar{y}_2} \mathbb{E}[Y_3(\bar{x}_2, \overline{m}_2) \,|\, \bar{x}_2, \overline{m}_2, \bar{y}_2]\, P(Y_2(x_1, m_1) = y_2 \,|\, x_1, m_1, y_1) \times$$

$$P(M_2(x_1^*) = m_2 \,|\, x_1^*, m_1)\, P(Y_1 = y_1)\, P(M_1 = m_1)$$

$$= \sum_{\overline{m}_2, \bar{y}_2} \mathbb{E}[Y_3 \,|\, \bar{x}_2, \overline{m}_2, \bar{y}_2]\, P(Y_2 = y_2 \,|\, x_1, m_1, y_1)\, P(Y_1 = y_1) \times$$

$$P(M_2 = m_2 \,|\, x_1^*, m_1)\, P(M_1 = m_1) \qquad \text{by C1}$$

For $T > 3$, iterating steps shown above, one obtains

$$\mathbb{E}[Y_t(\bar{x}_{t-1}, \overline{M}_{t-1}(\bar{x}_{t-2}^*))]$$

$$= \sum_{\overline{m}_{t-1}, \bar{y}_{t-1}} \mathbb{E}[Y_t \,|\, \bar{x}_{t-1}, \overline{m}_{t-1}, \bar{y}_{t-1}] \prod_{k=1}^{t-1} P(Y_k \,|\, \bar{x}_{k-1}, \overline{m}_{k-1}, \bar{y}_{k-1}) P(M_k \,|\, \bar{x}_{k-1}^*, \overline{m}_{k-1})$$

□

Notice that here we used sums, but they would be replaced by integrals in the case of continuous variables. Formulas (A.1) and (A.2) immediately follow from (A.3) by applying the definition of natural direct and indirect effects. Then, it can be proved that, if the true model is as in (2.1)-(2.3), and data are observed at three times, then the NDE of X on Y_3 is given by $\gamma_{_X}(x_2 - x_2^*) + \gamma_{_X}\gamma_{_Y}(x_1 - x_1^*)$, which is the sum of effects conveyed by paths $X_2 \to Y_3$ and $X_1 \to Y_2 \to Y_3$; NIE is given by $\beta_{_X}\gamma_{_Y}(x_1 - x_1^*)$, corresponding to $X_1 \to M_2 \to Y_3$.

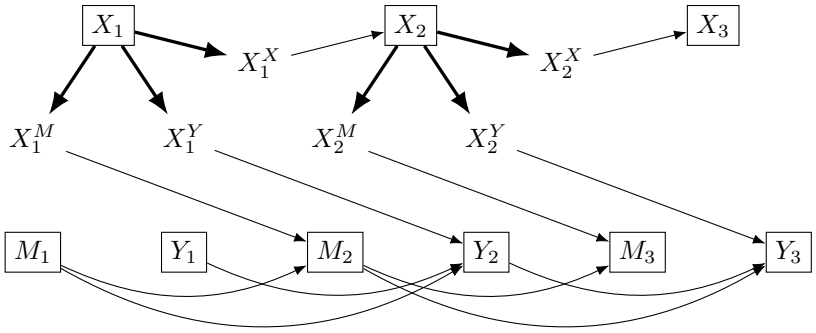

Figure A.1: Expanded graph of Figure 2.2 in the main text. Bold arrows represent deterministic relationships.

In the case of three waves, it is easy to understand along which paths direct and indirect effects propagate, but adding just one wave makes the task more complex, since the number of paths increases. For example, there may be direct effects conveyed straightforwardly from X at a certain point to Y at a subsequent time and direct effects passing through subsequent X's. Similarly, indirect effects may be conveyed by a single mediator or multiple mediators.

VanderWeele and Tchetgen Tchetgen (2017) show that, in a CLPM as that in MacKinnon (2008), interventional effects are identified, under static sequential ignorability. In absence of time-varying confounding they coincide with natural effects. Thus, in a case as that discussed so far, natural and interventional effects coincide.

Now suppose to be interested in estimating separable effects for the expanded graph shown in Figure A.1, corresponding to the path diagram in Figure 2.2 in the main text. Notice that, since each child of $\mathbf{X} = \{X_1, X_2, X_3\}$ in the original graph has exactly one component of X as parent in the expanded graph, it is an edge graph, using the same terminology as in Robins et al. (2020). Thus, using Corollary 1 in their paper, separable effects are identified.

Here we show how path specific effects can be written in terms of counterfactuals using separable components of X and that they are identified. For example, consider the graph in Figure 2.2 and the paths $\pi_1 = \{X_2 \to Y_3; X_1 \to Y_2 \to Y_3\}$ and $\pi_2 = \{X_1 \to M_2 \to Y_3\}$, which, assuming a binary exposure ($x = 1, x^* = 0$), correspond to the following counterfactuals

$$Y_3(\pi_1, x, x^*) = Y_3(X_1 = 1, X_2 = 1, Y_2(X_1 = 1), M_2(X_1 = 0))$$
$$Y_3(\pi_2, x, x^*) = Y_3(X_1 = 0, X_2 = 0, Y_2(X_1 = 0), M_2(X_1 = 1)),$$

respectively.

Notice that these counterfactuals coincide, as random variables, with $Y_3(X_1^M =$

$0, X_1^Y = 1, X_2^M = 0, X_2^Y = 1)$ and $Y_3(X_1^M = 1, X_1^Y = 0, X_2^M = 0, X_2^Y = 0)$, respectively. They are written in terms of the components of $\mathbf{X}$ shown in Figure A.1. This follows from Proposition 3 in Robins et al. (2020), since π_1 and π_2 are *edge-consistent*, i.e. they do not contain any recanting witness (Avin et al. 2005, Shpitser 2013). From Theorem 2 in Robins et al. (2020), it also follows that the two random variables are identified. Specifically,

$$\begin{aligned} Y_3(\pi_1, x, x^*) = Y_3(X_1 = 1, X_2 = 1, Y_2(X_1 = 1), M_2(X_1 = 0)) \equiv \\ Y_3(X_1^M = 0, X_1^Y = 1, X_2^M = 0, X_2^Y = 1) = \\ \sum_{m_2, y_2} P(Y_3 \,|\, y_2, m_2, x_2 = 1, x_1 = 1) P(y_2 \,|\, x_1 = 1) P(m_2 \,|\, x_1 = 0) \end{aligned}$$

$$\begin{aligned} Y_3(\pi_2, x, x^*) = Y_3(X_1 = 0, X_2 = 0, Y_2(X_1 = 0), M_2(X_1 = 1)), \equiv \\ Y_3(X_1^M = 1, X_1^Y = 0, X_2^M = 0, X_2^Y = 0) = \\ \sum_{m_2, y_2} P(Y_3 \,|\, y_2, m_2, x_2 = 0, x_1 = 0) P(y_2 \,|\, x_1 = 0) P(m_2 \,|\, x_1 = 1), \end{aligned}$$

where we marginalised over m_1 and y_1.

As regards the other SEMs discussed in the manuscript, factor CLPM, LGM and LDS, they are not non-parametrically identified, due to the presence of latent variables. We gave a detailed discussion about separable effects for LGMs in Chapter 4, showing why a non-parametric identification is impossible. We have not addressed the other classes of models, but the underlying logic is similar.

Appendix B

Proofs of formulas in Chapter 4

In this section we provide proofs of Equations (4.7) and (4.15). We consider a hypothetical randomised trial, where X^M and X^Y are randomised independently possibly to different values x^* and x, respectively.

B.1 Proof of Equation (4.7)

To prove Equation (4.7), let us start from $\mathbb{E}[Y_2(X^M = x^*, X^Y = x)]$.

$$
\begin{aligned}
&\mathbb{E}[Y_2(X^M = x^*, X^Y = x)] \\
&= \sum_{m_1, m_2, y_1} \mathbb{E}[Y_2(X^M = x^*, X^Y = x) \,|\, \overline{M}_2(X^M = x^*, X^Y = x) = \overline{m}_2, \\
&\qquad\qquad Y_1(X^M = x^*, X^Y = x) = y_1] \times \\
&\quad P(M_2(X^M = x^*, X^Y = x) = m_2 \,|\, Y_1(X^M = x^*, X^Y = x) = y_1, \\
&\qquad\qquad M_1(X^M = x^*, X^Y = x) = m_1) \times \\
&\quad P(Y_1(X^M = x^*, X^Y = x) = y_1 \,|\, M_1(X^M = x^*, X^Y = x) = m_1) \times \\
&\qquad\qquad P(M_1(X^M = x^*, X^Y = x) = m_1) \\
&= \sum_{m_1, m_2, y_1} \mathbb{E}[Y_2(X^M = x, X^Y = x) \,|\, \overline{M}_2(X^M = x, X^Y = x) = \overline{m}_2, \\
&\qquad\qquad Y_1(X^M = x, X^Y = x) = y_1] \times \\
&\quad P(M_2(X^M = x^*, X^Y = x^*) = \overline{m}_2 \,|\, Y_1(X^M = x^*, X^Y = x^*) = y_1, \\
&\qquad\qquad M_1(X^M = x^*, X^Y = x^*) = m_1) \times \\
&\quad P(Y_1(X^M = x, X^Y = x) = y_1 \,|\, M_1(X^M = x, X^Y = x) = m_1) \times \\
&\qquad\qquad P(M_1(X^M = x^*, X^Y = x^*) = m_1)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{m_1,m_2,y_1} \mathbb{E}[Y_2(X=x) \,|\, \overline{M}_2(X=x) = \overline{m}_2, Y_1(X=x) = y_1] \times \\
&\quad P(M_2(X=x^*) = m_2 \,|\, Y_1(X=x^*) = y_1, M_1(X=x^*) = m_1) \times \\
&\quad P(Y_1(X=x) = y_1 \,|\, M_1(X=x) = m_1) P(M_1(X=x^*) = m_1) \\
&= \sum_{m_1,m_2,y_1} \mathbb{E}[Y_2 \,|X=x,\, \overline{M}_2 = \overline{m}_2, Y_1 = y_1] \times \\
&\quad P(M_2 = m_2 \,|\, X=x^*, Y_1 = y_1, M_1 = m_1) \times \\
&\quad P(Y_1 = y_1 \,|\, X=x, M_1 = m_1) P(M_1 = m_1 \,|\, X=x^*)
\end{aligned}
$$

where the first equality is obtained applying the law of iterated expectations and the second equality follows from assumptions **A1** and **A2** since, for example

$$
\begin{aligned}
&P(M_1(X^M = x^*, X^Y = x) = m_1) \\
&\qquad = P(M_1 = m_1 \,|\, X^M = x^*, X^Y = x) \\
&\qquad = P(M_1 = m_1 \,|\, X^M = x^*, X^Y = x^*) \quad \text{by } \mathbf{A1} \\
&\qquad = P(M_1(X^M = x^*, X^Y = x^*) = m_1),
\end{aligned}
$$

and analogously for Y_1

$$
\begin{aligned}
&P(Y_1(X^M = x^*, X^Y = x) = y_1 \,|\, M_1(X^M = x^*,\, X^Y = x) = m_1) \\
&\qquad = P(Y_1 = y_1 \,|\, M_1 = m_1, X^M = x^*, X^Y = x) \\
&\qquad = P(Y_1 = y_1 \,|\, M_1 = m_1, X^M = x, X^Y = x) \quad \text{by } \mathbf{A2} \\
&\qquad = P(Y_1(X^M = x, X^Y = x) = y_1 \,|\, M_1(X^M = x,\, X^Y = x) = m_1).
\end{aligned}
$$

The same line of reasoning can be applied to the other terms in the formula. The third equality follows from 4.1 and the last one from the fact that

$$
\begin{aligned}
&P(Y_1(X=x) = y_1 \,|\, M_1(X=x) = m_1) \\
&\qquad = \frac{P(Y_1(X=x) = y_1, M_1(X=x) = m_1 \,|\, X=x)}{P(M_1(X=x) = m_1 \,|\, X=x)} \quad \text{by exchangeability} \\
&\qquad = \frac{P(Y_1 = y_1, M_1 = m_1 \,|\, X=x)}{P(M_1 = m_1 \,|\, X=x)} \quad \text{by consistency} \\
&\qquad = P(Y_1 = y_1 \,|\, X=x, M_1 = m_1)
\end{aligned}
$$

and this holds also for the other conditional densities appearing in the expression.

Iterating the same passages for any $t > 2$, it is easy to obtain Equation (4.7). See Appendix B of Stensrud et al. (2021) for a similar proof for a survival outcome in a

competing-event setting.

B.2 Proof of Equation (4.15)

In the following we use summation just for consistency with the notation adopted throughout Chapter 4, although we assume that the θ factors are Normally distributed and then integrals would be the correct choice.

$$
\begin{aligned}
&\mathbb{E}[Y_t(X^M = x^*, X^Y = x)] \\
&= \sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t(X^M = x^*, X^Y = x) \,|\, \theta_{1Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1Y}, \\
&\qquad \theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M} \\
&\qquad \theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}] \times \\
&\quad P(\theta_{1Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1Y} \,|\, \theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y}, \\
&\qquad \theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M}, \theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}) \times \\
&\quad P(\theta_{0Y}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0Y}) P(\theta_{1M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{1M}) \times \\
&\qquad P(\theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}) \\
&= \sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t(X^M = x, X^Y = x) \,|\, \theta_{1Y}(X^M = x, X^Y = x) = \tilde{\theta}_{1Y}, \\
&\qquad \theta_{0Y}(X^M = x, X^Y = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X^M = x, X^Y = x) = \tilde{\theta}_{1M} \\
&\qquad \theta_{0M}(X^M = x, X^Y = x) = \tilde{\theta}_{0M}] \times \\
&\quad P(\theta_{1Y}(X^M = x, X^Y = x) = \tilde{\theta}_{1Y} \,|\, \theta_{0Y}(X^M = x, X^Y = x) = \tilde{\theta}_{0Y}, \\
&\qquad \theta_{1M}(X^M = x, X^Y = x) = \tilde{\theta}_{1M}, \theta_{0M}(X^M = x, X^Y = x) = \tilde{\theta}_{0M}) \times \\
&\quad P(\theta_{0Y}(X^M = x, X^Y = x) = \tilde{\theta}_{0Y}) P(\theta_{1M}(X^M = x^*, X^Y = x^*) = \tilde{\theta}_{1M}) \times \\
&\qquad P(\theta_{0M}(X^M = x^*, X^Y = x^*) = \tilde{\theta}_{0M}) \\
&= \sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t(X = x) \,|\, \theta_{1Y}(X = x) = \tilde{\theta}_{1Y}, \theta_{0Y}(X = x) = \tilde{\theta}_{0Y}, \\
&\qquad \theta_{1M}(X = x) = \tilde{\theta}_{1M}, \theta_{0M}(X = x) = \tilde{\theta}_{0M}] \\
&\quad P(\theta_{1Y}(X = x) = \tilde{\theta}_{1Y} \,|\, \theta_{0Y}(X = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X = x) = \tilde{\theta}_{1M}, \\
&\qquad \theta_{0M}(X = x) = \tilde{\theta}_{0M}) \times \\
&\quad P(\theta_{0Y}(X = x) = \tilde{\theta}_{0Y}) P(\theta_{1M}(X = x^*) = \tilde{\theta}_{1M}) P(\theta_{0M}(X = x^*) = \tilde{\theta}_{0M}) \\
&= \sum_{\tilde{\theta}_{0M},\tilde{\theta}_{1M},\tilde{\theta}_{0Y},\tilde{\theta}_{1Y}} \mathbb{E}[Y_t \,|\, X = x,\, \theta_{1Y} = \tilde{\theta}_{1Y}, \theta_{0Y} = \tilde{\theta}_{0Y}, \theta_{1M} = \tilde{\theta}_{1M}, \theta_{0M} = \tilde{\theta}_{0M}]
\end{aligned}
$$

$$P(\theta_{1Y} = \tilde{\theta}_{1Y} \,|\, X = x, \theta_{1M} = \tilde{\theta}_{1M}, \theta_{0M} = \tilde{\theta}_{0M}) \times$$
$$P(\theta_{0Y} = \tilde{\theta}_{0Y} \,|\, X = x) P(\theta_{1M} = \tilde{\theta}_{1M} \,|\, X = x^*) P(\theta_{0M} = \tilde{\theta}_{0M} \,|\, X = x^*)$$

where the proof follows the same passages discussed for Equation (4.7). The first equality is obtained applying the law of iterated expectations and the second equality follows from assumptions **B1.1**, **B1.2**, and **B2** since, for example

$$\begin{aligned} P(\theta_{0M}(X^M = x^*, X^Y = x) = \tilde{\theta}_{0M}) & \\ &= P(\theta_{0M} = \tilde{\theta}_{0M} \,|\, X^M = x^*, X^Y = x) \\ &= P(\theta_{0M} = \tilde{\theta}_{0M} \,|\, X^M = x^*, X^Y = x^*) \quad \text{by } \mathbf{B1.1} \\ &= P(\theta_{0M}(X^M = x^*, X^Y = x^*) = \tilde{\theta}_{0M}), \end{aligned}$$

and analogously for any factor in $\mathbf{\Theta}^Y$ by using **B1.2**, and for Y at any time t by **B2**. The third equality follows from 4.1 and the last one from the fact that

$$\begin{aligned} & P(\theta_{1Y}(X = x) = \tilde{\theta}_{1Y} \,|\, \theta_{0Y}(X = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X = x) = \tilde{\theta}_{1M}, \theta_{0M}(X = x) = \tilde{\theta}_{0M}) \\ &= \frac{P(\theta_{1Y}(X = x) = \tilde{\theta}_{1Y},\, \theta_{0Y}(X = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X = x) = \tilde{\theta}_{1M}, \theta_{0M}(X = x) = \tilde{\theta}_{0M} \,|\, X = x)}{P(\theta_{0Y}(X = x) = \tilde{\theta}_{0Y}, \theta_{1M}(X = x) = \tilde{\theta}_{1M}, \theta_{0M}(X = x) = \tilde{\theta}_{0M} \,|\, X = x)} \\ & \qquad\qquad \text{by exchangeability} \\ &= \frac{P(\theta_{1Y} = \tilde{\theta}_{1Y}\, \theta_{0Y} = \tilde{\theta}_{0Y}, \theta_{1M} = \tilde{\theta}_{1M}, \theta_{0M} = \tilde{\theta}_{0M} \,|\, X = x)}{P(\theta_{0Y} = \tilde{\theta}_{0Y}, \theta_{1M} = \tilde{\theta}_{1M}, \theta_{0M} = \tilde{\theta}_{0M} \,|\, X = x)} \quad \text{by consistency} \\ &= P(\theta_{1Y} = \tilde{\theta}_{1Y} \,|\, X = x, \theta_{1M} = \tilde{\theta}_{1M}, \theta_{0M} = \tilde{\theta}_{0M}) \end{aligned}$$

and this line of reasoning ca be extended to the other conditional densities appearing in the expression.

Appendix C

The COVCO-Basel study: an alternative decomposition of income

In Chapter 5, we proposed two alternative decompositions of income, but in the main text we discussed only the one more similar to the approach used for binary exposures. Here, we show the separable mediational effects obtained using the other decomposition, i.e. we estimated

$$SDE = \mathbb{E}[(Y(X^M = 2,000,\, X^Y = 6,000))] - \mathbb{E}(Y(X^M = 2,000,\, X^Y = 3,000))$$
$$SIE = \mathbb{E}[(Y(X^M = 4,000,\, X^Y = 6,000))] - \mathbb{E}[(Y(X^M = 2,000,\, X^Y = 6,000))],$$

corresponding to the intervention moving monthly income from 5,000 to 10,000 CHF, and

$$SDE = \mathbb{E}[(Y(X^M = 0,\, X^Y = 3,000))] - \mathbb{E}(Y(X^M = 0,\, X^Y = 0))$$
$$SIE = \mathbb{E}[(Y(X^M = 2,000,\, X^Y = 3,000))] - \mathbb{E}[(Y(X^M = 0,\, X^Y = 3,000))],$$

corresponding to the intervention changing monthly income from 0 to 5,000 CHF. The tables below show the effects estimated using these decompositions.

In both cases, the direct effects are always negative and significant, as it happens with the other decomposition, but their magnitude is smaller. This is not surprising, since the increment in the X^Y component is just 3,000 CHF, not 5,000 as in the decomposition discussed in the main text. This is likely the same reason why all indirect effects become non-significant, except for that mediated by economic worries at the last occasion under the intervention shifting income from 5,000 to 10,000 CHF. Again, it is possible to observe a slight increase of the direct effects over time, and

Table C.1: Point estimates, standard errors and confidence intervals of separable direct and indirect effects of income on depression mediated by one kind of worry at a time, under an intervention changing income from 5,000 CHF to 10,000 CHF, using an alternative decomposition of income.

Worries	**Effect**	**Fortnight**		
		8	10	12
Economic	SDE	**-0.067** (0.026)	**-0.070** (0.026)	**-0.073** (0.027)
		(-0.115, -0.012)	(-0.109, -0.016)	(-0.112, -0.014)
	SIE	-0.010 (0.007)	-0.010 (0.006)	**-0.011** (0.006)
		(-0.023, 0.004)	(-0.021, 0.001)	(-0.024, -0.002)
Health	SDE	**-0.074** (0.027)	**-0.079** (0.028)	**-0.080** (0.030)
		(-0.123, -0.024)	(-0.135, -0.027)	(-0.136, -0.021)
	SIE	-0.006 (0.008)	-0.005 (0.007)	-0.006 (0.006)
		(-0.022, 0.010)	(-0.019, 0.010)	(-0.019, 0.004)
Social	SDE	**-0.062** (0.026)	**-0.065** (0.025)	**-0.068** (0.028)
		(-0.107, -0.018)	(-0.109, -0.015)	(-0.117, -0.012)
	SIE	-0.009 (0.012)	-0.012 (0.010)	-0.013 (0.010)
		(-0.031, 0.015)	(-0.032, 0.008)	(-0.029, 0.004)
Cultural	SDE	**-0.076** (0.027)	**-0.077** (0.026)	**-0.081** (0.027)
		(-0.128, -0.026)	(-0.128, -0.028)	(-0.135, -0.034)
	SIE	-0.005 (0.009)	-0.004 (0.007)	-0.003 (0.006)
		(-0.020, 0.011)	(-0.017, 0.008)	(-0.015, 0.011)

that the magnitude of effects under the second intervention is somewhat higher than that of effects obtained under the first one, but this is more evident in the results obtained following the 'traditional' decomposition.

These results drive us to delve more deeply into the concept of separability for non-binary exposure/treatments. Leaving aside income, which is a very special kind of exposure, there are many other examples of possible non-binary treatments, such as the already mentioned dose of a drug, the number of physical activity hours or cigarettes smoked in a day, the weekly hours of work or the daily caloric intake. All these examples pose two problems: how to conceive two separate treatments, in order to carry out a reference (even hypothetical) four-arm trial, and how to decompose the unique X variable in its components. The additive way, that we described in Chapter 5 and put in practice here, is probably the most intuitive one, but we believe final judgement on this complex issue can only be reached through future, specifically aimed, research work.

Table C.2: Point estimates, standard errors and confidence intervals of separable direct and indirect effects of income on depression mediated by one kind of worry at a time, under an intervention changing income from 0 CHF to 5,000 CHF, using an alternative decomposition of income.

Worries	**Effect**	**Fortnight**		
		8	10	12
Economic	SDE	**-0.073** (0.026)	**-0.075** (0.026)	**-0.079** (0.028)
		(-0.122, -0.032)	(-0.122, -0.032)	(-0.126, -0.035)
	SIE	-0.010 (0.007)	-0.010 (0.007)	-0.010 (0.007)
		(-0.022, 0.002)	(-0.023, 0.003)	(-0.025, 0.000)
Health	SDE	**-0.076** (0.026)	**-0.079** (0.026)	**-0.081** (0.027)
		(-0.125, -0.025)	(-0.135, -0.029)	(-0.137, -0.026)
	SIE	-0.006 (0.009)	-0.008 (0.008)	-0.006 (0.007)
		(-0.024, 0.011)	(-0.023, 0.008)	(-0.021, 0.007)
Social	SDE	**-0.064** (0.027)	**-0.066** (0.027)	**-0.068** (0.027)
		(-0.121, -0.012)	(-0.121, -0.021)	(-0.115, -0.015)
	SIE	-0.007 (0.012)	-0.010 (0.012)	-0.011 (0.010)
		(-0.030, 0.015)	(-0.030, 0.014)	(-0.031, 0.006)
Cultural	SDE	**-0.079** (0.029)	**-0.082** (0.030)	**-0.086** (0.030)
		(-0.130, -0.023)	(-0.136, -0.017)	(-0.142, -0.025)
	SIE	-0.004 (0.009)	-0.003 (0.007)	-0.003 (0.007)
		(-0.019, 0.010)	(-0.015, 0.009)	(-0.015, 0.011)

Bibliography

Aalen, O. O. and Frigessi, A. (2007), 'What can Statistics Contribute to a Causal Understanding?', *Scand. J. Stat.* **34**(1), 155–168.

Aalen, O. O., Røysland, K. and Gran, J. M. (2012), 'Causality, mediation and time: a dynamic viewpoint', *J. R. Stat. Soc. A* **175**(4), 831–861.

Aalen, O. O., Røysland, K., Gran, J. M., Kouyos, R. and Lange, T. (2016), 'Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms', *Stat. Methods Med. Res.* **25**(5), 2294–2314.

Aalen, O. O., Stensrud, M. J., Didelez, V., Daniel, R., Røysland, K. and Strohmaier, S. (2020), 'Time-dependent mediators in survival analysis: Modeling direct and indirect effects with the additive hazards model', *Biom. J.* **62**, 532–549.

Albert, J. M., Geng, C. and Nelson, S. (2016), 'Causal mediation analysis with a latent mediator', *Biom. J.* **58**(3), 535–548.

Albert, J. M., Li, Y., Sun, J., A.Woyczynski, W. and Nelson, S. (2019), 'Continuous-time causal mediation analysis', *Stat. Med.* **38**(22), 4334–4347.

Alwin, D. F. and Hauser, R. M. (1975), 'The Decomposition of Effects in Path Analysis', *Am. Sociol. Rev.* **40**(1), 37–47.

Andrews, R. M. and Didelez, V. (2021), 'Insights into the Cross-world Independence Assumption of Causal Mediation Analysis', *Epidemiol.* **32**, 209 – 219.

Asparouhov, T. and Muthén, B. (2019), 'Latent Variable Centering of Predictors and Mediators in Multilevel and Time-Series Models', *Struct. Equ. Model.* **26**(1), 119–142.

Avin, C., Shpitser, I. and Pearl, J. (2005), Identifiability of path-specific effects, *in* M. Kaufmann, ed., 'Proceedings of the 19th international joint conference on artificial intelligence', Vol. 19, San Francisco, pp. 257–363.

Baron, R. M. and Kenny, D. A. (1986), 'The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic and Statistical Considerations', *J. Personal. Soc. Psychol.* **51**(6), 1173–1182.

Bauer, D. J. (2003), 'Estimating Multilevel Linear Models as Structural Equation Models', *J. Educ. Behav. Stat.* **28**(2), 135–167.

Bauer, D. J., Preacher, K. J. and Gil, K. M. (2006), 'Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations', *Psychol. Methods* **11**(2), 142–163.

Bickel, R. (2007), *Multilevel analysis for applied research: It's just regression!*, Guilford Press, New York: NY.

Bind, M.-A. C., VanderWeele, T. J., Coull, B. A. and Schwartz, J. D. (2016), 'Causal mediation analysis for longitudinal data with exogenous exposure', *Biostat.* **17**(1), 122–134.

Blood, E. A., Cabral, H., Heeren, T. and Cheng, D. M. (2010), 'Performance of mixed effects models in the analysis of mediated longitudinal data', *BMC Med. Res. Methodol.* **10**(16).

Blood, E. A. and Cheng, D. M. (2011), 'The Use of Mixed Models for the Analysis of Mediated Data with Time-Dependent Predictors', *J. Environ. Public Health* **2011**.

Bollen, K. A. (1987), 'Total, Direct, and Indirect Effects in Structural Equation Models', *Sociol. Methodol.* **17**, 37–69.

Bollen, K. A. (1989), *Structural Equations with Latent Variables*, Wiley, New York: NY.

Bollen, K. A. and Curran, P. J. (2004), 'Autoregressive Latent Trajectory (ALT) Models A Synthesis of Two Traditions', *Sociol. Methods Res.* **32**(3), 336–383.

Bollen, K. A. and Curran, P. J. (2006), *Latent curve models. A structural equation perspective*, Wiley, Hoboken, NJ.

Bollen, K. A. and Zimmer, C. (2010), An Overview of the Autoregressive Latent Trajectory (ALT) Model, *in* K. van Montfort, J. Oud and A. Satorra, eds, 'Longitudinal Research with Latent Variables', Springer, Berlin, Heidelberg, pp. 153–176.

Cheong, J. (2011), 'Accuracy of Estimates and Statistical Power for Testing Mediation in Latent Growth Curve Modeling', *Struct. Equ. Model.* **18**(2), 195–211.

Cheong, J. and MacKinnon, D. P. (2012), Mediation/Indirect Effects in Structural Equation Modeling, *in* R. H. Hoyle, ed., 'Handbook of Structural Equation Models', The Gilford Press, New York, NY, pp. 417–435.

Cheong, J., MacKinnon, D. P. and Khoo, S. T. (2003), 'Investigation of Mediational Processes Using Parallel Process Latent Growth Curve Modeling', *Struct. Equ. Model.* **10**(2), 238.

Chou, C., Bentler, P. M. and Pentz, M. A. (1998), 'Comparisons of two statistical approaches to study growth curves: The multilevel model and the latent curve analysis', *Struct. Equ. Model.* **5**(3), 247–266.

Clare, P. J., Dobbins, T. A. and Mattick, R. P. (2019), 'Causal models adjusting for time-varying confounding-a systematic review of the literature', *Int. J. Epidemiol.* **48**(1), 254–265.

Cole, D. A. and Maxwell, S. E. (2003), 'Testing Mediational Models With Longitudinal Data: Questions and Tips in the Use of Structural Equation Modeling', *J. Abnorm. Psychol.* **112**(4), 558–577.

Collins, M. L., Graham, J. J. and Flaherty, P. B. (1998), 'An alternative framework for defining mediation', *Multivar. Behav. Res.* **33**(2), 295–312.

Curran, P. J. (2003), 'Have Multilevel Models Been Structural Equation Models All Along?', *Multivar. Behav. Res.* **38**(4), 529–569.

Curran, P. J. and Bauer, D. J. (2011), 'The Disaggregation of Within-Person and Between-Person Effects in Longitudinal Models of Change', *Annu. Rev. Psychol.* **62**, 583–619.

Curran, P. J. and Bollen, K. A. (2001), The best of both worlds: Combining autoregressive and latent curve models, *in* L. M. Collins and A. G. Sayer, eds, 'Decade of behavior. New methods for the analysis of change', Am. Psychol. Assoc., pp. 107–135.

Curran, P. J., Lee, T., Howard, A. L., Lane, S. and MacCallum, R. (2012), Disaggregating within-person and between-person effects in multilevel and structural equation growth models, *in* J. R. Harring and G. R. Hancock, eds, 'Advances in longitudinal methods in the social and behavioral sciences', IAP Information Age Publishing, pp. 217–253.

Curran, P. J., Obeidat, K. and Losardo, D. (2010), 'Twelve Frequently Asked Questions About Growth Curve Modeling', *J. Cogn. Dev.* **11**(2), 121–136.

Daniel, R., Cousens, S., De Stavola, B., Kenward, M. G. and Sterne, J. A. C. (2013), 'Methods for dealing with time-dependent confounding', *Stat. Med.* **32**, 1584–1618.

Daniel, R. M., De Stavola, B. L. and Cousens, S. N. (2011), 'gformula: Estimating causal effects in the presence of time-varying confounding or mediation using the g-computation formula', *Stata J.* **11**(4), 479–517.

Daniels, M. J., Roy, J. A., Kim, C., Hogan, J. W. and Perri, M. G. (2012), 'Bayesian Inference for the Causal Effect of Mediation', *Biom.* **68**(4), 1028–1036.

Dawid, P. (2000), 'Causal Inference Without Counterfactuals', *J. Am. Stat. Assoc.* **95**(450), 407–424.

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B. and Micali, N. (2014), 'Mediation Analysis With Intermediate Confounding: Structural Equation Modeling Viewed Through the Causal Inference Lens', *Am. J. Epidemiol.* **181**(1), 64–80.

Deboeck, P. R. and Preacher, K. J. (2016), 'No Need to be Discrete: A Method for Continuous Time Mediation Analysis', *Struct. Equ. Model.* **23**(1), 61–75.

Didelez, V. (2019*a*), Causal Concepts and Graphical Models, *in* M. Maathuis, M. Drton, S. Lauritzen and M. Wainwright, eds, 'Handbook of Graphical Models', Chapman & Hall/CRC, chapter 15, pp. 353–380.

Didelez, V. (2019*b*), 'Defining causal meditation with a longitudinal mediator and a survival outcome', *Lifetime Data Anal.* **25**(4), 593–610.

Didelez, V., Dawid, P. A. and Geneletti, S. (2006), Direct and indirect effects of sequential treatments, *in* 'Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence', p. 138–146.

Domènech-Abella, J., Mundó, J., Leonardi, M., Chatterji, S., Tobiasz-Adamczyk, B., Koskinen, S., Ayuso-Mateos, J. L. and Haro, J. M. (2018), 'The association between socioeconomic status and depression among older adults in Finland, Poland and Spain: A comparative cross-sectional study of distinct measures and pathways', *J. of Affect. Disord.* **241**, 311–318.

Duncan, T. E. and Duncan, S. C. (2004), 'An Introduction to Latent Growth Curve Modeling', *Behav. Ther.* **35**, 333–363.

Eckardt, P. (2012), 'Propensity Score Estimates in Multilevel Models for Causal Inference', *Nurs. Res.* **61**(3), 213–223.

Ellwardt, L., Aartsen, M., Deeg, D. and Steverink, N. (2013), 'Does loneliness mediate the relation between social support and cognitive functioning in later life?', *Soc. Sci. Med.* **98**, 116–124.

Enders, C. K. and Tofighi, D. (2007), 'Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue', *Psychol. Methods* **12**(2), 121–138.

Feller, A. and Gelman, A. (2015), Hierarchical Models for Causal Effects, *in* 'Emerging Trends in the Social and Behavioral Sciences', John Wiley & Sons, Ltd, pp. 1–16.

Ferrer, E. and McArdle, J. (2003), 'Alternative Structural Models for Multivariate Longitudinal Data Analysis', *Struct. Equ. Model.* **10**(4), 493–524.

Finney, J. M. (1972), 'Indirect Effects in Path Analysis', *Sociol. Methods Res.* **1**(2), 175–186.

Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. O. (2006), 'Dynamic path analysis – a new approach to analyzing time-dependent covariates', *Lifetime Data Anal.* **12**, 143–167.

Fox, J. (1980), 'Effect Analysis in Structural Equation Models: Extensions and Simplified Methods of Computation', *Sociol. Methods Res.* **9**(3), 3–28.

Frangakis, C. E. and Rubin, D. B. (2002), 'Principal Stratification in Causal Inference', *Biom.* **58**(1), 21–29.

Freeman, A., Tyrovolas, S., Koyanagi, A., Chatterji, S., Leonardi, M., Ayuso-Mateos, J. L., Tobiasz-Adamczyk, B., Koskinen, S., Rummel-Kluge, C. and Haro, J. M. (2016), 'The role of socio-economic status in depression: results from the COURAGE (aging survey in Europe)', *BMC Public Health* **16**(1).

Fuentes, A., Lüdtke, O. and Robitzsch, A. (2021), 'Causal Inference with Multilevel Data: A Comparison of Different Propensity Score Weighting Approaches', *Multivar. Behav. Res.* pp. 1–24.

Gelman, A. and Hill, J. (2007), *Data analysis using regression and multilevel/hierarchical model*, Cambridge University Press, New York: NY.

Geneletti, S. (2007), 'Identifying direct and indirect effects in a non-counterfactual framework', *J. R. Stat. Soc. B* **69**, 199–215.

Gitelman, A. I. (2005), 'Estimating Causal Effects From Multilevel Group-Allocation Data', *J. Educ. Behav. Stat.* **30**(4), 397–412.

Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M. and Pickles, A. (2017), 'Tutorial: The Practical Application of Longitudinal Structural Equation Mediation Models in Clinical Trials', *Psychol. Methods* **23**(2), 191–207.

Goldstein, H. (1979), *The design and analysis of longitudinal studies*, Academic Press, London.

Goodman, L. A. (1960), 'On the Exact Variance of Products', *J. Am. Stat. Assoc.* **55**(292), 708–713.

Graham, D. J., McCoy, E. J. and Stephens, D. A. (2014), 'Quantifying Causal Effects of Road Network Capacity Expansions on Traffic Volume and Density via a Mixed Model Propensity Score Estimator', *J. Am. Stat. Assoc.* **109**(508), 1440–1449.

Greene, V. L. (1977), 'An Algorithm For Total and Indirect Causal Effects', *Polit. Methodol.* **4**(4), 369–381.

Gu, F., Preacher, K. J. and Ferrer, E. (2014), 'A State Space Modeling Approach to Mediation Analysis', *J. Educ. Behav. Stat.* **39**(2), 117–143.

Gunzler, D., Tang, W., Lu, N., Wu, P. and Tu, X. M. (2014), 'A Class of Distribution-Free Models for Longitudinal mediation analysis', *Psychom.* **79**(4), 543–568.

Haavelmo, T. (1943), 'The statistical implications of a system of simultaneous equations', *Econom.* **11**, 1–12.

Hamaker, E. L., Kuiper, R. M. and Grasman, R. P. P. P. (2015), 'A critique of the cross-lagged panel model', *Psychol. Methods* **20**(1), 102–116.

Harman, H. (1960), *Modern factor analysis*, University of Chicago Press, Chicago.

Hayes, A. F. (2009), 'Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium', *Commun. Monogr.* **76**(4), 408–420.

Hayes, A. F. (2018), *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*, 2 edn, Guilford Press, New York, NY.

Heck, R. H. (2009), Multilevel Modeling With SEM, *in* G. A. Marcoulides and R. E. Schumacker, eds, 'New Developments and Techniques in Structural Equation Modeling', Lawrence Erlbaum Associates, Mahwah, New Jersey: NJ, pp. 89–127.

Heck, R. H. and Thomas, S. L. (2020), *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches*, 4 edn, Routledge, New York: NY.

Henderson, C. R. (1953), 'Estimation of Variance and Covariance Components', *Biom.* **9**, 226–252.

Henderson, C. R. (1973), 'Sire evaluation and genetic trends', *J. Anim. Sci.* **1973**, 10–41.

Hernán, M. and Robins, J. M. (2020), *Causal Inference: What If*, Chapman and Hall/CRC, Boca Raton.

Hill, J. (2013), Multilevel models and causal inference, *in* 'The SAGE handbook of multilevel modeling', SAGE Publications, Ltd, chapter 12, pp. 201–220.

Hoebel, J., Maske, U. E., Zeeb, H. and Lampert, T. (2017), 'Social Inequalities and Depressive Symptoms in Adults: The Role of Objective and Subjective Socioeconomic Status', *PLoS One* **12**(1), 1–13.

Hoffman, L. and Stawski, R. S. (2009), 'Persons as Contexts: Evaluating Between-Person and Within-Person Effects in Longitudinal Analysis', *Res. Hum. Dev.* **6**(2-3), 97–120.

Holland, P. W. (1986), 'Statistics and Causal Inference', *J. Am. Stat. Assoc.* **81**(396), 945–960.

Hong, G. and Raudenbush, S. W. (2006), 'Evaluating Kindergarten Retention Policy: A Case Study of Causal Inference for Multilevel Observational Data', *J. Am. Stat. Assoc.* **101**(475), 901–910.

Huber, P., Ronchetti, E. and Victoria-Feser, M.-P. (2004), 'Estimation of generalized linear latent variable models', *J. R. Stat. Soc. B* **66**, 893–908.

Hudson, C. G. (2005), 'Socioeconomic Status and Mental Illness: Tests of the Social Causation and Selection Hypotheses', *Am. J. Orthopsychiatr.* **75**(1), 3–18.

Imai, K., Tingley, D. and Yamamoto, T. (2013), 'Experimental designs for identifying causal mechanisms', *J. R. Stat. Soc. A* **176**(1), 5–51.

Jo, B., Wang, C.-P. and Ialongo, N. S. (2009), 'Using latent outcome trajectory classes in causal inference', *Stat. Interface* **2**(4), 403–412.

Jöreskog, K. G. (1969), 'A general approach to confirmatory maximum likelihood factor analysis', *Psychom.* **34**, 183–202.

Jöreskog, K. G. (1970), 'A general method for analysis of covariance structures', *Biom.* **57**(2), 239–251.

Jöreskog, K. G. (1973), A general method for estimating a linear structural equation system, *in* A. Goldberger and O. D. Duncan, eds, 'Structural equation models in the social sciences', Academic Press, New York, pp. 85–112.

Jöreskog, K. G. (1978), 'Structural analysis of covariance and correlation matrices', *Psychom.* **43**, 443–477.

Jöreskog, K. G. and Sörbom, D. (2001), *LISREL 8 user's reference guide*, Scientific Software International, Chicago.

Judd, C. M. and Kenny, D. A. (1981), 'Process Analysis: Estimating Mediation in Treatment Evaluations', *Eval. Rev.* **5**(5), 602–619.

Kaplan, D. (2009), *Structural Equation Modeling: Foundations and Extensions*, 2 edn, SAGE Publications, Inc., Thousand Oaks: CA.

Kaplan, D. and Elliott, P. R. (1997), 'A didactic example of multilevel structural equation modeling applicable to the study of organizations', *Struct. Equ. Model.* **4**(1), 1–24.

Keidel, D., Schaffner, E., Vermes, T., Imboden, M. and Probst-Hensch, N. (2021), COVCO-Basel Interim scientific report - July 2020 until January 2021, Technical report, Swiss Tropical and Public Health Institute.

Kenny, D. A., Korchmaros, J. D. and Bolger, N. (2003), 'Lower Level Mediation in Multilevel Models', *Psychol. Methods* **8**(2), 115–128.

Kim, C., Zigler, C. M., Daniels, M. J., Choirat, C. and Roy, J. A. (2019), 'Bayesian Longitudinal Causal Inference in the Analysis of the Public Health Impact of Pollutant Emissions'. arXiv:1901.00908 [stat.ME].

Koopmans, T., Rubin, H. and Leipnik, R. (1950), Measuring the equation systems of dynamic economics (cowles commission monograph 10), *in* T. Koopmans, ed., 'Statistical inference in dynamic economic models', Wiley, New York, pp. 53–237.

Kourouklis, D., Verropoulou, G. and Tsimbos, C. (2020), 'The impact of wealth and income on the depression of older adults across european welfare regimes', *Ageing Soc.* **40**(11), 2448–2479.

Kreft, I. G., de Leeuw, J. and Aiken, L. S. (1995), 'The Effect of Different Forms of Centering in Hierarchical Linear Models', *Multivar. Behav. Res.* **30**(1), 1–21.

Krull, J. L., Cheong, J., Fritz, M. S. and MacKinnon, D. P. (2016), Moderation and mediation in interindividual longitudinal analysis, *in* D. Cicchetti, ed., 'Dev. Psychopathol.', John Wiley & Sons Inc, pp. 922–985.

Krull, J. L. and MacKinnon, D. P. (1999), 'Multilevel mediation modeling in group-based intervention studies', *Eval. Rev.* **23**(4), 418–444.

Krull, J. L. and MacKinnon, D. P. (2001), 'Multilevel Modeling of Individual and Group Level Mediated Effects', *Multivar. Behav. Res.* **36**(2), 249–277.

Lai, K. (2018), 'Estimating Standardized SEM Parameters Given Nonnormal Data and Incorrect Model: Methods and Comparison', *Struct. Equ. Model.* **25**(4), 600–620.

Laird, N. M. and Ware, J. H. (1982), 'Random effects models for longitudinal data', *Biom.* **38**, 963–974.

Lange, T., Vansteelandt, S. and Bekaert, M. (2012), 'A Simple Unified Approach for Estimating Natural Direct and Indirect Effects', *Am. J. Epidemiol.* **176**(3), 190–195.

Lawley, D. N. (1940), 'The estimation of factor loadings by the method of maximum likelihood', *Proc. R. Soc. Edinb.* **60**(1), 64–82.

le Cessie, S., Debeij, J., Rosendaal, F. R., Cannegieter, S. C. and Vandenbroucke, J. P. (2012), 'Quantification of Bias in Direct Effects Estimates Due to Different Types of Measurement Error in the Mediator', *Epidemiol.* **23**(4), 551–560.

Lee, E. J. and Stone, S. I. (2012), 'Co-Occurring Internalizing and Externalizing Behavioral Problems: The Mediating Effect of Negative Self-Concept', *J. Youth Adolesc.* **41**, 717–731.

Lee, S. (1990), 'Multilevel analysis of structural equation models', *Biom.* **77**(4), 763–772.

Lee, S. and Song, X. (2001), 'Hypothesis Testing and Model Comparison in Two-level Structural Equation Models', *Multivar. Behav. Res.* **36**(4), 639–655.

Lee, S. and Zhu, H. (2002), 'Maximum likelihood estimation of nonlinear structural equation models', *Psychom.* **67**(2), 189–210.

Lin, S.-H. and VanderWeele, T. J. (2017), 'Interventional Approach for Path-Specific Effects', *J. Causal Inference* **5**(1).

Lin, S.-H., Young, J. G., Logan, R. and VanderWeele, T. J. (2017), 'Mediation analysis for a survival outcome with time - varying exposures, mediators, and confounders', *Stat. Med.* pp. 1–14.

Lin, S., Young, J., Logan, R., Tchetgen Tchetgen, E. J. and VanderWeele, T. J. (2017), 'Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders', *Epidemiol.* **28**(2), 266–274.

Lin, V., McGrath, S., Zhang, Z., Petito, L. C., Logan, R. W., Hernán, M. A. and Young, J. G. (2020), 'gfoRmula: An R Package for Estimating the Effects of Sustained Treatment Strategies via the Parametric g-formula', *Patterns (NY)* **1**(3).

Liu, L. C., Flay, B. R. and Aban Aya Investigators (2009), 'Evaluating Mediation in Longitudinal Multivariate Data: Mediation Effects for the Aban Aya Youth Project Drug Prevention Program', *Prev. Sci.* **10**(3), 197–207.

Loeys, T., Moerkerke, B., Raes, A., Rosseel, Y. and Vansteelandt, S. (2014), 'Estimation of Controlled Direct Effects in the Presence of Exposure-Induced Confounding and Latent Variables', *Struct. Equ. Model.* **21**(3), 396–407.

Loh, W. W., Moerkerke, B., Tom Loeys, L. P., Crombez, G. and Vansteelandt, S. (2020), 'Estimation of Controlled Direct Effects in Longitudinal Mediation Analyses with Latent Variables in Randomized Studies', *Multivar. Behav. Res.* **55**(5), 763–785.

Lutz, S. M., Sordillo, J. E., Hokanson, J. E., Wu, A. C. and Lange, C. (2020), 'The effects of misspecification of the mediator and outcome in mediation analysis', *Genet. Epidemiol.* **44**(4), 400–403.

MacCallum, C. R. and Austin, T. J. (2000), 'Applications of structural equation modeling in psychological research', *Annu. Rev. Psych.* **51**(1), 201–226.

MacKinnon, D. P. (2008), *Introduction to Statistical Mediation Analysis*, Taylor and Francis Group, New York.

Maxwell, S. E. and Cole, D. A. (2007), 'Bias in Cross-Sectional Analyses of Longitudinal Mediation', *Psychol. Methods* **12**(1), 23–44.

Maxwell, S. E., Cole, D. A. and Mitchell, M. A. (2011), 'Bias in Cross-Sectional Analyses of Longitudinal Mediation: Partial and Complete Mediation Under an Autoregressive Model', *Multivar. Behav. Res.* **46**, 816–841.

Mayer, A., Umbach, N., Flunger, B. and Kelava, A. (2017), 'Effect Analysis Using Nonlinear Structural Equation Mixture Modeling', *Struct. Equ. Model.* **24**(4), 556–570.

McArdle, J. (2001), A latent difference score approach to longitudinal dynamic structural analysis., *in* R. Cudeck, S. D. Toit and D. Sörbom, eds, 'Structural equation modeling: Present and future', Scientific Software International, Lincolnwood, IL, pp. 342–380.

McArdle, J. and Hamagami, F. (2001), Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data, *in* L. M. Collins and A. G. Sayer, eds, 'Decade of behavior. New methods for the analysis of change', Am. Psychol. Assoc., pp. 139–175.

McArdle, J. J. (1978), A structural view of structural models. Paper presented at the Winter Workshop on Latent Structure Models Applied to Developmental Data. University of Denver.

McArdle, J. J. (1979*a*), The development of general multivariate software, *in* 'Proceedings of the Association for the Development of Computer-Based Instructional Systems', University of Akron Press, Akron, Ohio, pp. 824–862.

McArdle, J. J. (1979*b*), Reticular analysis modeling (RAM) theory: the simplicity and generality of structural equations. Paper presented at the American Psychological Association Annual Meeting, New York.

McArdle, J. J. (2005), The Development of the RAM Rules for Latent Variable Structural Equation Modeling, *in* A. Maydeu-Olivares and J. J. McArdle, eds, 'Contemporary Psychometrics', Lawrence Erlbaum Associates, Mahwah, New Jersey: NJ, pp. 225–273.

McArdle, J. J. and McDonald, R. P. (1984), 'Some algebraic properties of the Reticular Action Model for moment structures', *Br. J. Math. Stat. Psychol.* **37**, 234–251.

McArdle, J. J. and Nesselroade, J. R. (2013), An Overview of Latent Curve and Latent Change Score Analyses, *in* I. B. Weiner, ed., 'Handbook of Psychology, Second Edition', John Wiley & Sons, Inc.

McDonald, R. P. (1978), 'A simple comprehensive model for the analysis of covariance structures', *Br. J. Math. Stat. Psychol.* **31**(1), 59–72.

McDonald, R. P. and Goldstein, H. (1989), 'Balanced versus unbalanced designs for linear structural relations in two-level data', *Br. J. Math. Stat. Psychol.* **42**(2), 215–232.

McDonald, R. P. and Swaminathan, H. (1972), Structural analyses of dispersion matrices based on a very general model with a rapidly convergent procedure for the evaluation of parameters. Informal publication of M.E.C.A., O.I.S.E, Toronto.

McNeish, D. and Matta, T. (2018), 'Differentiating between mixed-effects and latent-curve approaches to growth modeling', *Behav. Res. Methods* **50**, 1398–1414.

Mealli, F. and Mattei, A. (2012), 'A Refreshing Account of Principal Stratification', *Int. J. Biostat.* **8**(1), Article 8.

Mehta, P. D. and Neale, M. C. (2005), 'People Are Variables Too: Multilevel Structural Equations Modeling', *Psychol. Methods* **10**(3), 259–284.

Mehta, P. D. and West, S. G. (2000), 'Putting the Individual Back Into Individual Growth Curves', *Psychol. Methods* **5**(1), 23–43.

Meredith, W. and Tisak, J. (1984), "Tuckerizing" growth curves. Paper presented at the annual meeting of the Psychometric Society, Santa Barbara, CA.

Meredith, W. and Tisak, J. (1990), 'Latent curve analysis', *Psychom.* **55**(1), 107–122.

Miočević, M., Gonzalez, O., Valente, M. J. and MacKinnon, D. P. (2018), 'A Tutorial in Bayesian Potential Outcomes Mediation Analysis', *Struct. Equ. Model.* **25**(1), 121–136.

Mittinty, M. N. and Vansteelandt, S. (2020), 'Longitudinal Mediation Analysis Using Natural Effect Models', *Am. J. Epidemiol.* **189**(11), 1427–1435.

Morgan, S. L. and Winship, C. (2015), *Counterfactuals and causal inference: methods and principles for social research*, 2 edn, Cambridge University Press, New York, NY.

Muthén, B. and Asparouhov, T. (2015), 'Causal Effects in Mediation Modeling: An Introduction With Applications to Latent Variables', *Struct. Equ. Model.* **22**(1), 12–23.

Muthén, B. O. (1989), 'Latent variable modeling in heterogeneous populations', *Psychom.* **54**(4), 557–585.

Muthén, B. O. (1991), 'Multilevel factor analysis of class and student achievement components', *J. Educ. Meas.* **28**(4), 338–354.

Muthén, B. O. (1994), 'Multilevel covariance structure analysis', *Sociol. Methods Res.* **22**(3), 376–398.

Muthén, B. O. and Satorra, A. (1989), Multilevel aspects of varying parameters in structural models, *in* R. D. Bock, ed., 'Multilevel analysis of educational data', Academic Press, San Diego, pp. 87–99.

Niku, J., Brooks, W., Herliansyah, R., Hui, F. K. C., Taskinen, S. and Warton, D. I. (2019), 'Efficient estimation of generalized linear latent variable models', *Plos One* **14**(5), e0216129.

Niku, J., Warton, D. I., Hui, F. K. C. and Taskinen, S. (2017), 'Generalized Linear Latent Variable Models for Multivariate Count and Biomass Data in Ecology', *J. Agric. Biol. Environ. Stat.* **22**, 498–522.

O'Laughlin, K. D., Martin, M. J. and Ferrer, E. (2018), 'Cross-Sectional Analysis of Longitudinal Mediation Processes', *Multivar. Behav. Res.* **53**(3), 375–402.

Osafo Hounkpatin, H., Wood, A. M., Brown, G. D. A. and Dunn, G. (2015), 'Why does Income Relate to Depressive Symptoms? Testing the Income Rank Hypothesis Longitudinally', *Soc. Indic. Res.* **124**, 637–655.

Paccagnella, O. (2006), 'Centering or Not Centering in Multilevel Models? The Role of the Group Mean and the Assessment of Group Effects', *Eval. Rev.* **30**(1), 66–85.

Pakpahan, E., Hoffmann, R. and Kr$\acute{}$oger, H. (2017), 'Statistical methods for causal analysis in life course research: an illustration of a crosslagged structural equation model, a latent growth model, and an autoregressive latent trajectories model', *Int. J. Soc. Res. Methodol.* **20**(1), 1–19.

Patel, V., Burns, J. K., Dhingra, M., Tarver, L., Kohrt, B. A. and Lund, C. (2018), 'Income inequality and depression: a systematic review and meta-analysis of the association and a scoping review of mechanisms', *World Psychiatr.* **17**(1), 76–89.

Pearl, J. (1995), 'Causal diagrams for empirical research (with discussion)', *Biom.* **82**(4), 669–710.

Pearl, J. (2001), Direct and indirect effects, *in* J. S. Breese and D. Koller, eds, 'Proceedings of 7th Conference on Uncertainty in artificial intelligence', Morgan Kaufmann, San Francisco, CA, pp. 411 – 420.

Pearl, J. (2009), *Causality*, 2 edn, Cambridge University Press, Cambridge.

Pearl, J. (2010), 'An introduction to causal inference', *Int. J. Biostat.* **6**(2), Article 7.

Pearl, J. (2012), The Mediation Formula: A guide to the assessment of causal pathways in nonlinear models, *in* C. Berzuini, P. Dawid and L. Bernardinelli, eds, 'Causality: Statistical Perspectives and Applications', Wiley, chapter 12, pp. 151–179.

Pearl, J. (2014), Graphical Models for Probabilistic and Causal Reasoning, *in* A. Tucker, T. Gonzalez, H. Topi and J. Diaz-Herrera, eds, 'Computing Handbook', 3 edn, Vol. 1, Chapman and Hall/CRC.

Pituch, K. A., Murphy, D. L. and Tate, R. L. (2010), 'Three-level models for indirect effects in school- and class-randomized experiments in education', *J. Exp. Educ.* **78**(1), 60–95.

Pituch, K. A. and Stapleton, L. M. (2012), 'Distinguishing Between Cross- and Cluster- Level Mediation Processes in the Cluster Randomized Trial', *Sociol. Methods Res.* **41**(4), 630–670.

Pituch, K. A., Stapleton, L. M. and Kang, J. Y. (2006), 'A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials', *Multivar. Behav. Res.* **41**(3), 367–400.

Pituch, K. and Stapleton, L. (2011), Hierarchical Linear and Structural Equation Modeling Approaches to Mediation Analysis in Randomized Field Experiments, *in* M. Williams and W. P. Vogt, eds, 'The SAGE Handbook of Innovation in Social Research Methods', SAGE Publications Ltd, London, pp. 590–619.

Preacher, K. J. (2011), 'Multilevel SEM Strategies for Evaluating Mediation in Three-Level Data', *Multivar. Behav. Res.* **46**(4), 691–731.

Preacher, K. J. (2015), 'Advances in Mediation Analysis: A Survey and Synthesis of New Developments', *Annu. Rev. Psych.* **66**(1), 825–852.

Preacher, K. J., Wichman, A. L., MacCallum, R. C. and Briggs, N. E. (2008), *Latent Growth Curve Modeling*, SAGE Publications, Thousand Oaks, CA.

Preacher, K. J., Zhang, Z. and Zyphur, M. J. (2011), 'Alternative Methods for Assessing Mediation in Multilevel Data: The Advantages of Multilevel SEM', *Struct. Equ. Model.* **18**(2), 161–182.

Preacher, K. J., Zyphur, M. J. and Zhang, Z. (2010), 'A General Multilevel SEM Framework for Assessing Multilevel Mediation', *Psychol. Methods* **15**(3), 209–233.

Rappaport, L. M., Amstadter, A. B. and Neale, M. C. (2020), 'Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling', *Struct. Equ. Model.* **27**(2), 318–329.

Raudenbush, S. W. and Bryk, A. S. (2002), *Hierarchical linear models: Applications and data analysis methods*, 2 edn, Sage, Thousand Oaks: CA.

Raudenbush, S. W. and Schwartz, D. (2020), 'Randomized Experiments in Education, with Implications for Multilevel Causal Inference', *Annu. Rev. Stat. Appl.* **7**(1), 177–208.

Richardson, T. S. and Robins, J. M. (2013), 'Single world intervention graphs (SWIGs): A unification of counterfactual and graphical approaches to causality', (Working Paper 128).

Robins, J. (1986), 'A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survivor effect', *Math. Model.* **7**, 1393–1512.

Robins, J. M. (2003), Semantics of causal DAG models and the identification of direct and indirect effects, *in* P. Green, N. Hjort and S. Richardson, eds, 'Highly Structured Stochastic Systems', Oxford University Press, Oxford, UK, pp. 70–81.

Robins, J. M. and Greenland, S. (1992), 'Identifiability and Exchangeability for Direct and Indirect Effects', *Epidemiol.* **3**(2), 143–155.

Robins, J. M., Hernán, M. and Brumback, B. (2000), 'Marginal structural models and causal inference in epidemiology', *Epidemiol.* **11**, 550–560.

Robins, J. M. and Richardson, T. S. (2011), Alternative graphical causal models and the identification of direct effects, *in* P. Shrout, K. Keyes and K. Ornstein, eds, 'Causality and psychopathology: finding the determinants of disorders and their cures', Oxford University Press, Oxford, pp. 103–158.

Robins, J. M., Richardson, T. S. and Shpitser, I. (2020), 'An Interventionist Approach to Mediation Analysis'. arXiv:2008.06019 [stat.ME].

Robins, J. M. and Wasserman, L. (1997), Estimation of effects of sequential treatments by reparameterizing directed acyclic graphs, *in* 'Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence, UAI '97', AUAI Press, pp. 409–420.

Roesch, S. C., Norman, G. J., Adams, M. A., Kerr, J., Sallis, J. F., Ryan, S., Calfas, K. J. and Patrick, K. (2009), 'Latent Growth Curve Modeling of Adolescent Physical Activity: Testing Parallel Process and Mediation Models', *J. Health Psychol.* **14**, 313–325.

Rovine, M. J. and Molenaar, P. C. M. (1998), 'A nonstandard method for estimating a linear growth model in LISREL', *Int. J. Behav. Dev.* **22**(3), 453–473.

Rovine, M. J. and Molenaar, P. C. M. (2000), 'A Structural Modeling Approach to a Multilevel Random Coefficients Model', *Multivar. Behav. Res.* **35**(1), 51–88.

Rovine, M. J. and Molenaar, P. C. M. (2001), A structural equations modeling approach to the general linear mixed model, *in* L. M. Collins and A. G. Sayer, eds,

'New methods for the analysis of change', Am. Psychol. Assoc., Washington, DC, pp. 65–96.

Rubin, D. B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *J. Educ. Psychol.* **66**(5), 688–701.

Rubin, D. B. (1978), 'Bayesian inference for causal effects: the role of randomization', *Ann. Stat.* **6**(1), 34–58.

Rubin, D. B. (2004), 'Direct and Indirect Causal Effects via Potential Outcomes', *Scand. J. Stat.* **31**, 161–170.

Ryan, O., Kuiper, R. M. and Hamaker, E. L. (2018), A Continuous-Time Approach to Intensive Longitudinal Data: What, Why, and How?, *in* K. van Montfort, J. H. L. Oud and M. C. Voelkle, eds, 'Continuous Time Modeling in the Behavioral and Related Sciences', Springer, chapter 2, pp. 27–54.

Ryu, E. and West, S. G. (2009), 'Level-Specific Evaluation of Model Fit in Multilevel Structural Equation Modeling', *Struct. Equ. Model.* **16**(4), 583–601.

Schuurman, N. K., Ferrer, E., de Boer-Sonnenschein, M. and Hamaker, E. L. (2016), 'How to Compare Cross-Lagged Associations in a Multilevel Autoregressive Model', *Psychol. Methods* **21**(2), 206–221.

Selig, J. P. and Little, T. D. (2012), Autoregressive and cross-lagged panel analysis for longitudinal data, *in* B. Laursen, T. D. Little and N. A. Card, eds, 'Handbook of developmental research methods', The Guilford Press, New York, NY, pp. 265–278.

Selig, J. P. and Preacher, K. J. (2009), 'Mediation Models for Longitudinal Data in Developmental Research', *Res. Hum. Dev.* **6**(2-3), 144–164.

Serang, S., Grimm, K. J. and Zhang, Z. (2019), 'On the Correspondence between the Latent Growth Curve and Latent Change Score Models', *Struct. Equ. Model.* **26**(4), 623–635.

Shardell, M. and Ferrucci, L. (2018), 'Joint mixed-effects models for causal inference with longitudinal data', *Stat. Med.* **37**, 829–846.

Shpitser, I. (2012), Graph-based criteria of identifiability of causal questions, *in* C. Berzuini, P. Dawid and L. Bernardinelli, eds, 'Causality: Statistical Perspectives and Applications', Wiley, chapter 6, pp. 59–71.

Shpitser, I. (2013), 'Counterfactual Graphical Models for Longitudinal Mediation Analysis With Unobserved Confounding', *Cogn. Sci.* **37**, 1011–1035.

Shpitser, I. (2019), Identification in Graphical Causal Models, *in* M. Maathuis, M. Drton, S. Lauritzen and M. Wainwright, eds, 'Handbook of Graphical Models', Chapman & Hall/CRC, chapter 16, pp. 381–404.

Shpitser, I. and Pearl, J. (2008), 'Complete Identification Methods for the Causal Hierarchy', *J. Mach. Learn. Res.* **9**, 1941–1979.

Shpitser, I. and Tchetgen Tchetgen, E. (2016), 'Causal inference with a graphical hierarchy of interventions', *Ann. Stat.* **44**(6), 2433–2466.

Shpitser, I. and VanderWeele, T. J. (2011), 'A complete graphical criterion for the adjustment formula in mediation analysis', *Int. J. Biostat.* **7**(1), article 16.

Shpitser, I., VanderWeele, T. and Robins, J. M. (2010), On the validity of covariate adjustment for estimating causal effects, *in* 'Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence, UAI 2010', AUAI Press, pp. 527–536.

Simone, M. and Lockhart, G. (2019), 'Empirical Sample Size Guidelines for Use of Latent Difference Score Mediation', *Struct. Equ. Model.* **26**(4), 636–645.

Sjölander, A. (2009), 'Bounds on natural direct effects in the presence of confounded intermediate variables', *Stat. Med.* **28**, 558–571.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*, Chapman & Hall/CRC, Boca Raton: FL.

Snijders, T. A. B. and Bosker, R. J. (2012), *Multilevel analysis: An introduction to basic and advanced multilevel modeling*, 2 edn, Sage, Thousand Oaks: CA.

Spirtes, P., Glymour, C. and Scheines, R. (1993), Causation, prediction, and search, *in* 'Lecture Notes in Statistics', Vol. 81, Springer.

Steen, J. and Vansteelandt, S. (2019), Mediation Analysis, *in* M. Maathuis, M. Drton, S. Lauritzen and M. Wainwright, eds, 'Handbook of Graphical Models', Chapman & Hall/CRC, chapter 17, pp. 405–437.

Stensrud, M. J., Hernán, M. A., Tchetgen Tchetgen, E. J., Robins, J. M., Didelez, V. and Young, J. G. (2021), 'A generalized theory of separable effects in competing event settings', *Lifetime Data Anal.* **27**, 588–631.

Stensrud, M. J., Young, J. G., Didelez, V., Robins, J. M. and Hernán, M. A. (2020), 'Separable effects for causal inference in the presence of competing events', *J. Am. Stat. Assoc.* .

Strohmaier, S., Røysland, K., Hoff, R., Borgan, Ø., Pedersen, T. R. and Aalen, O. O. (2015), 'Dynamic path analysis – a useful tool to investigate mediation processes in clinical survival trials', *Stat. Med.* **34**(29), 3866–3887.

Tian, J. and Pearl, J. (2003), On the identification of causal effects, Technical report, Department of Computer Science, University of California, Los Angeles.

Tofighi, D. (2010), Multilevel Mediation Analysis: Statistical Assumptions and Centering, PhD thesis, Arizona State University.

Tofighi, D., Hsiao, Y.-Y., Kruger, E. S., MacKinnon, D. P., Horn, M. L. V. and Witkiewitz, K. (2019), 'Sensitivity analysis of the no-omitted confounder assumption in latent growth curve mediation models', *Struct. Equ. Model.* **26**(1), 94–109.

Tsai, T.-L., Shau, W.-Y. and Hu, F.-C. (2006), 'Generalized Path Analysis and Generalized Simultaneous Equations Model for Recursive Systems with Responses of Mixed Types', *Struct. Equ. Model.* **13**(2), 229–251.

Usami, S., Hayes, T. and McArdle, J. J. (2015), 'On the Mathematical Relationship Between Latent Change Score and Autoregressive Cross-Lagged Factor Approaches: Cautions for Inferring Causal Relationship Between Variables', *Multivar. Behav. Res.* **50**(6), 676–687.

Usami, S., Hayes, T. and McArdle, J. J. (2016), 'Inferring Longitudinal Relationships Between Variables: Model Selection Between the Latent Change Score and Autoregressive Cross-Lagged Factor Models', *Struct. Equ. Model.* **23**(3), 331–342.

Usami, S., Murayama, K. and Hamaker, E. L. (2019), 'A Unified Framework of Longitudinal Models to Examine Reciprocal Relations', *Psychol. Methods* **24**(5), 637 – 657.

Valeri, L., Lin, X. and VanderWeele, T. J. (2014), 'Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model', *Stat. Med.* **33**(28), 4875–4890.

VanderWeele, T. J. (2008), 'Simple relations between principal stratification and direct and indirect effects', *Stat. Prob. Lett.* **78**(17), 2957–2962.

VanderWeele, T. J. (2011), 'Principal Stratification – Uses and Limitations', *Int. J. Biostat.* **7**(1), Article 28.

VanderWeele, T. J. (2015), *Explanation in Causal Inference*, Oxford University Press.

VanderWeele, T. J., Asomaning, K., Tchetgen Tchetgen, E. J., Han, Y., Spitz, M. R., Shete, S., Wu, X., Gaborieau, V., Wang, Y., McLaughlin, J., Hung, R. J., Brennan, P., Amos, C. I., Christiani, D. C. and Lin, X. (2012a), 'Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction', *Am. J. Epidemiol.* **175**(10), 1013–1020.

VanderWeele, T. J. and Hernán, M. A. (2012), Causal effects and natural laws: Towards a conceptualization of causal counterfactuals for nonmanipulable exposures, with application to the effects of race and sex, *in* C. Berzuini, P. Dawid and L. Bernardinelli, eds, 'Causality: Statistical Perspectives and Applications', Wiley, chapter 9, pp. 101–113.

VanderWeele, T. J. and Tchetgen Tchetgen, E. J. (2017), 'Mediation analysis with time varying exposures and mediators', *J. R. Stat. Soc. B* **79**(3), 917–938.

VanderWeele, T. J., Valeri, L. and Ogburn, E. L. (2012b), 'The role of measurement error and misclassification in mediation analysis', *Epidemiol.* **23**(4), 561–564.

VanderWeele, T. J. and Vansteelandt, S. (2009), 'Conceptual issues concerning mediation, interventions and composition', *Stat. Interface* **2**, 457–468.

VanderWeele, T. J., Vansteelandt, S. and Robins, J. M. (2014), 'Effect decomposition in the presence of an exposure-induced mediator-outcome confounder', *Epidemiol.* **25**, 300–306.

Vansteelandt, S., Bekaert, M. and Lange, T. (2012), 'Imputation strategies for the estimation of natural direct and indirect effects', *Epidemiol. Methods* **1**(1), 131–158.

Vansteelandt, S. and Daniel, R. (2017), 'Interventional effects for mediation analysis with multiple mediators', *J. Causal Inference* **28**(2), 258–265.

Vansteelandt, S. and Joffe, M. (2014), 'Structural Nested Models and G-estimation: The Partially Realized Promise', *Stat. Sci.* **29**(4), 707–731.

Vansteelandt, S., Linder, M., Vandenberghe, S., Steen, J. and Madsen, J. (2019), 'Mediation analysis of time-to-event endpoints accounting for repeatedly measured mediators subject to time-varying confounding', *Stat. Med.* **38**, 4828–4840.

Verbeke, G. and Molenberghs, G. (2000), *Linear mixed models for longitudinal data*, Springer, New York: NY.

von Soest, T. and Hagtvet, K. (2011), ' Mediation Analysis in a Latent Growth Curve Modeling Framework', *Struct. Equ. Model.* **18**(2), 289–314.

Wall, M. (2009), Maximum Likelihood and Bayesian Estimation for Nonlinear Structural Equation Models, *in* R. E. Millsap and A. Maydeu-Olivares, eds, 'The SAGE handbook of quantitative methods in psychology', SAGE Publications Ltd, pp. 540–577.

Wang, L. and Maxwell, S. E. (2015), 'On Disaggregating Between-Person and Within-Person Effects With Longitudinal Data Using Multilevel Models', *Psychol. Methods* **20**(1), 63–83.

Wehmeyer, M. L., Palmer, S. B., Shogren, K., Williams-Diehm, K. and Soukup, J. H. (2013), 'Establishing a Causal Relationship Between Intervention to Promote Self-Determination and Enhanced Student Self-Determination', *J. Special Educ.* **46**(4), 195–210.

West, E. A., Anker, D., Amati, R. and *et al.* (2020), 'Corona Immunitas: study protocol of a nationwide program of SARS-CoV-2 seroprevalence and seroepidemiologic studies in Switzerland', *Int. J. Public Health* **65**, 1529–1548.

Wright, S. (1920), 'The relative importance of heredity and environment in determining the piebald pattern of guinea pigs', *Proc. Natl. Acad. Sci.* **6**(6), 320–332.

Wright, S. (1921), 'Systems of mating: I. the biometric relations between parent and offspring', *Genet.* **6**(3), 111–123.

Wright, S. (1934), 'The Method of Path Coefficients', *Ann. Math. Stat.* **5**(3), 161–215.

Wu, W., Carroll, I. A. and Chen, P.-Y. (2018), 'A Single-Level Random Coefficients Cross-Lagged Panel Models for Longitudinal Mediation Analysis', *Behav. Res. Methods* **50**, 2111–2124.

Young, J. G., Hernán, M. and Robins, J. M. (2014), 'Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data', *Epidemiol. Methods* **3**(1), 1–19.

Yuan, K. and Bentler, P. M. (2007), 'Multilevel covariance structure analysis by fitting multiple single-level models', *Sociol. Methodol.* **37**(1), 53–82.

Yuan, Y. and MacKinnon, D. P. (2009), 'Bayesian Mediation Analysis', *Psychol. Methods* **14**(4), 301–322.

Zhang, Q. and Phillips, B. (2018), 'Three-Level Longitudinal Mediation with Nested Units: How does an Upper-Level Predictor Influence a Lower-Level Outcome via an Upper- Level Mediator Over Time?', *Multivar. Behav. Res.* **53**(5), 655–675.

Zhang, Q., Wang, L. and Bergeman, C. S. (2018), 'Multilevel autoregressive mediation models: Specification, estimation, and applications', *Psychol. Methods* **23**, 278–297.

Zhang, Z., Zyphur, M. J. and Preacher, K. J. (2009), 'Testing Multilevel Mediation Using Hierarchical Linear Models Problems and Solutions', *Organ. Res. Methods* **12**(4), 695–719.

Zhao, X., Lynch Jr., J. G. and Chen, Q. (2010), 'Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis', *J. Consum. Res.* **37**, 197–206.

Zheng, W. and van der Laan, M. J. (2017), 'Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes', *J. Causal Inference* **5**(2).

Zhu, S., Sagherian, K., Wang, Y., Nahm, E. and Friedmann, E. (2021), 'Longitudinal Moderated Mediation Analysis in Parallel Process Latent Growth Curve Modeling in Intervention Studies', *Nurs. Res.* **70**(3), 184–192.

Zigler, C. K. and Ye, F. (2019), 'A Comparison of Multilevel Mediation Modeling Methods: Recommendations for Applied Researchers', *Multivar. Behav. Res.* **54**(3), 338–359.

Outputs of the research activity

During my PhD, I authored and co-authored the following papers

- Di Maria, C., Abbruzzo, A. and Lovison, G. (2022) 'Networks as mediating variables: a Bayesian latent space approach'. Statistical Methods and Applications. Accepted.
- Rubino, C., Di Maria, C., Abbruzzo, A. and Ferrante, M. (2022), 'Socio-economic inequality, interregional mobility and mortality among cancer patients: a mediation analysis approach', Socio-Economic Planning Sciences. Accepted.
- Di Maria, C. and Lovison, G. (2022), 'Causal mediation analysis in R: three packages illustrated and compared'. Working paper.
- Di Maria, C., Abbruzzo, A. and Lovison, G. (2021) 'Discrete-time Longitudinal Mediation Analysis: a Review'. Submitted to International Statistical Review
- Di Maria, C., Abbruzzo, A. and Lovison, G. (2021) 'Structural and multilevel linear models for longitudinal mediation analysis: a unification'. In progress
- Di Maria, C., Abbruzzo, A. and Lovison, G. (2021) 'Revisiting multilevel mediation analysis through a definition variable approach'. In progress
- Di Maria, C. and Didelez, V. (2021) 'Longitudinal mediation analysis with latent variables: a separable effect approach.' In progress

Conference Papers

- Di Maria, C. (2021) 'Does self-efficacy affect academic results? A separable-effect mediation analysis', Book of Short Papers-SIS 2021, 1382-1387.
- Di Maria, C., Abbruzzo, A. and Lovison, G. (2020) 'Analysing the mediating role of a network: a Bayesian latent space approach', Book of Short Papers-SIS 2020, 503-508.

I also participated in the following conferences as a speaker

- 50th Meeting of the Italian Statistical Society, 21-25 June 2021.
  Title of the talk: "Does self-efficacy influence academic results? A separable-effect mediation analysis"

- 2021 World Meeting of the International Society for Bayesian Analysis (ISBA), 23 June - 2 July 2021.
  Title of the talk: "Estimating the Bayesian posterior distribution of indirect effects in causal longitudinal mediation analysis" – joint work with Antonino Abbruzzo and Gianfranco Lovison.

- European Causal Inference Meeting (EuroCIM) 2021 - Day 4, 2 September 2021.
  Title of the talk: "Longitudinal mediation analysis with latent variables: a separable effect approach" – joint work with Vanessa Didelez.

- Workshop 'From Data to Causes', 6-7 October 2021.
  Same talk given at the EuroCIM in the form of a lightning presentation ($\sim$5 min).