Interventions in dynamic systems: A causal approach to continuous-time mediation analysis

Oisín Ryan

Utrecht University, Utrecht, The Netherlands

Author Note:

The work of this authors was supported by a grant from the Netherlands Organization for Scientific Research (NWO Onderzoekstalent Grant 406-15-128)

Correspondence concerning this paper should be addressed to Oisín Ryan, Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands. Email: o.ryan@uu.nl.

Abstract

Mediation analysis plays a central role in the investigation of causal mechanisms, where the total, direct and indirect influence of one variable on another is typically assessed using *path-tracing rules*. Two recent developments promise to revolutionize mediation analysis: first, the *interventionist causal inference framework* has allowed researchers to define and investigate mediation in terms of interventions on variables rather than path-tracing rules; second, researchers have begun to use *Continuous-Time (CT) models* to investigate mediation in longitudinal data. This paper shows how CT mediation can be treated using an interventionist approach. The links between interventionist and path-tracing definitions in traditional longitudinal SEM models and in CT models are explored in detail. The treatment of mediation from diverse strands of literature is unified, by showing how total, direct and indirect effects in different frameworks can be defined as qualitatively different types of interventions on the values of variables.

Keywords: Vector autoregressive VAR(1) model, repeated measurements data, continuous-time SEM, interventionist causal inference, path-specific effects.

Interventions in dynamic systems: A causal approach to continuous-time mediation analysis

Introduction

Mediation analysis is a fundamental component of social science research, the most common which researchers use to investigate causal mechanisms: to what degree does X influence Y *directly*, and to what degree is this relationship the result of *indirect* pathways through some mediator(s) M? Mediation analysis has a long history in psychology , where path-specific effects are typically defined in terms of products of regression coefficients, using *path-tracing rules* in SEM models (Baron & Kenny, 1986; Bollen, 1987).

More recently, mediation has been considered within the broader framework of *interventionist causal inference*, a principled approach to linking statistical or associational information to explicitly causal quantities (Robins & Greenland, 1992; Pearl, 2012, 2014). This framework defines the causal effect of interest in terms of a hypothetical experiment or *intervention* on the values of variables: The target of inference is defined, not as the associational parameters themselves, but as a causally meaningful quantity, which researchers can learn about using those parameters. This in turn allows for the clarification of the conditions necessary to identify causal effects from observational or experimental data (Pearl, 2009; Hernán, 2018). Although these conditions may be difficult to verify in observational data, approaching causal inference in this way allows researchers to make predictions about the effects of interventions, which can potentially be verified or falsified by performing the relevant experiment.

In longitudinal settings, mediation has typically been studied using SEM models in which a set of variables measured at a particular occasion are regressed on that same set of variables at a previous occasion. These models are known variously as cross-lagged panel models (CLPMs) or discrete-time (DT) first-order vector auto-regressive (VAR(1)) models. Cole & Maxwell (2003); MacKinnon (2008) discuss the calculation of direct and indirect effects in such models using path-tracing rules. VanderWeele & Tchetgen Tchetgen (2017) in turn define interventionist effects in such settings, showing how both approaches often

lead to comparable definitions.

More and more authors are advocating for the use of Continuous-Time (CT) models in longitudinal settings in the social sciences Oud & Delsing (e.g., 2010); Boker, Montpetit, et al. (e.g., 2010); Chow et al. (e.g., 2005). CT models are appealing for both practical reasons, in that they deal well with unequal time-intervals between observation waves, and conceptual reasons, in that they may better reflect our substantive notions about the underlying data-generating process (Boker, 2002; Aalen et al., 2016; Voelkle et al., 2012). Deboeck & Preacher (2016) and Aalen et al. (2012, 2018) have specifically discussed CT models for mediation analyses, demonstrating the application of path-tracing rules in this setting. These authors argue that very different conclusions about the mediation mechanism result when using CT, rather than DT models: for instance, it is argued that the *total* effect derived from the usual DT-VAR(1) should actually be interpreted as a *direct* effect. What is lacking from the current literature is a treatment of CT mediation within the framework of interventionist causal inference. Deboeck & Preacher (2016) do not discuss whether the effects they define are related to interventionist quantities, while Aalen et al. (2016) in fact advocate a path-tracing rather than interventionist interpretation of mediation in CT processes.

In this paper it is shown that path-tracing mediation effects from DT- and CT-VAR(1) models can all be described in terms of *variable-setting interventions* in a CT system. Specifically, it is shown that the direct effect from the DT model reflects one or more intervention on the mediator at *acute* moments in time: In contrast, the direct and indirect effects in CT models represent interventions on all values of the mediator over a particular *interval* of time. It is shown how these effects can be calculated for any length of time-interval between observations, with any number of variables and paths. By bridging the gap between path-tracing and interventionist causal frameworks in this dynamic setting, researchers are armed with the tools to define and calculate the effects which pertain to their research questions.

The first three sections of this paper consist of a review of existing approaches and an integration of diverse strands of literature, which is necessary for the developments that form the major contribution of this paper. First, a brief introduction to the core concepts in interventionist causal mediation analysis is given. Second, the DT-VAR(1) model is described, along with both the path-tracing and interventionist definitions of mediation effects. Third, the CT model is introduced and its application to mediation analysis is discussed. With these basics in place, the interventionist treatment of CT mediation analysis is given: The different types of interventions which we can potentially learn about from the CT-VAR(1) model are described, and these are related back to the path-tracing definitions given by previous authors. This is followed by a brief discussion of potential applications, the connection to other works concerning causal inference for CT processes, and future directions.

1. A Primer on Causal Inference, Interventions and Mediation

In this paper the general framework of interventionist causal inference, largely based on the specific work of Judea Pearl and James Robins, is used (Pearl, 2009; Robins, 2003).[1] In this framework X is defined as a cause of Y if changing the value of X results in a change in the expected value of Y: the causal effect of X on Y is the difference in *value* of Y following an *intervention*. That is, a (hypothetical) experimental manipulation, to set the value of X. These causal dependencies can be represented in graphical form using Directed Acyclical Graphs (DAGs), which describe the conditional-dependency structure amongst a set of variables. In such a graph, a one-headed arrow between variables, $X \to Y$, denotes that X has a causal influence on Y. DAGs do not imply any model for the form of the relationships between variables, and as such do not make any assumptions regarding, for instance, Gaussian distributions or linearity. As such, given a DAG one still needs to estimate the causal effects of interest, either by nonparametric means, or by fitting some

[1] A detailed treatment of these and other causal inference frameworks, and their relationship to mediation analysis can be found in Hernan & Robins (2013) and VanderWeele (2015), respectively. Here we introduce here only the key concepts and nomenclature.

kind of model to the data. One popular choice in the case of normally distributed variables is linear SEM, which we will utilize to estimate conditional expectations throughout. For a discussion of the relationship between SEM models and DAGs see Bollen & Pearl (2013).

The interventionist treatment of a simple cross-sectional mediation model is given here, to show how total, direct and indirect effects can be defined as interventionist quantities, the assumptions necessary to identify these effects from data, and how these interventionist quantities can be related back to path-tracing effects from linear SEM. The logic of this approach will be used for more complex causal structures, that is, those described by DT- and CT-VAR(1) models, later in the paper. A discussion of the assumptions necessary to identify the causal effects from data is beyond the scope of the current paper. Throughout we will make the simplified assumptions that a) there are no unobserved confounders of any variables in a given system, b) intervening to set variables to values is possible and c) such interventions do not change how these variables are structurally related. For a more detailed treatment we refer readers to Pearl (2009); Dawid (2010); VanderWeele (2015); Robins & Richardson (2010).

1.1 Cross-sectional Mediation effects as interventions

Let us define a DAG, G, representing a classic cross-sectional mediation structure, shown in Figure 1a. We define the *total effect* of X on Y as quantifying the expected change in the outcome variable Y, given that we intervene to set the value of the variable X from some constant $X = x^*$ to some other constant $X = x$.[2] This can be expressed as

$$TE = E[Y|X = x] - E[Y|X = x^*]. \tag{1}$$

The *controlled direct effect* is defined as the expected change in Y, given that we intervene to set the value of variable X from x^* to x, while in both instances holding the value of the

[2] While these expected values in principle could be defined using Pearls *Do-Operator*, i.e $E(Y|Do(X = x))$, for simplicity we will take that the simplified set of assumptions mentioned above mean that the expected value of Y given that we intervene to set X to x is the same as the expected value of Y given that we observe $X = x$, $E(Y|Do(X = x)) = E(Y|X = x)$.

mediator M fixed at a level of m.[3] This can be expressed as

$$CDE = E[Y|X = x, M = m] - E[Y|X = x^*, M = m]. \tag{2}$$

The definition of an indirect effect as an intervention is less straightforward. Pearl (2001) defines the *natural indirect effect* (NIE) as the change in expected value of Y given that X is fixed at x^*, but the mediator M is changed to the value it would have taken on if X had been set to x, given as $M(X = x)$: this can be expressed as

$$NIE = E[Y|X = x^*, M(X = x)] - E[Y|X = x^*, M(X = x^*)]. \tag{3a}$$

The above definition of the NIE is considered problematic, in that, in order to observe $Y|X = x^*, M(X = x)$, we must intervene to set $X = x$, observe $M(X = x)$, then return X to a value of x^*: In contrast the direct and total effect, only require interventions to set variables to some pre-determined constant value. For a discussion on this topic see Robins & Greenland (1992) and Robins & Richardson (2010).

Robins & Richardson (2010) offers an alternative definition of the natural indirect effect: For the NIE to be meaningful, it should be theoretically justified that the predictor variable X can be split into two components, one which influences Y (that is, X^a) and one which influences M (that is, X^b), both of which are *fully determined* by X. Following Robins & Richardson (2010), this theoretical decomposition of X can be represented using an extended DAG, G', shown in Figure 1b, where deterministic relationships are denoted by hollow edges. The original and extended graphs, G and G' respectively, represent statistically indistinguishable explanations for observational data, as in these cases X^a is always equal to X^b. As such the existence of extended DAG G' must be justified from a

[3] Two kinds of direct effect can be defined, the controlled direct effect, in which the mediator is set to some constant m, and the natural direct effect, in which the mediator value is a function of $x*$. These will be equivalent in the case of linear models with no interaction effects. Such models will be the focus of the current paper, and so we introduce only the controlled direct effect here (cf. VanderWeele, 2015).

theoretical standpoint: It represents our substantive notion that, in experimental settings, we could intervene on the system to force X^b to take on a different value than X^a.[4]

Given the extended DAG G', we can define the NIE in terms of an intervention to change the value of X^b from x^* to x, while keeping X^a fixed. Alternatively, we can define this as the CDE (see Equation (2)), n which X^a is treated as the mediator. It is this latter definition of the NIE which will be used throughout the rest of this paper. In the cross-sectional case the NIE can thus be expressed as

$$NIE = E[Y|X = x, X^a = x^*] - E[Y|X = x^*, X^a = x^*]. \tag{3b}$$

1.2 Linking Path-tracing to interventions

Given the relevant assumptions about our causal system of interest, and assuming that the conditional dependencies above can be expressed using a linear SEM model, we can relate the path-tracing definitions and interventionist definitions of total, direct and indirect effects in a straightforward way. Take it that the dependencies in the DAG in Figure 1a can be represented by the linear SEM model

$$\begin{bmatrix} E[X] \\ E[M|X] \\ E[Y|X,M] \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} X \\ M \\ Y \end{bmatrix}.$$

In the general SEM literature, mediation effects are defined by the well known *path-tracing rules* (cf. Wright, 1934; Bollen, 1987). As such, the total, direct and indirect effects are defined in terms of (products of) regression ($\boldsymbol{\beta}$) coefficients. The path-tracing total effect is

[4] Robins & Richardson (2010) discusses an example where X is cigarette smoking (binary), M is blood pressure and Y is lung cancer. Cigarettes have multiple ingredients: let us take it that the nicotine ingredient causes increases in blood pressure X^b, while the other ingredients contribute directly to developing lung cancer X^a. The indirect effect in this case could refer to the difference in lung cancer levels given participants i) stop smoking cigarettes entirely ($X = 0 \Rightarrow X^a = X^b = 0$) and ii) stop smoking cigarettes but start using nicotine patches ($X = 0 \Rightarrow X^a = 0$ but $X^b = 1$).

made up of all the paths between X and Y: this is given as $\beta_{31} + \beta_{21}\beta_{32}$. The corresponding direct effect is that part of the total effect which does not contain paths through M, given as β_{31}. The indirect effect is that part of the total effect which only contains paths through M, that is $\beta_{21}\beta_{32}$.[5]

We can calculate the *interventionist* total, direct and indirect effects, by using the linear SEM model above to evaluate the relevant (conditional) expected values of Y. These are given as

$$\begin{aligned} TE &= E[Y|X = x] - E[Y|X = x^*] \\ &= (\beta_{31} + \beta_{21}\beta_{32})(x - x^*) \\ CDE &= E[Y|X = x, M = m] - E[Y|X = x^*, M = m] \\ &= \beta_{31}(x - x^*) \\ NIE &= E[Y|X = x, X^a = x^*] - E[Y|X = x^*, X^a = x^*] \\ &= \beta_{21}\beta_{32}(x - x^*). \end{aligned}$$

As can be seen, if the SEM model above is appropriate, the interventionist and path tracing definitions of mediated effects differ only by a factor of $(x - x*)$, that is, only in that the interventionist effects are defined relative to a pair of values of X.

2. Mediation and Time-Varying Variables: Current best practice

Now that we have covered the basic concepts linking interventionist and path-tracing approaches to mediation, we can examine mediation in a dynamic context, that is, with time-varying X, M and Y. A popular model used for data made up of repeated measurement waves is the DT-VAR(1) model, also known as the Cross-Lagged Panel model (cf. Rogosa, 1980; Selig & Little, 2012; Hamaker et al., 2015; Hamaker & Dolan, 2009).

[5] Total effects are found by taking the sum of increasing powers of $\boldsymbol{\beta}$; indirect effects are found by first removing the direct pathways from $\boldsymbol{\beta}$, (i.e. setting $\beta_{32} = 0$), then taking the sum of increasing powers of this altered $\boldsymbol{\beta}$ matrix

The DT-VAR(1) is a linear SEM model, and has been suggested for use with mediation analysis by, amongst others, Cole & Maxwell (2003) and MacKinnon (2008). In this section, the DT-VAR(1) model is introduced, both path-specific and interventionist definitions of mediation in the context of this model are reviewed, and the well-known time-interval problem associated with the DT-VAR(1) is discussed.

2.1 The DT-VAR(1) model

The DT-VAR(1) process can be written in general as

$$\boldsymbol{Z}_\tau = \boldsymbol{\Phi}\boldsymbol{Z}_{\tau-1} + \boldsymbol{\epsilon}_\tau, \tag{4}$$

where, given V variables, the $V \times 1$ column vector of random variables $\boldsymbol{Z}$ at occasion τ are a function of the $V \times 1$ column vector of those same variables at a previous occasion, $\boldsymbol{Z}_{\tau-1}$, and their dependencies are encoded in the $V \times V$ matrix of autoregressive and cross-lagged parameters, $\boldsymbol{\Phi}$. Here we treat the random variables $\boldsymbol{Z}$ as though they are centered around their long-term average value, $\boldsymbol{\mu}$, and so omit any intercept term. The $V \times 1$ column vector $\boldsymbol{\epsilon}_\tau$ represent the random shocks or innovations, with mean zero and variance-covariance matrix $\boldsymbol{\Psi}$ (Hamilton, 1994). Throughout the current article the elements of $\boldsymbol{\Phi}$ are denoted ϕ_{jk}, representing the effect of variable k at measurement occasion $\tau - 1$ on variable j at measurement occasion τ ($j = 1, \ldots, V$, $k = 1, \ldots, V$). It is an assumption of the DT-VAR(1) model that the occasions τ are equally spaced in time: the same amount of time elapses between each consecutive occasion. Furthermore it is assumed that the auto-regressive and cross-lagged parameters $\boldsymbol{\Phi}$ are equal across occasions.

For simplicity, and without loss of generality, we will focus on the simplest case of a VAR(1) process in which mediation can be studied: that is, when we have three variables, and a lower triangular $\boldsymbol{\Phi}$ matrix. This is also the type of system examined by Cole & Maxwell (2003). Figure 2a represents the DAG of such a process, and the autoregressive

and cross-lagged parameters are given by

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & 0 & 0 \\ \phi_{21} & \phi_{22} & 0 \\ \phi_{31} & \phi_{32} & \phi_{33} \end{bmatrix}$$

meaning that the $M_{\tau-1} \rightarrow X_\tau$, $Y_{\tau-1} \rightarrow M_\tau$ and $Y_{\tau-1} \rightarrow X_\tau$ effects are zero.

2.2 Path-tracing and interventions in the VAR(1) Model

Cole & Maxwell (2003) approach the issue of path-specific mediation effects using linear SEM path-tracing rules. Cross-lagged effects ϕ_{jk} are interpreted as direct effects of $Z_{k,\tau-1}$ on $Z_{j,\tau}$. The value of direct and indirect effects depends on the number of measurement occasions which separate them (i.e., $\Delta\tau$): The total, direct and indirect effect of X_0 on Y_1 will not generally be the same as the corresponding path-specific effect of X_0 on Y_2. The direct effect of X_0 on Y_1 is given as ϕ_{31}, and the direct effect of X_0 on Y_2 is given as $\phi_{31}\phi_{33} + \phi_{11}\phi_{31}$. Similarly, there is no indirect effect of X_0 on Y_1, while the indirect effect of X_0 on Y_2 through M_1 is given as $\phi_{21}\phi_{32}$.

The interventionist definitions of direct and indirect effect outlined above for the cross-sectional case can be extended to time-varying predictors, mediators and outcome variables (VanderWeele & Tchetgen Tchetgen, 2017; VanderWeele, 2015). As in the cross-sectional case, we will here define these effects in terms of differences in conditional expectations given interventions; we will then show the equivalence of path-tracing and interventionist quantities, in the case where the assumptions of linear SEM are met. For a discussion on the relevant identifying assumptions, see VanderWeele (2015, chapter 6).

2.2.1 Total Effect. Let us define the total effect of X_τ on $Y_{\tau+\Delta_\tau}$ (the value that Y takes on Δ_τ occasions later) as TE_{Δ_τ}. This represents the expected change in value of $Y_{\tau+\Delta_\tau}$ given an intervention to set X_τ from a value of x^* to x, written as

$$TE_{\Delta_\tau} = E[Y_{\tau+\Delta_\tau}|X_\tau = x] - E[Y_{\tau+\Delta_\tau}|X_\tau = x^*]. \tag{5a}$$

Taking $\Delta_\tau = 1$, we can use the VAR(1) model defined above to come to an expression for TE_1

$$\begin{aligned} TE_1 &= E[Y_{\tau+1}|X_\tau = x] - E[Y_{\tau+1}|X_\tau = x^*]. \\ TE_1 &= \phi_{31}x + \phi_{32}E[M_\tau] + \phi_{33}E[Y_\tau] - (\phi_{31}x^* + \phi_{32}E[M_\tau] + \phi_{33}E[Y_\tau]) \\ TE_1 &= \phi_{31}(x - x^*). \end{aligned} \tag{5b}$$

This shows that the path-tracing definition of the direct effect reflects the effect on $Y_{\tau+1}$ of a variable-setting intervention on the value of X_τ.

2.2.2 Controlled Direct Effect. We can define the CDE_{Δ_τ} as the expected change in $Y_{\tau+\Delta\tau}$ given that we intervene to set X_τ from x^* to x, and intervene to hold M constant at some level m at each intervening measurement occasion between τ and $\tau + \Delta\tau$, that is, $M_{\tau+1} = \cdots = M_{\tau+(\Delta_\tau-1)} = m$. In the general case we can write

$$\begin{aligned} CDE_{\Delta_\tau} =& E[Y_{\tau+\Delta_\tau}|X_\tau = x, \{M_{\tau+1} \dots M_{\tau+(\Delta_\tau-1)}\} = m] \\ &- E[Y_{\tau+\Delta_\tau}|X_\tau = x^*, \{M_{\tau+1} \dots M_{\tau+(\Delta_\tau-1)}\} = m] \end{aligned} \tag{6a}$$

where $\{M_{\tau+1} \dots M_{\tau+(\Delta_\tau-1)}\}$ indicates the values of M at each equally spaced measurement occasion $M_{\tau+k}$ for k from 0 to Δ_τ.[6]

Based on this we can define, for instance, CDE_2, as the expected change in $Y_{\tau+2}$, given an intervention on X_τ, and interventions on the value of $M_{\tau+1}$ and. This can be written as

$$CDE_2 = E[Y_{\tau+2}|X_\tau = x, M_{\tau+1} = m] - E[Y_{\tau+2}|X_\tau = x^*, M_{\tau+1} = m]. \tag{6b}$$

We can calculate this effect by by expressing each of the expected values above using the

[6] Due to the fact that M_τ is independent of X_τ, and the fact that Y_τ is independent of M_τ in our DAG, this is also equivalent to intervening on the values of $\{M_\tau, M_{\tau+1}, \dots M_{\tau+\Delta_\tau}\}$. Both formulations lead to equivalent results.

VAR(1) model. That is,

$$E[Y_{\tau+2}|X_\tau, M_{\tau+1}] = \phi_{31}E[X_{\tau+1}] + \phi_{32}E[M_{\tau+1}] + \phi_{33}E[Y_{\tau+1}]$$

where, plugging in the VAR(1) model for $E(X_{\tau+1})$ and $E(Y_{\tau+1})$, we come to

$$E[Y_{\tau+2}|X_\tau, M_{\tau+1}] = \phi_{31}(\phi_{11}E[X_\tau]) + \phi_{32}E[M_{\tau+1}] + \phi_{33}(\phi_{31}E[X_\tau] + \phi_{32}E[M_\tau] + \phi_{33}E[Y_\tau]).$$

Now, to evaluate $E[Y_{\tau+2}|X_\tau = x, M_{\tau+1} = m]$, we can replace $E[X_\tau]$ with x and $E[M_{\tau+1}]$ with m,

$$E[Y_{\tau+2}|X_\tau = x, M_{\tau+1} = m] = \phi_{31}(\phi_{11}x) + \phi_{32}m + \phi_{33}(\phi_{31}x + \phi_{32}E[M_\tau] + \phi_{33}E[Y_\tau]).$$

from which it can easily be seen that

$$CDE_2 = (\phi_{31}\phi_{11} + \phi_{33}\phi_{31})(x - x^*) \tag{6c}$$

showing that the CDE_2 exactly equals the direct effect given by path-tracing rules.

2.2.3 Natural Indirect Effect. The natural indirect effect NIE_{Δ_τ} can be defined as the expected change in $Y_{\tau+\Delta\tau}$ given that we hold X_τ constant at x^*, but that $M_{\tau+1}, \dots M_{\tau+(\Delta_\tau-1)}$ take on the values they would have taken on had X_τ been set to x. In the general case we can write

$$\begin{aligned} NIE_{\Delta_\tau} = {} & E[Y_{\tau+\Delta_\tau}|X_\tau = x^*, \{M_{\tau+1}, \dots M_{\tau+(\Delta_\tau-1)}\}(X_\tau = x)] \\ & - E[Y_{\tau+\Delta_\tau}|X_\tau = x^*, \{M_{\tau+1}, \dots M_{\tau+(\Delta_\tau-1)}\}(X_\tau = x^*)] \end{aligned} \tag{7}$$

where $\{M_{\tau+1}, \dots M_{\tau+(\Delta_\tau-1)}\}(X_\tau = x^*)$ denotes that the mediator variable at discrete occasions $\tau + 1$ to $\tau + (\Delta_\tau - 1)$ take on the values they would have taken on had X_τ taken on a value of x^*.

As we did in the cross-sectional case, it is helpful to think of the indirect effect as referring to a scenario in which we separate the $Y_{\tau+1}$ -causing component of X_τ, X_τ^a, from the $M_{\tau+1}$ -causing component of X_τ, X_τ^b. The relevant extended DAG is shown in Figure 2b. We can represent the lagged relationships in this DAG in matrix form: just as G' denoted the extended causal graph of G, we will denote the matrix of parameters in the extended graph as $\mathbf{\Phi'}$, where the first and second rows of $\mathbf{\Phi'}$ denote the paths entering and leaving X_τ^a and X_τ^b respectively. This extended matrix is given by

$$\mathbf{\Phi'} = \begin{bmatrix} \phi_{11} & 0 & 0 & 0 \\ 0 & \phi_{11} & 0 & 0 \\ 0 & \phi_{21} & \phi_{22} & 0 \\ \phi_{31} & 0 & \phi_{32} & \phi_{33} \end{bmatrix} \tag{8}$$

where the parameters are given in terms of their positions in the original $\mathbf{\Phi}$ matrix. As shown by the DAG in Figure 2b, the parent variable X_τ has autoregression of ϕ_{11}, and fully determines each of X_τ^a and X_τ^b at each measurement occasion, denoted by the hollow arrows: for simplicity, the parent variable X_τ is omitted from this matrix.

The NIE_{Δ_τ} can thus be defined in relation to the extended DAG as

$$NIE_{\Delta_\tau} = E[Y_{\tau+\Delta_\tau}|X_\tau = x, X_\tau^a = x^*] - E[Y_{\tau+\Delta_\tau}|X_\tau = x^*, X_\tau^a = x^*] \tag{9a}$$

Using this, NIE_2 can be expressed as

$$\begin{aligned} NIE_2 &= E[Y_{\tau+2}|X_\tau = x, X_\tau^a = x^*] - E[Y_{\tau+2}|X_\tau = x, X_\tau^a = x^*] \\ &= \phi_{21}\phi_{32}(x - x^*) \end{aligned} \tag{9b}$$

Again we see that the NIE_2 is equal to the associated path-tracing indirect effect.

2.3 The time-interval problem

Despite the popularity and widespread use of this model, the DT-VAR(1) model has a major drawback. This is the issue of *time-interval dependency*: referring to the fact that estimates of lagged parameters $\mathbf{\Phi}$ depend on the time-interval Δt with which measurements τ are taken. This dependency has long been noted in the literature (Gollob & Reichardt, 1987), and is discussed in relation to this model by Cole & Maxwell (2003) and Reichardt (2011).

Take for instance that the model in Figure 2a is the "true" data-generating model, which would be obtained from observations taken with some uniform time-interval between each measurement, $\Delta t_\tau = \Delta t$ for all τ. Cole & Maxwell (2003) show that if researchers observe the same process with a different interval than Δt, the resulting model will lead to different estimates of the total, direct and indirect effects. In Figure 2c we show the cross-lagged parameters we would obtain using an observation interval of $2\Delta t$, in terms of the original parameters. Using an interval of $2\Delta t$, the *direct* effect of X_0 on Y_2, given by $\phi_{11}\phi_{31} + \phi_{31}\phi_{32} + \phi_{21}\phi_{32}$, is in fact equal to the *total* made up of pathways through the latent, unobserved values of X_1, M_1, and Y_1 from the "true" data-generating model.

Reichardt (2011) provides a further treatment of this phenomenon in relation to the VAR(1) model, showing that conclusions regarding the existence or non-existence of direct and indirect effects can change depending on the time-intervals between measurement occasions. Other researchers have taken this as a motivation for moving away from the DT-VAR(1) model towards alternative continuous-time models (e.g., Oud & Delsing, 2010).

3. Continuous-Time systems and mediation

The time-interval dependency problem results from using the *discrete-time* version of the VAR(1): DT models treat time only with respect to the ordering of measurement occasions, and ignore the actual length of time between these occasions. However the DT-VAR(1) should be viewed as an approximation or simplification of a *dynamic process*, that is, a system of time-varying variables, which evolves and varies *continuously* over time.

Here we describe and motivate an alternative *continuous-time* (CT) approach to modeling time-varying processes; we describe how this is related to the DT-VAR(1); we show how such processes can be depicted using causal graphs; and we review path-tracing based approaches to mediation analysis in the CT-VAR(1) model.

3.1 The Continuous-Time VAR(1)

The continuous-time equivalent of the (discrete-time) VAR(1) system can be represented in two ways: as a first-order stochastic differential equation (SDE) or as a CT-VAR(1), the integral form of the SDE. Both formulations are presented here. Take it that we have V time-varying processes $Z_1, \dots Z_V \in \boldsymbol{Z}$. These processes can be said to make up a $V-$dimensional *system*. The position of the V-dimensional system at a certain point in time, relative to the equilibrium position $\boldsymbol{\mu}$ is denoted $\boldsymbol{Z}(t)$. Without loss of generality, take it the equilibrium position $\boldsymbol{\mu}$ is zero. The first-order SDE, describes how the *rate of change* of the process with respect to time $\frac{d\boldsymbol{Z}(t)}{dt}$ is a function of $\boldsymbol{Z}(t)$. The former can also be thought of as a vector of *velocities*, describing in what direction and with what magnitude the system will move an instant later in time (i.e., the ratio of the *change in position over some time-interval*, to the*length of that time-interval*, as the length of the time-interval approaches zero).

The first-order SDE is given as

$$\frac{d\boldsymbol{Z}(t)}{dt} = \boldsymbol{A}\boldsymbol{Z}(t) + \boldsymbol{\Omega}\frac{d\boldsymbol{W}(t)}{dt} \tag{10a}$$

where $\boldsymbol{Z}(t)$ and $\frac{d\boldsymbol{Z}(t)}{dt}$ are $V \times 1$ column vectors described above, and the $V \times V$ matrix $\boldsymbol{A}$ represents the *drift* matrix relating $\frac{d\boldsymbol{Z}(t)}{dt}$ to $(\boldsymbol{Z}(t))$. The diagonal elements of $\boldsymbol{A}$, relating the position in a certain dimension to the velocity in that same dimension, are referred to as *auto-effects* while the off-diagonal elements are referred to as *cross-effects*. The second part on the right-hand side of Equation (10a) represents the stochastic innovation part of the model, a Wiener process with variance-covariance matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^{\boldsymbol{T}}$, sometimes referred to

as the diffusion matrix (for details see Voelkle et al., 2012). For simplicity we will assume that the stochastic innovations are uncorrelated.

The first-order SDE can equivalently be written in integral form, as the Continuous-Time VAR(1) system (also known as the Ornstein-Uhlenbeck model, cf. Oravecz et al., 2009). We can write this model as

$$\boldsymbol{Z}(t + \Delta t) = \boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{Z}(t) + \boldsymbol{w}(\Delta t) \tag{10b}$$

where Δt represents the length of the time-interval between consecutive measurements, $\boldsymbol{A}$ has the same meaning as above, $\boldsymbol{Z}(t)$ represents the value of our set of variables at the current point in time t, $\boldsymbol{Z}(t + \Delta t)$ represents the values of our set of variables some time interval Δt later, and these two are related by $e^{\boldsymbol{A}\Delta t}$, the matrix exponential of the drift matrix multiplied by the length of the time-interval. The $V \times 1$ column vector $\boldsymbol{w}(\Delta t)$ represents the stochastic innovations, the integral form of the Wiener process in Equation (10a). These innovations are normally distributed with a variance-covariance matrix that is a function of the time-interval between measurements Δt, the drift matrix $\boldsymbol{A}$, and the diffusion matrix $\boldsymbol{\Omega}\boldsymbol{\Omega}^{\boldsymbol{T}}$ (cf. Voelkle et al., 2012). Throughout we make an assumption of stability equivalent to that described in DT case: we assume that $\boldsymbol{A}$ is invariant over time.

3.2 Relating the DT- and CT-VAR(1)

By comparing Equations 4 and 10b we can see that the CT- and DT- VAR(1) models are closely related. Recalling that we can re-write the DT-VAR(1) effects matrix $\boldsymbol{\Phi}$ as dependent on the length of the time-interval $\boldsymbol{\Phi}(\Delta t)$, the CT and DT effects matrices are related by the equality

$$e^{\boldsymbol{A}\Delta t} = \boldsymbol{\Phi}(\Delta t). \tag{11}$$

Based on this, the DT-VAR(1), in which observations are equally spaced in time, can be seen as an approximation or simplification of the CT-VAR(1) model: If there are equal

intervals between measurements, then DT-VAR(1) model parameters will be an estimate of $\boldsymbol{\Phi}(\Delta t)$ for whatever Δt is used in data collection. However if this assumption of the DT-VAR(1) is broken, then the estimated $\boldsymbol{\Phi}$ will be a mix of $\boldsymbol{\Phi}(\Delta t)$ for different values of Δt.

Conceptually, the first-order SDE represents a different way to express causal dependencies between variables over time than the DT-VAR(1). In the first-order SDE, the effects matrix $\boldsymbol{A}$ encodes the *local dependencies* between processes, that is, how the values of the process at a particular point in time $\boldsymbol{Z}(t)$ relate to the values those processes will take on a very small step forward in time later, determined by the derivative $\frac{d\boldsymbol{Z}(t)}{dt}$ (Aalen, 1987; Didelez, 2008).[7] Given the local dependency structure from the first-order SDE, we can extrapolate the dependencies between $\boldsymbol{Z}_\tau$ and $\boldsymbol{Z}_{\tau-1}$ for any arbitrary time-interval between occasions Δt. This is what is expressed by the $e^{\boldsymbol{A}\Delta t}$ term in the CT-VAR(1) model.

The key conceptual difference between the DT- and CT-VAR(1) systems is that the CT system encodes the substantive belief that the processes $\boldsymbol{Z}(t)$ take on some value at every point in time t. Thus, for any two sets of consecutive observations, there are an infinite amount of unobserved latent values of the variables of interest, which could have been observed between these measurement occasions. Furthermore, it is assumed that the processes of interest continue to vary and exert influence on one another in between measurement occasions. Various authors have argued that such a CT model is a more realistic conceptual model for, amongst others, psychological or biological processes, which are unlikely to evolve or influence one another in discrete-steps over time (e.g. Oravecz et al., 2011; Boker, 2002; Aalen et al., 2012)

We can represent the CT-VAR(1) model graphically in two different ways. First, it can be represented as a SEM model or DAG with an infinite number of latent values of X, M and Y between measurement occasions, as in Figure 3a Deboeck & Preacher

[7] Note that local dependency in this context refers to two processes being locally independent in time. This is different from the use of local independence given by, for example Borsboom (2008), in which this refers to the independence of observed variables conditional on a common cause latent variable

(represented similarly by 2016). In this representation, each latent value of $X(t)$ has an effect on the value of $M(t + \Delta t)$ as $\Delta t \to 0$. That is, the cross lagged arrows between latent variables in this graph represent *local dependencies*, as described by the drift matrix $\boldsymbol{A}$. Here we depict a typical mediation scenario, that is, a lower triangular $\boldsymbol{A}$ matrix. Equivalently, we can represent the dependencies in $\boldsymbol{A}$ using a *local independence graph* (LIG, cf. Didelez, 2007, 2008; Aalen et al., 2012) shown in Figure 3b. This is a more parsimonious representation, where nodes represent processes, and arrows between nodes represent local dependencies i.e., non-zero elements of $\boldsymbol{A}$.

3.3 Mediation in the CT-VAR(1) using path-tracing rules

Deboeck & Preacher (2016) argue that for CT mediation models, the parameters $e^{\boldsymbol{A}\Delta t}$, and as such, $\boldsymbol{\Phi}(\Delta t)$ for any given Δt should be interpreted as total effects rather than direct effects. This argument follows the same reasoning as above when we took it that the DT-VAR(1) model with interval Δt was the true data-generating model in comparison to the DT model with $2\Delta t$: Any $\boldsymbol{\Phi}(\Delta t)$ represents the combination of all pathways between a pair of variables through unobserved latent values of X, M and Y, as depicted in Figure 3a. As such, in the general case, two processes which are time-locally independent may still produce lagged values that are dependent, due to indirect pathways ($A_{jk} = 0 \not\Rightarrow \boldsymbol{\Phi}(\Delta t)_{jk} = 0$) (see also Aalen et al., 2012, 2016, 2018).

Deboeck & Preacher (2016), and equivalently Aalen et al. (2012) and Aalen et al. (2018), thus offer an alternative method for calculating the direct and indirect effects based on path-tracing. They suggest calculating path-specific effects by first altering the $\boldsymbol{A}$ matrix to remove pathways between variables which you wish to exclude, then applying the matrix exponential function $\boldsymbol{e}^{\boldsymbol{A}\Delta t}$ to this altered drift matrix to obtain the path-specific lagged effect of interest. This logic is similar to the method of Bollen (1987) for general SEM path-tracing, where he suggests finding direct and indirect effects by setting paths which are not of interest to zero, and taking the sum of increasing powers of the effects

matrix.[8]

Given the drift matrix

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Deboeck & Preacher suggest that to find the direct effect of $X(t)$ on $Y(t + \Delta t)$, we would set a_{21} (representing the relationship between $X(t)$ and the *rate of change* of $M(t)$) and a_{32} (representing the relationship between $M(t)$ and the rate of change of $Y(t)$) to zero, before taking the matrix exponential of this altered drift matrix. Applying Equation 11 to this altered drift matrix, we find a direct effect parameter of

$$\phi_{31}^{D}(\Delta t) = \frac{a_{31}(e^{a_{11}\Delta t} - e^{a_{33}\Delta t})}{a_{11} - a_{33}} \tag{12a}$$

where superscript D denotes that this is the ϕ parameter which results from only direct pathways between $X(t)$ and $Y(t + \Delta t)$.

In order to find the corresponding indirect effect of $X(t)$ on $Y(t + \Delta t)$ through intermediate values of M, A_{31} is set to zero, leading to

$$\phi_{31}^{I}(\Delta t) = -\frac{a_{21}a_{32}(-a_{22}e^{a_{11}\Delta t}) + a_{11}e^{a_{22}\Delta t} + a_{22}e^{a_{33}\Delta t} - a_{33}e^{a_{22}\Delta t} - a_{11}e^{a_{33}\Delta t} + a_{33}e^{a_{11}\Delta t})}{(a_{22} - a_{11})(a_{22} - a_{33})(a_{11} - a_{33})} \tag{12b}$$

as the indirect effect, where the superscript I denotes that this is the value of $\phi_{31}(\Delta t)$ resulting from only indirect pathways. Summing both these terms leads to the total effect, which can be shown to equal $\phi_{31}(\Delta t)$ derived from an unaltered drift matrix.

This definition of path specific effects fits within the tradition of path analysis, and possesses intuitively appealing and logical characteristics in comparison to the DT concept of direct, indirect and total effect. For instance, if $a_{31} = 0$ this suggest that the rate of

[8] The matrix exponential function is also defined as taking the (weighted) sum of increasing powers of the target matrix.

change of the process $Y(t)$ is independent of the value of $X(t)$ given the value of $M(t)$, that is, there is no direct link between $X(t)$ and $Y(t)$. However, the corresponding discrete-time parameter $\phi_{31}(\Delta t)$ will in general not be zero if there are indirect paths between $X(t)$ and $Y(t)$, (i.e., $a_{21} \neq 0$ and $a_{32} \neq 0$), calling into question the interpretation of this parameter as a *direct effect*. In this conceptualization, the lagged direct effect of $X(t)$ on $Y(t + \Delta t)$ will only be non-zero when the process Y is time-locally dependent on the process X.

3.4 Interventions and CT mediation: current state of affairs

Numerous authors have shown that an interventionist approach can be used to assess the effects of interventions on dynamic systems which evolve continuously over time (Didelez, 2008; Eichler & Didelez, 2012; Røysland et al., 2012; Sokol & Hansen, 2014). However, while Deboeck & Preacher (2016) and Aalen et al. (2012, 2018) both offer alternative conceptualizations of total, direct and indirect effects for CT-VAR(1) systems, they do so strictly within the path-tracing framework. In the case of Deboeck & Preacher (2016), no connection is drawn to interventionist causal framework.Aalen et al. (2012) posits that it is not useful to define these effects in an interventionist setting, arguing that a mechanistic approach which focuses on path-tracing rather is more suitable in the context of CT models.

This current state-of-affairs is lacking for a number of reasons. First, an interventionist treatment of CT mediation would be hugely beneficial in providing a clear and (hypothetically) falsifiable definition of total, direct and indirect effects, something which is lacking in path-tracing approach. If path-specific effects in CT models could be defined in terms of variable interventions, it would be possible to make use of other tools from the interventionist literature to help in the identification of these effects.

Second, no general approach to calculating path-specific effects in CT models is given in the literature: with more than three variables, and a non-triangular drift matrix, there are numerous possible combinations of paths which could be omitted or included in a path-tracing direct or indirect effect. An interventionist definition of these effects would

help clarify this calculation greatly.

Third, it is difficult to reconcile the claim of Deboeck & Preacher (2016) that the total effect from DT models in fact represents a direct effect, with the interventionist treatment of these effects from VanderWeele & Tchetgen Tchetgen (2017). In the current paper we have shown the clear links between the DT- and CT-VAR(1), as well as the links between the interventionist and path-tracing definitions of mediation in the DT-VAR(1). As such, it is not clear in what sense the DT direct effect, given as DE_{Δ_τ} given in Equation (6a), can be construed as misleading or incorrect. These three shortcomings of the current literature are addressed comprehensively in the next section.

4. Variable-setting Interventions in CT systems

The goal of this section is to investigate whether each of the total, direct and indirect effects derived from path-tracing rules for CT models can be defined as variable-setting interventions in CT systems. In doing so, the relationship between variable-setting operations on CT systems, and the total, direct and indirect effects defined for DT systems in section 2, is also elucidated. As the equivalence between path-tracing and variable setting operations in DT systems has already been shown, these are referred to simply as DT mediation effects. Finally, after treating each of the total, direct and indirect in turn, we will re-examine the argumentation of Deboeck & Preacher (2016) regarding the interpretation of DT direct and total effects, described above.

This section introduces two different types of variable-setting operations on CT systems: *acute* interventions and *interval* interventions. Both refer to interventions on some putative mediator: the acute intervention refers to setting the value of a mediator to a constant at one or more particular moment(s) in time t; the interval intervention refers to setting the value of a mediator to a constant at every moment in time, over an interval Δt. Using this terminology, two different *direct effects* and two different *indirect effects* can be defined, each reflecting both an *acute* and *interval* interventions.

It is shown that the CT path-tracing direct effect reflects an interval intervention,

and that the DT direct effect reflect an acute intervention. It is further shown that the CT path-tracing indirect effect reflects an interval intervention, defined using *extended causal graphs*, analogous to those shown in sections 1 and 2. However, it is also shown that the DT indirect effect does not reflect any intervention in a CT system, as it is derived from a different extended causal graph.

The total, direct and indirect effect, based on variable-setting interventions in CT systems, are each described in turn in the following subsections. The definitions and calculation of direct and indirect effects, using path-tracing rules and using variable-setting interventions, as well as the proof of their equivalence, is given in Appendix A (for direct effects) and Appendix B (for indirect effects). The latter also describes the construction of indistinguishable extended local independence graphs, used in defining the variable-setting indirect effect. R functions for the calculation of total, direct and indirect effects in the CT-VAR(1) are given in the supplementary materials. In the main text the definitions of the total, direct and indirect effects are given briefly, along with their substantive interpretation, and an illustration using a numerical example. The numerical example is taken from Deboeck & Preacher (2016), with drift matrix

$$\boldsymbol{A} = \begin{bmatrix} -.357 & 0 & 0 \\ .771 & -.511 & 0 \\ -.450 & .729 & -.693 \end{bmatrix} \tag{13a}$$

and corresponding auto-regressive and cross-lagged parameters, for $\Delta t = 1$

$$e^{\boldsymbol{A}*1} = \boldsymbol{\Phi}(\Delta t = 1) = \begin{bmatrix} .7 & 0 & 0 \\ .5 & .6 & 0 \\ -.1 & .4 & .5 \end{bmatrix}. \tag{13b}$$

4.1 Total Effects

Recall the definition of total effects in a DT-VAR(1) system from Equation (5a). We can define an equivalent total effect given a CT data-generating system, which will be a function of the time-interval length, Δt. Let us define the *total effect* of $X(t)$ on $Y(t+\Delta t)$ as the effect on $Y(t+\Delta t)$ of intervening to set the value of $X(t)$ from x^* to x, given as $TE(\Delta t)$. This total effect is defined as

$$\begin{aligned} TE(\Delta t) &= E[Y(t+\Delta t)|X(t)=x] - E[Y(t+\Delta t)|X(t)=x^*] \\ &= \phi_{31}(\Delta t)(x-x^*). \end{aligned} \tag{14}$$

Comparing this to the DT total effect in Equation (5a) we can see that the DT total effect is equivalent to the CT total effect.[9] In general, the DT and CT total effects will be equal for any positive integer multiplication of Δt. For instance, with our example mediation system in Equations (13a) and (13b), the DT total effect would be calculated as

$$\begin{aligned} TE_2 &= E[Y_{\tau+2}|X_\tau = x] - E[Y_{\tau+2}|X_\tau = x^*]. \\ &= \mathbf{\Phi}^2\langle 3,1\rangle(x-x^*) \end{aligned}$$

where $\langle 3,1\rangle$ denotes the third row and first column of $\mathbf{\Phi}^2$. Now, recalling that $\mathbf{\Phi} = e^{\boldsymbol{A}\Delta t}$, and substituting

$$\begin{aligned} TE_2 &= (e^{\boldsymbol{A}\Delta t})^2\langle 3,1\rangle(x-x^*) \\ &= e^{\boldsymbol{A}2\Delta t}\langle 3,1\rangle(x-x^*) \end{aligned}$$

[9] Here, and throughout, it should be emphasized that this is only the case when the time-interval between subsequent observations is equal to Δt, that is, $\mathbf{\Phi} = \mathbf{\Phi}(\Delta t)$.

which, comparing this expression to Equation (14), shows that

$$TE_2 = TE(2\Delta t)$$

and in general that

$$TE_{\Delta_\tau} = TE(\Delta_\tau \Delta t)$$

for whatever Δt which is reflected by $\mathbf{\Phi}$. In other words, the CT total effect describes the same intervention, with the same result, as the total effect derived from the DT system.

This is the case provided that Δt is reflects an integer multiple of the time that elapses between occasions, Δt_τ, for the DT system. As such, an important difference between the DT and CT total effects is that the DT model does not lend itself to predictions or inferences regarding the total effects at a time-interval shorter than that used in data collection, or at non-integer multiples of that time interval ($\Delta t = .8$ or $\Delta t = 1.3$ etc.) whereas the CT model does allow us to make predictions about the effects of interventions at shorter time-intervals.

We can visualize the total effect as the difference between the position of two different trajectories of the Y process, at some point in time $t + \Delta t$, given two different starting conditions for the X process, $X(t)$. These trajectories are shown in Figure 4 with starting values of $x^* = 1$, $x = 2$, and $M(t) = Y(t) = 0$. Let us denote $Y(1)|X(0) = x^*$ as $Y^*(1)$. The total effect of setting $X(0)$ from x^* to x on $E(Y(1))$, given as $TE(1)$, is shown as the vertical distance between $Y(1)$ and $Y^*(1)$, denoted by the line connecting both red dots at $t = 1$. In this case $TE(1) = -.1$, exactly equal to ϕ_{31} in Equation (13b). We further show the total effect for $\Delta t = 2$, that is, $TE(2) = .08$. Note that the total effect at the longer time interval is now positive rather than negative, showing that it is not only the *magnitude* of the total effect which is a function of the time-interval, but also the *direction* in this case.

4.2 Direct effects

To define different direct effects in an interventionist framework for CT systems, it will prove useful to distinguish between two kinds of interventions on a putative mediator: *acute* and *interval* interventions. It is then shown that the path-tracing direct effect from the DT model is equivalent to a particular *acute* intervention on the mediator. We then show that the path-tracing direct effect from the CT model in fact represents an *interval* intervention.

4.2.1 Acute Interventions and DT direct effects: a simple case. The direct effect described in Equation 6a in the context of a DT-VAR(1) model (i.e. CDE_{Δ_τ}) refers to the effect on Y at occasion $\tau + \Delta_\tau$ of an intervention to set the value of X_τ from x^* to x, combined with interventions to set the value of M to m at each equally spaced intermediate measurement occasion, that is, some discrete moments in time, $\tau + 1, \ldots \tau + (\Delta_\tau - 1)$. For example, the CDE_2 consisted of the aforementioned intervention on X_τ, combined with interventions to set the value of $M_{\tau+1}$. We can define an equivalent kind of direct effect as the CDE_{Δ_τ}, but for a CT system. As was the case for the total effect, the value of this direct effect will depend on the length of the time interval Δt between observations, as well as the specific *acute moments in time* at which the mediator is intervened on.

Let us take it that we are interested in the difference in expected value of $Y(t + \Delta t)$, given an intervention to set $X(t)$ from x^* to x, combined with an intervention to set the value of the mediator M at exactly one moment in time, half way between t and $t + \Delta t$, given as $M(t + \Delta t/2)$, to a value of m. This would be given by the difference between two expected values,

$$E[Y(t+\Delta t)|X(t) = x, M(t + \frac{\Delta t}{2}) = m] - E[Y(t+\Delta t)|X(t) = x^*, M(t + \frac{\Delta t}{2}) = m] \quad (15)$$

where we will define this difference as an *acute direct effect* or $aCDE$, as it refers to an intervention on the mediator at an acute moment in time, $M(t + \Delta t/2)$. We will denote this

direct effect as $aCDE(\Delta t, 1)$, where Δt refers to the time interval of interest, and the one refers to the number of *equally spaced intermediate occasions* at which we intervene on M.

We can evaluate this expression by plugging the CT-VAR(1) parameters for the expected values above. This leads to

$$aCDE(\Delta t, 1) = (\phi_{11}(\Delta t/2)\phi_{31}(\Delta t/2) + \phi_{31}(\Delta t/2)\phi_{33}(\Delta t/2))(x - x^*). \tag{16}$$

We can see that this expression exactly equivalent to the discrete-time CDE_2, given that $\mathbf{\Phi} = \mathbf{\Phi}(\Delta s)$, where in this case $\Delta s = \Delta t/2$. For instance, taking $\Delta t = 2$, we get

$$\begin{aligned} aCDE(2, 1) &= (\phi_{11}(1)\phi_{31}(1) + \phi_{31}(1)\phi_{33}(1))(x - x^*) \\ &= (\phi_{11}\phi_{31} + \phi_{31}\phi_{33})(x - x^*) \\ &= CDE_2 \end{aligned}$$

showing that the DT direct effect in fact reflects this specific acute intervention on the mediator in the CT system. Using our numerical example, we can visualize $aCDE(2, 1) = -.12$ as the difference between the positions of $Y(2)$ and $Y*(2)$ in Figure 5a, with values of $x^* = 1$, $x = 2$, with $Y(0) = 0$ and $M(0) = 0$. We can see from Figure 5a that the value of $M(1)$ and $M^*(1)$ is set to zero; this effects the rate of change of $Y(2)$ and $Y^*(2)$, and results in the vertical distance $Y(2) - Y^*(2)$ being equal to exactly $-.12$.

4.2.2 Acute Interventions and DT direct effects: the general case. We can generalize our definition of the acute controlled direct effect, by supposing that we wish to intervene on the mediator at q equally spaced intermediate measurement occasions. We can then define the Acute Controlled Direct Effect or $aCDE(\Delta t, q)$ as the expected change in $Y(t + \Delta t)$ given that we intervene to set $X(t)$ from x^* to x, and intervene to set the values of $M(t + \frac{p\Delta t}{q+1})$ for all integer values of p from one to $q + 1$ to m, given as $\{M(t + \frac{p\Delta t}{q+1}\} = m$. For example, if $q = 2$, we would intervene to set

$M(t + \frac{1\Delta t}{3}) = M(t + \frac{2\Delta t}{3}) = m$. This is given in the general case as

$$aCDE(\Delta t, q) = E[Y(t + \Delta t)|X(t) = x, \{M(t + \frac{p\Delta t}{q+1})\} = m] - E[Y(t + \Delta t)|X(t) = x^*, \{M(t + \frac{p\Delta t}{q+1})\} = m]. \quad (17)$$

Comparing this to the corresponding DT direct effect, CDE_{Δ_τ}, it is clear that they are equal whenever Δt is equal to Δt_τ, and q is equal to $\Delta_\tau - 1$. That is, as we showed above, $aCDE(2, 1) = CDE_2$, but also $aCDE(3, 2) = CDE_3$, $aCDE(4, 3) = CDE_4$ and so forth.

Analogously to the total effect, the $aCDE(\Delta t, q)$ is a more general quantity than the discrete-time CDE_{Δ_τ} as it allows us to make predictions about the effects of interventions for any choice of Δt or q. So for example, we can make predictions about what would happen to the expected value of $Y(t + 2)$, given an intervention to set $X(t)$ from x^* to x, combined with an intervention to set $M(t + \frac{2}{4}) = M(t + \frac{4}{4}) = M(t + \frac{6}{4})$. This set of interventions would correspond to $aCDE(2, 3)$, visualized in Figure 5b. Here we can see that the vertical distance corresponding to the value of $aCDE(2, 3)$ has increased along with the number of acute interventions on the mediator, $aCDE(2, 3) = -.22$.

4.2.3 Interval Interventions and CT direct effects. Now that we have defined the general $aCDE(\Delta t, q)$ as reflecting intervention(s) on the mediator at q *acute* moments in time, we can derive an expression for an intervention to set the mediator to a constant at every moment in time over an interval. Let us define the history of process M from time point t until time point $t + \Delta t$, given as $\bar{M}(t, t + \Delta t)$, as the set of values that M takes on at every point in time between t and $t + \Delta t$; that is, $\bar{M}(t, t + \Delta t)$ contains the value of $M(s)$ for $t \leq s \leq t + \Delta t$, $s \in \mathbb{R}$.

Thus we can define the *interval controlled direct effect* $iCDE(\Delta t)$ as the difference in expected value of $Y(t + \Delta t)$ given an intervention to change $X(t)$ from x^* to x, while

intervening to keep $\bar{M}(t, t+\Delta t)$ equal to a constant m. This is given as

$$iCDE(\Delta t) = E[Y(t+\Delta t)|X(t) = x^*, \bar{M}(t, t+\Delta t) = m] - E[Y(t+\Delta t)|X(t) = x, \bar{M}(t, t+\Delta t) = m] \tag{18}$$

The interval CDE is the limiting form of the $aCDE(\Delta t, q)$ as the number of acute moments at which we intervene on the mediator, given by q, approaches infinity

$$iCDE(\Delta t) = \lim_{q\to\infty} aCDE(\Delta t, q).$$

In Appendix A it is shown that this variable-setting intervention is exactly equal to the *path-tracing direct effect* suggested by Deboeck & Preacher (2016) and Aalen et al. (2012), and described in section 3.3. The appendix also describes general procedures for finding both the path-tracing direct effect and equivalently the $iCDE(\Delta t)$.

We can illustrate the interval $iCDE(\Delta t)$, its interpretation, and its equivalence to the path-tracing CT direct effect, hereby referred to as $pDE(\Delta t)$, using our numerical example. From the numerical example, we can calculate $pDE(2) = .321$. In Figure 5c we have plotted the trajectories of our processes under $aCDE(2, 9)$ interventions, that is, with 9 acute interventions on the mediator, and in Figure 5d we show the trajectories of our processes under $iCDE(\Delta t)$. In both cases we use the same numerical example and starting values as for the previous figures, with $m = 0$. First, we can note that, as shown in Figure 5d, the $iCDE(2)$ is exactly equal to $pDE(2)$, $iCDE(2) = pDE(2) = .321$.

Comparing the trajectories of $Y(t)$ and $Y^*(t)$ in Figures 5a, 5b and 5c, we see that, as q increases, the trajectories, and the difference between them, become more and more similar to those in Figure 5d. It is clear then that the $iCDE(2)$ not only represents a related, but qualitatively different intervention on our system as the $aCDE(2, 2)$, but also results in a much larger difference in $Y(t + 2)$, $iCDE(2) = -.321$, $aCDE(2, 2) = -.22$. In general these effects may be of different signs as well as different absolute values.

4.4.4 A note on Deboeck & Preacher (2016). Now that we have defined the $aCDE(\Delta t)$ and $iCDE(\Delta t)$, and shown their equivalence to the DT direct effect and CT path-tracing direct effect respectively, we can re-examine the argument of Deboeck & Preacher (2016) that $\phi_{31}(\Delta t)$, should be interpreted as a total effect rather than a direct effect. It can now be clarified that the authors were only correct in part: indeed $\phi_{31}(\Delta t)$ does not reflect the $iCDE(\Delta t)$, as it consists of pathways through latent values of the mediator M for $M(s)$, $t < s < t + \Delta t$. However, $\phi_{31}(\Delta t)$ does reflect another type of intervention-based direct effect, consisting of interventions on $M(t)$ and $M(t+\Delta t)$, given as

$$\begin{aligned}\phi_{31}(\Delta t)(x - x^*) =& E[Y(t + \Delta t)|X(t) = x, M(t) = M(t + \Delta t) = m] \\ &- E[Y(t + \Delta t)|X(t) = x^*, M(t) = M(t + \Delta t) = m].\end{aligned}$$

It happens that, as $M(t)$ is independent of $X(t)$, and as $Y(t + \Delta)$ is independent of $M(t + \Delta t)$, that this set of interventions produces the same change in expected value as the total effect intervention

$$\phi_{31}(\Delta t)(x - x^*) = E[Y(t + \Delta t)|X(t) = x] - E[Y(t + \Delta t)|X(t) = x^*].$$

However it should also be noted that this is entirely consistent with the interpretation of this parameter in DT settings: it is clear form section 2.2.2 that $TE_1 = CDE_1$. In the general case (i.e. for $\Delta_\tau > 1$), it is clear from above that both DT and CT path-tracing formulations of direct effects are correct, but simply refer to qualitatively different types of interventions on the mediator.

4.3 Indirect Effects

As was the case in defining indirect effects as interventions in a cross-sectional (section 1.1) and DT-VAR(1) system (section 2.2.3), to define indirect effects as interventions in a CT system we must posit that the X variable can be decomposed into

two components. These two components are X^a, on which only Y is time-locally dependent; and X^b, on which only M is time-locally dependent. This decomposition can be represented using extended causal graphs: the extended DAG is shown in Figure 6a, and the corresponding extended LIG is shown in Figure 6b. In this extended system, the parent process $X(t)$ has auto-regression, but the child processes $X^a(t)$ and $X^b(t)$ do not: The value of these at each moment in time is determined completely by the parent process, except in the case of an intervention to set the value of $X^a(t)$ and $X^b(t)$ to a constant.

The drift matrix for this extended LIG is given by

$$\boldsymbol{A'} = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ 0 & a_{11} & 0 & 0 \\ 0 & a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{32} & a_{33} \end{bmatrix} \tag{19a}$$

where the parameters are given in terms of their positions in the original drift matrix. In this matrix the parent process $X(t)$ is omitted, and the first and second row contain the local-dependency paths entering (indirectly through $X(t)$) and leaving $X^a(t)$ and $X^b(t)$ respectively. As in the cross-sectional and DT cases, this extended drift matrix represents a statistically indistinguishable model from the unextended drift matrix: In the absence of any intervention to set them to different values, $X^a(t)$ is always equal to $X^b(t)$ (for details see Appendix B).

Taking the matrix exponential of this extended drift matrix we come to

$$\boldsymbol{\Phi'}(\Delta t) = \begin{bmatrix} \phi_{11}(\Delta t) & 0 & 0 & 0 \\ 0 & \phi_{11}(\Delta t) & 0 & 0 \\ 0 & \phi_{21}(\Delta t) & \phi_{22}(\Delta t) & 0 \\ \phi^a_{31}(\Delta t) & \phi^b_{31}(\Delta t) & \phi_{32}(\Delta t) & \phi_{33}(\Delta t) \end{bmatrix} \tag{19b}$$

where the parameters are given again in terms of their positions in the original $\boldsymbol{\Phi}(\Delta t)$

matrix, and again the first and second rows and columns represent X^a and X^b respectively. Here we have denoted the cross-lagged parameter from $X^a(t)$ to $Y(t + \Delta t)$ as $\phi^a_{31}(\Delta t)$, and we have denoted the cross-lagged parameter from $X^b(t)$ to $Y(t + \Delta t)$ as $\phi^b_{31}(\Delta t)$. Both parameters sum to the parameter $\phi_{31}(\Delta t)$ from the unaltered drift matrix, i.e. $\phi^a_{31}(\Delta t) + \phi^b_{31}(\Delta t) = \phi_{31}(\Delta t)$ (Deboeck & Preacher, 2016; Aalen et al., 2012).

4.3.1 Interventions and the DT Indirect Effect. Now we can examine whether the natural indirect effect from the DT model, NIE_{Δ_τ} from Equation (9a), correctly describes an intervention in a CT system. Recalling the definition of the NIE_{Δ_τ}, we can define a conceptually similar indirect effect on a CT system as the effect on $Y(t + \Delta t)$ of changing $X(t)$ from x^* to x, while intervening to keep $X^a(t)$ at a value of x^*. This is given as

$$NIE(\Delta t) = E[Y(t+\Delta t)|X(t) = x, X^a(t) = x^*] - E[Y(t+\Delta t)|X(t) = x^*, X^a(t) = x^*]. \quad (20)$$

Note that this intervention sets only the value of X^a at time t to zero, and no other values of X^a. Unlike the $aCDE(\Delta t, q)$, this path-specific effect is not equivalent to its discrete time equivalent, the NIE_{Δ_τ}, which in Equation (9b) was given as $\phi_{21}\phi_{32}(x - x^*)$.

This is because the posited extended CT system in Figure 6 is qualitatively different than the extended DT system in Figure 2b. In the extended CT system, $X(s)$ determines the value of $X^a(s)$ at all points in time $s > t$, whereas in the extended DT system, the X_τ determines the values of X^a at occasions X^a_τ: this is not compatible with a CT system as it posits some kind of discontinuity in $X^a(t)$ in between discrete occasions. The substantive notion that the system of interest varies continuously over time leads in turn to a different substantively meaningful extended causal system. The drift matrix for the extended CT system in general will not result in the same matrix of cross-lagged and auto-regressive parameters (i.e. $\mathbf{\Phi'}(\Delta t)$) as the extended DT system (i.e. $\mathbf{\Phi'}$) except in the trivial case of a lower triangular drift matrix, in which X and M, or M and Y are locally independent (i.e.

$a_{21} = 0$ or $a_{32} = 0$).

Due to this continuous influence from $X(s)$ to $X^a(s)$, the latter will be exactly equal to the former for all $s > t$. As such, $NIE(\Delta t)$ will be approximately zero except for infinitesimally small values of Δt. In other words the value that X^a takes on immediately after time point t will be identical in both scenarios, so the effect of setting X^a to a different value at the time point t on the trajectory of the Y process will be negligible.

4.3.2 Interval Interventions and CT Indirect Effects. The path-tracing indirect effect of Deboeck & Preacher (2016) and Aalen et al. (2012), described in section 3.4, can be shown to be equivalent to an interval version of the $NIE(\Delta t)$ defined above. While the general definition and calculation of the path-tracing CT indirect effect is given in Appendix B, let us here denote this simply as $pIE(\Delta t)$.

We can now define the *interval* natural indirect effect over an interval Δt, denoted $iNIE(\Delta t)$, as the effect on the expected value of $Y(t + \Delta t)$ of intervening to change $X(t)$ from x^* to x, while intervening to keep $\bar{X^a}(t, t + \Delta t)$ fixed at some constant c. This is given as

$$\begin{aligned} iNIE(\Delta t) =& E[Y(t + \Delta t)|X(t) = x, \bar{X^a}(t, t + \Delta t) = c] \\ & - E[Y(t + \Delta t)|X(t) = x^*, \bar{X^a}(t, t + \Delta t) = c]. \end{aligned}$$

This definition follows from the fact that the $iNIE(\Delta t)$ can be seen as equivalent to an interval direct effect $iCDE(\Delta t)$ in the extended system, with X^a treated as the mediator. As such, the equivalence of the path-tracing and variable-setting definitions of indirect effects

$$iNIE(\Delta t) = pIE(\Delta t) \tag{21}$$

also follows immediately from the proofs in Appendix A.

This equivalence between path-tracing variable-setting indirect effects be illustrated with our numerical example. In this case, $pIE(2) = .401$, following from the calculation

described in section 3.4 and Appendix B. In Figure 7 we have plotted the trajectories of our processes under the interventions described by $iNIE(\Delta t)$, with $x = 2$, $x^* = 1$, $c = 0$, with $M(0) = 0$ and $Y(0) = 0$ respectively. The vertical distance between $Y(2)$ and $Y^*(2)$ represents $iNIE(2) = .401$, exactly the parameter acquired for $psIE(2)$. Furthermore, we can see that the interval indirect effect $iNIE(2) = .401$ and interval direct effect $iCDE(2) = -.321$ sum to the total effect, $TE(2) = .08$.

4.3.3 On the interpretation of extended LIGs. The treatment of indirect effects using extended causal systems is a natural extension of Robins (2003) and Robins & Richardson (2010) to a continuous-time dynamic setting. However it should be stressed again that the existence of such an extended system, and the accompanying decomposition of the $X(t)$ process, must be substantively motivated in each case. This can be illustrated with an example: take $X(t)$ to be stress, $M(t)$ to be anxiety, and $Y(t)$ to be physical discomfort. We could posit an extended causal system in which we decompose stress into two components: a hormonal or biological response ($X^a(t)$) which tends to cause physical discomfort directly by inducing sweating and shaking, and a cognitive component, $X^b(t)$, which influences feelings of anxiety; high levels of anxiety in turn result in higher levels of perceived physical discomfort. In normal settings, both components are co-present as stressful situations induce a biological and cognitive response. Suppose that (at least theoretically), we could apply some treatment to inhibit the biological component $X^a(t)$ from activating in response to stressful scenarios. We could then ask the question: If we could stop the experience of stress releasing hormones which cause a physical reaction (i.e. set $\bar{X^a}(t, t + \Delta t) = 0$), what would the effect of increasing stress $X(t)$ be on feelings of physical comfort some time later $Y(t + \Delta t)$? We could make predictions about the effect of such an intervention using the $iNIE(\Delta t)$ from the extended drift matrix. However, whether this effect is meaningful in any way depends entirely on the plausibility of the extended causal system, which must be assessed on a case-by-case basis.

5. Conclusions

In this paper we have shown how path-tracing definitions and interventionist definitions of total, direct and indirect effects relate to one another in a CT system. We have introduced the terminology of *acute* and *interval* interventions to describe different interventions and effects we can derive. We can summarize our specific findings as follows

1. The total effect derived from DT path-tracing is equivalent to the total effect derived from CT path-tracing. Both of these reflect the effect of an intervention to change an initial value of $X(t)$ on the expected value of $Y(t + \Delta t)$, given as $TE(\Delta t)$

2. The direct effect from DT path-tracing rules describes a particular case of an *acute* intervention on the mediator, $aCDE(\Delta t, 2)$. That is, the effect on $Y(t + \Delta t)$ of changing the value of $X(t)$, combined with an intervention to set $M(t + \Delta t/2)$ to a constant.

3. The direct effect from CT path-tracing represents an *interval* intervention on the mediator, $iCDE(\Delta t)$. That is, the effect on $Y(t + \Delta t)$ of changing the value of $X(t)$, while holding the value of the mediator M to a constant at every moment in time in that interval.

4. DT and CT direct effects differ, both numerically and qualitatively, as they describe quite different interventions on a CT system.

5. Indirect effects can be defined in terms of interventions on a CT system, making use of a extended causal graphs. This graph represents our notion that the X variable can be decomposed into an $M-$causing and a $Y-$causing component

6. The indirect effect from CT path-tracing represents an *interval* intervention on the $Y-$ causing component of our system, X^a. That is, the effect on $Y(t + \Delta t)$ of changing the value of $X(t)$, and intervening to hold X^a to a constant at every moment in time in that interval

While the interpretation of both acute and interval, direct and indirect effects are illustrated in the main text by use of a tri-variate mediation model, with lower triangular drift matrix, each of the statements above are proved in the general case, that is, for a drift matrix with any number of processes and potentially including reciprocal relationships, in Appendix A (for direct effects) and Appendix B (for indirect effects). Appendix A and B further describe the calculation of each type of direct and indirect effect respectively, with the latter also describing the creation of suitable extended graphs and drift matrices. Although the main text focuses on the case of a single mediating variable, this is generalizable in a straightforward way to the case of multiple mediators.

6. Discussion

In this paper it has been shown how total, direct and indirect effects in CT-VAR(1) models can be treated in terms of the interventionist causal inference framework. The connection between path-tracing and interventionist accounts of mediation in both DT- and CT-VAR(1) models has been elaborated on, and we have outlined the similarities and dissimilarities that arise when treating causal mechanisms as DT or CT in nature. R functions which allow researchers to easily calculate the CT total, direct, and indirect effects described in this paper can be found in the supplementary materials.

The major benefit of this work is that it will allow researchers to make more informed predictions about the effects of different types of interventions. This is of particular benefit to the quickly emerging m-health and e-health approaches in the social, behavioral and health sciences (Mohr et al., 2014; Naslund et al., 2015). M-health technology allows researchers to collect large amounts of intensive longitudinal data about behavioral or psychological processes, as well as allowing for the administration and monitoring of interventions on these processes. Using this technology, some researchers have been using experience-sampling data to design personalized interventions in psychological systems (e.g., Kramer et al., 2014; Hartmann et al., 2015). Other researchers have been conducting micro-randomized trials to design interventions which most effectively achieve positive

changes in behavior or cognition, also known as just-in-time interventions (Sarker et al., 2014, 2017). The interventionist approach to CT mediation analysis described in the current paper could potentially aid researchers in more accurately predicting the effects of those interventions, as well as identifying what part of the system should be targeted for intervention.

The significance of this work goes beyond studies in which a specific mediation hypothesis, or the design of interventions, is of primary interest. In the network approach to psychopathology, researchers approach psychopathologies as large systems of symptoms which mutually exert direct causal influence on one another (Borsboom & Cramer, 2013; McNally et al., 2015; Epskamp et al., in press). Network structures further allow for the investigation of multivariate relationships through the use of centrality measures, which quantify the contribution of different symptoms, directly and indirectly, to the system as a whole. When experience sampling data of symptoms are available, these network structures are typically estimated using DT-VAR(1) models, in which the cross-lagged parameters are treated as direct causal links between symptoms, and centrality measures are a function of these parameters (cf. Bringmann et al., 2013, 2015, 2016). However, as has been detailed in the current paper, the interpretation of these parameters as direct effects may be questionable. Furthermore, the interventionist approach to total and indirect effects may lead to new ways of approaching centrality measures in such systems using a CT approach. As such, a CT-VAR(1) approach to network analysis, combined with an interventionist view of causal mechanisms, is a promising avenue for further research.

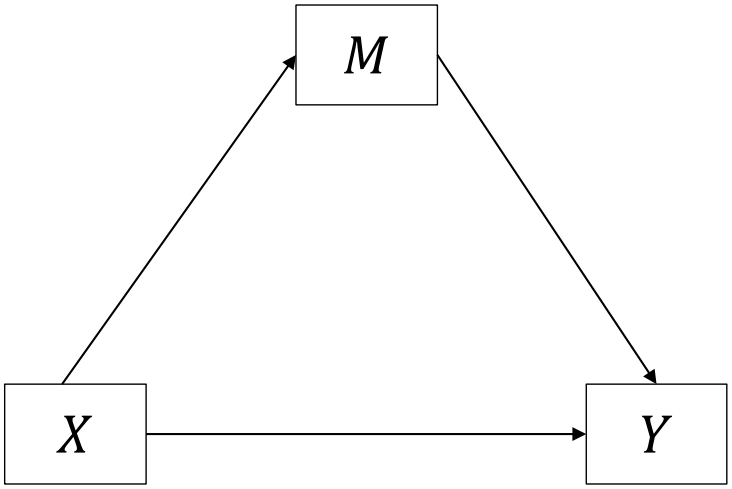
The use of CT approaches for causal inference in dynamic systems is appealing as such models reflect our conceptual ideas about how systems evolve and vary over time. These processes are likely to take on values, and exert influence on one another, continuously and smoothly over time, not in discrete jumps (Boker, 2002). As such, the use of CT models over DT models reflects the use use of substantive knowledge in combination with observed data (Aalen et al., 2018). Amongst others, Didelez (2008), Eichler & Didelez

(2012), Røysland et al. (2012) and Sokol & Hansen (2014) have presented a formal interventionist treatment of causal inference in general CT systems in much greater detail, often utilizing stochastic process theory. In contrast, a simplified approach to this topic, using only the basic principles of the interventionist approach and dynamic systems is given in this paper, to facilitate a more accessible analysis of CT mediation. Future research should make greater use of this literature, for instance to derive identifiability conditions for the effects described in the current paper.

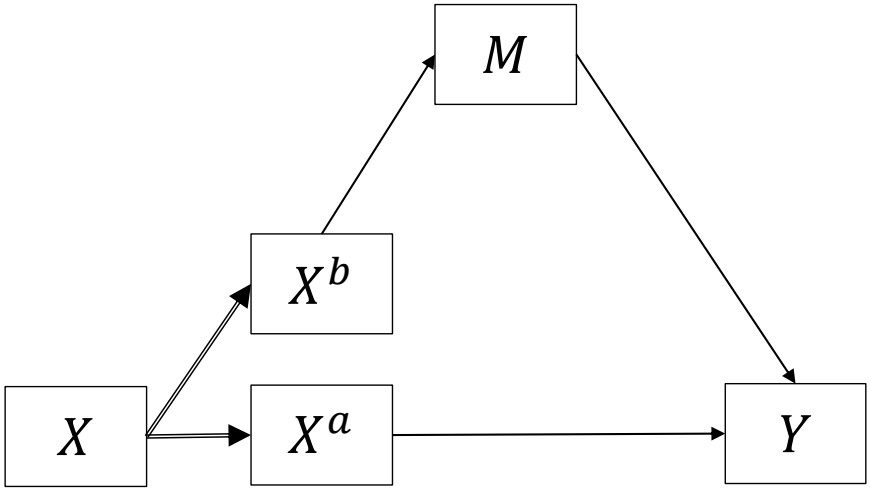
In this paper the discussion of the DT- and CT-VAR(1) models have been kept as general as possible. As such, the characteristics of different types of longitudinal data to which these models can be applied has not been elaborated on. Typically, the VAR(1) terminology is used in the context of single-subject data, while the terminology of CLPMs is used in the context of multiple-subject panel data. In the latter case, it is generally recommended that researchers make use of multi-level extensions, to distinguish within and between participant variance, and to allow for differences in both means and lagged parameters between participants (Hamaker et al., 2015). While multi-level extensions of CT models have been described (Oravecz et al., 2009; Driver & Voelkle, 2017), further research is needed to examine the issue of causal inference in these applications (cf. Vansteelandt, 2007). Furthermore, intensive longitudinal data can take on many different forms: In the case of self-report data, questions may take the form of daily diary entries, momentary assessments, or retrospective questions based on experiences since the last measurement. Research on the applicability or inapplicability of CT models to these different data types is necessary to move the conversation regarding CT versus DT models forward.

The development of new estimation packages means that the application of the CT modeling approach has never been easier (Boker et al., 2004; Boker, Deboeck, et al., 2010; Driver et al., 2017; Ou et al., 2017). The call for a more widespread adoption of CT models is matched by a growing number of applications of these models in substantive areas (e.g., Toharudin et al., 2008; Boker, Montpetit, et al., 2010; Steele & Ferrer, 2011; Angraini et

al., 2014; Chow et al., 2018). For many researchers, however, a knowledge gap remains: How can CT models be interpreted, and how are they related to the DT approaches which are still most prevalent? By connecting traditional social science methodology, interventionist causal inference, and CT dynamical systems, the current paper aids in bridging this gap, and contributes to pushing the CT modeling approach further towards the forefront of longitudinal modeling in the social and behavioral sciences sciences.

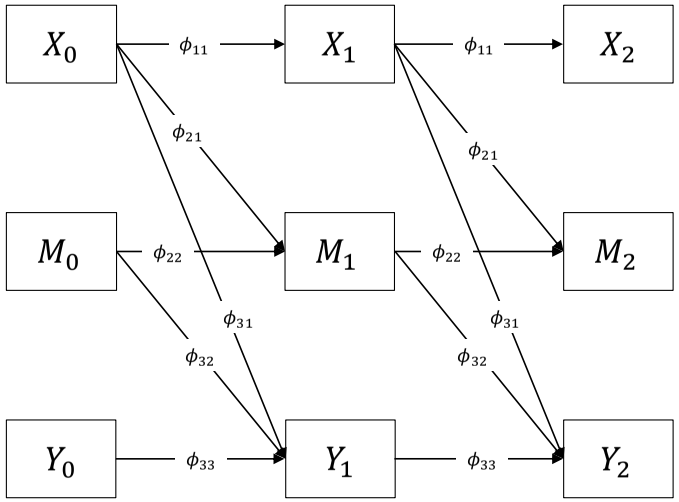


(a) DAG

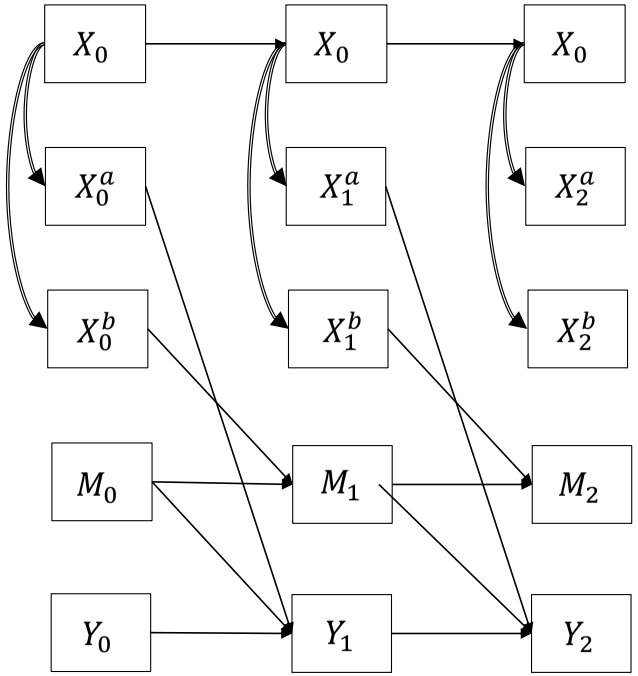


(b) Extended DAG

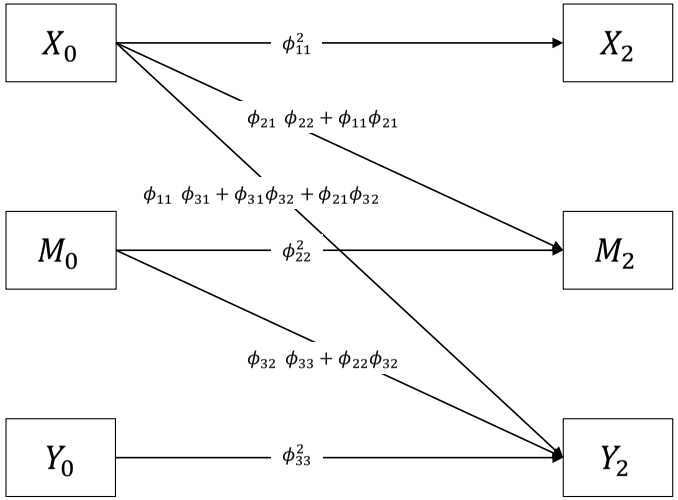
Figure 1. DAGs representing a cross-sectional mediation structure



(a) DT-VAR(1) mediation process with interval Δt

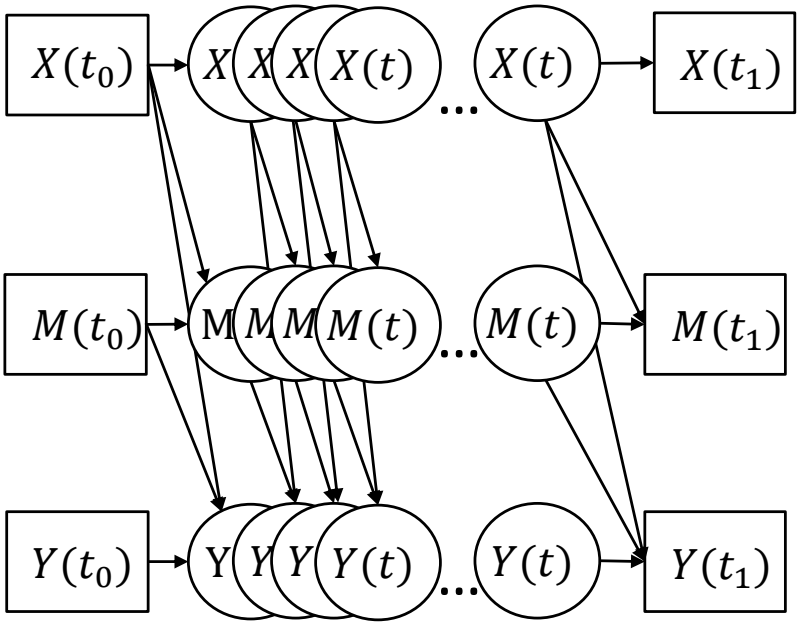


(b) Extended graph for DT-VAR(1) mediation process

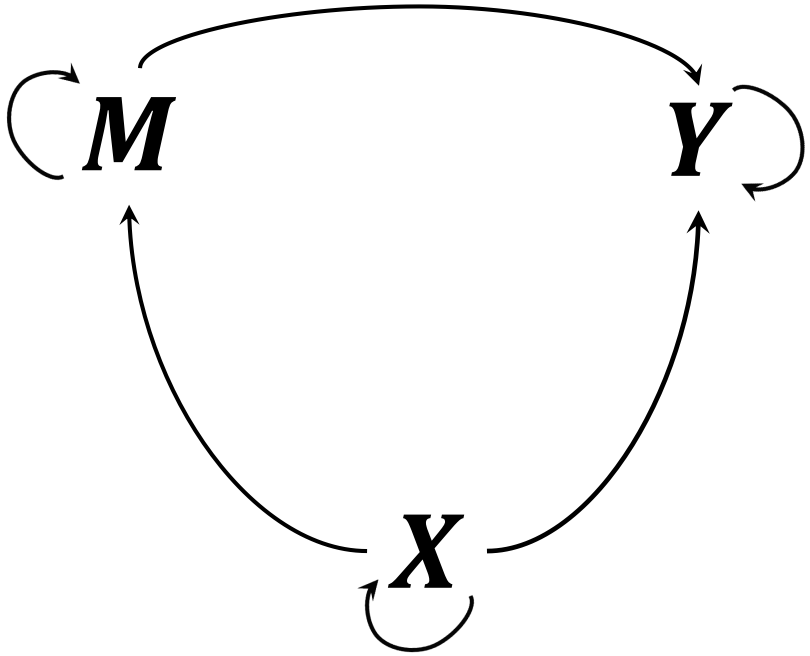


(c) DT-VAR(1) mediation with interval $2\Delta t$

Figure 2. DAG and SEM representations of DT-VAR(1) mediation models



(a) SEM representation



(b) Local Independence Graph

Figure 3. Two representations of CT causal structure, for a mediation model with lower-triangular drift matrix

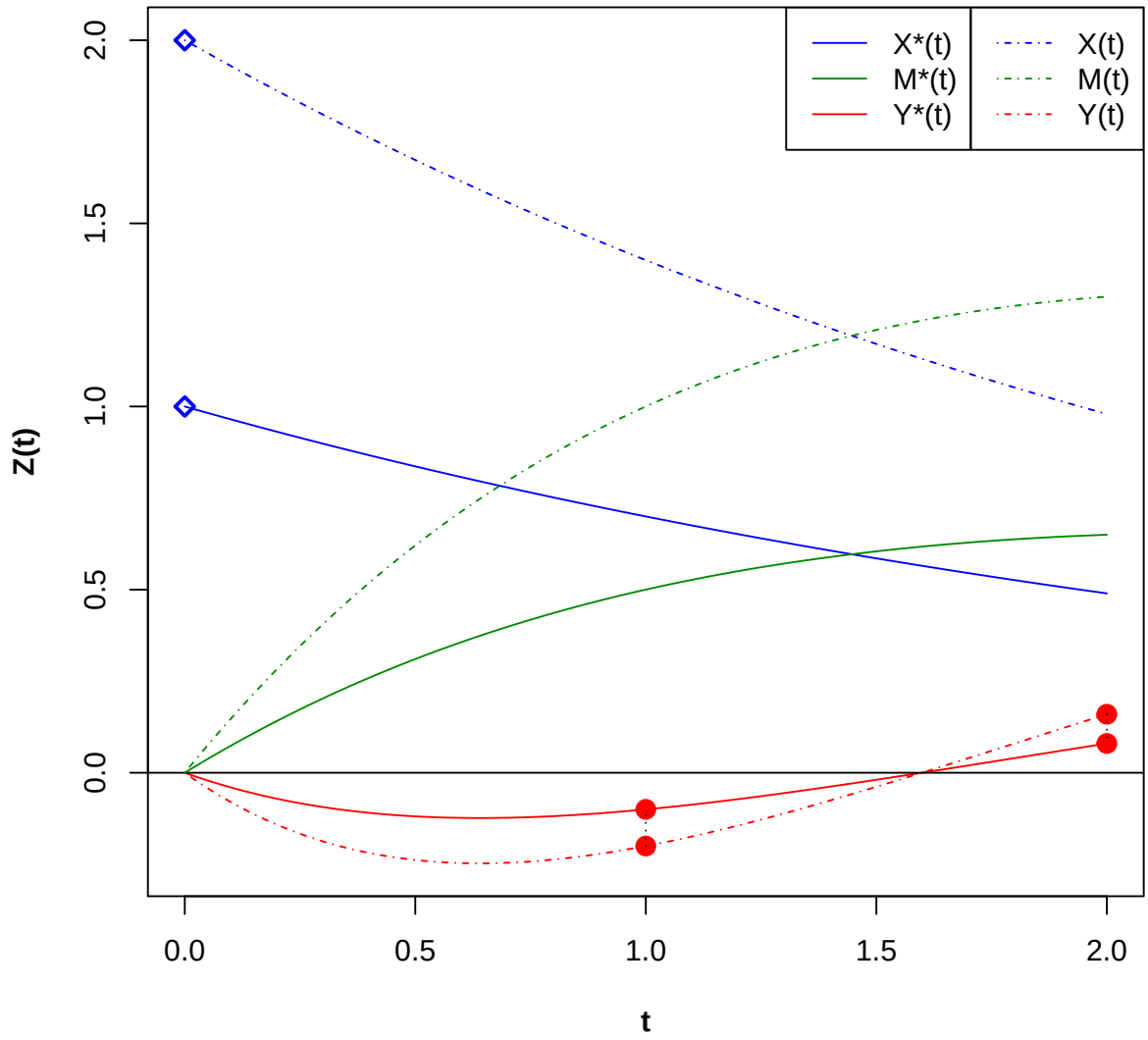
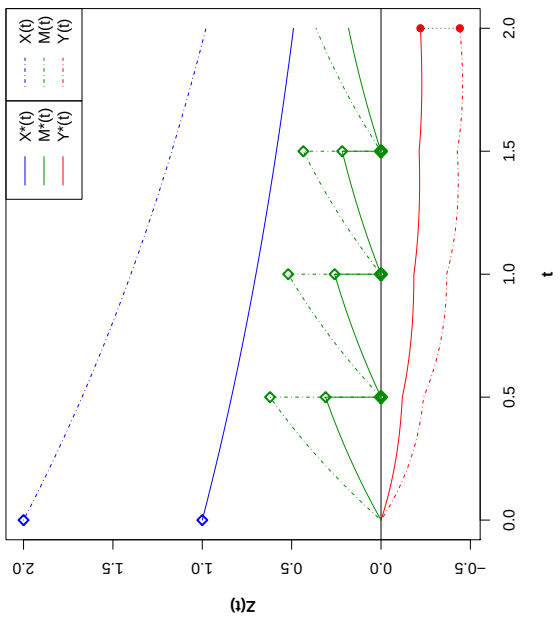
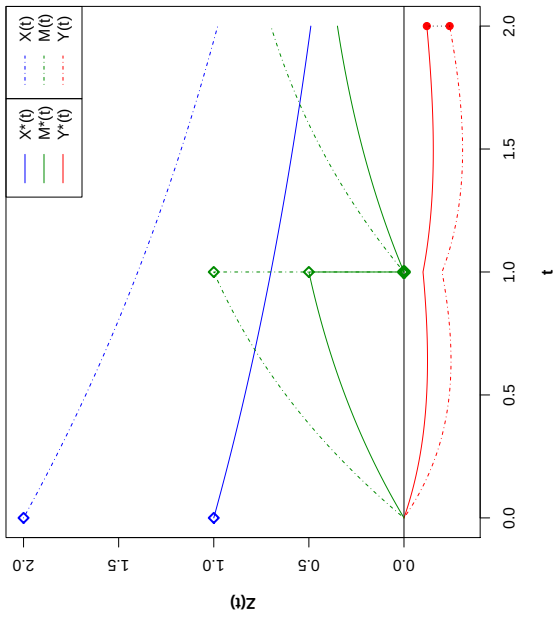


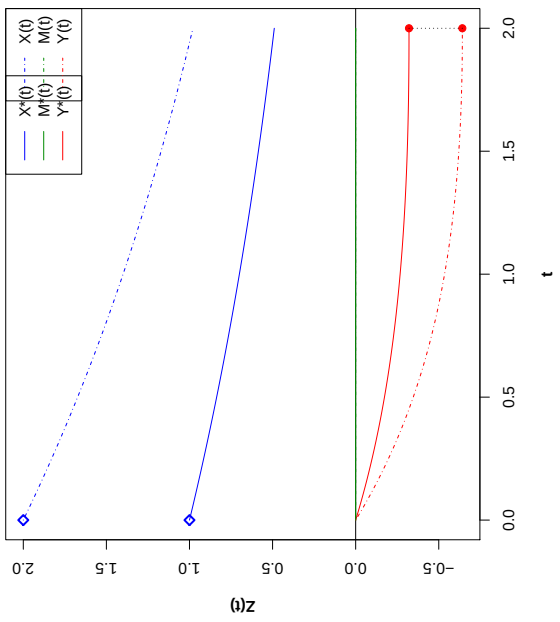
Figure 4. Trajectories of the multivariate process $\boldsymbol{Z}(t)$ under the two interventions described by the total effect, $TE(\Delta t)$. The two dotted lines connecting pairs of red dots at $t = 1$ and $t = 2$ represent $TE(1)$ and $TE(2)$ respectively. The two blue diamonds represent the different values of $X(0)$ in both sets of trajectories.



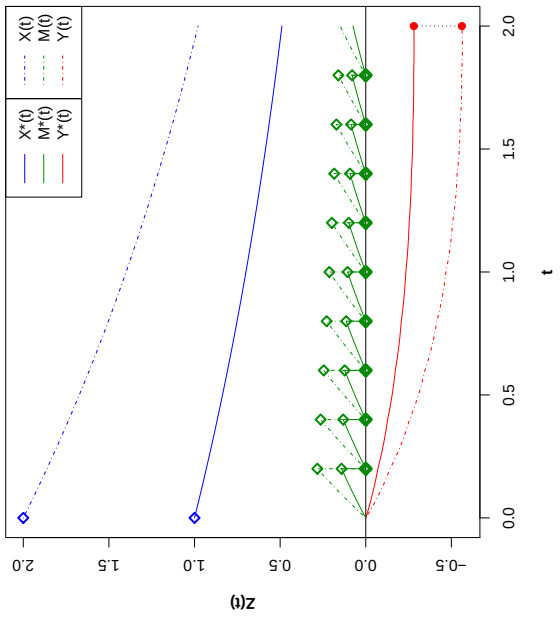
(a) $aCDE(2,1)$



(b) $aCDE(2,3)$

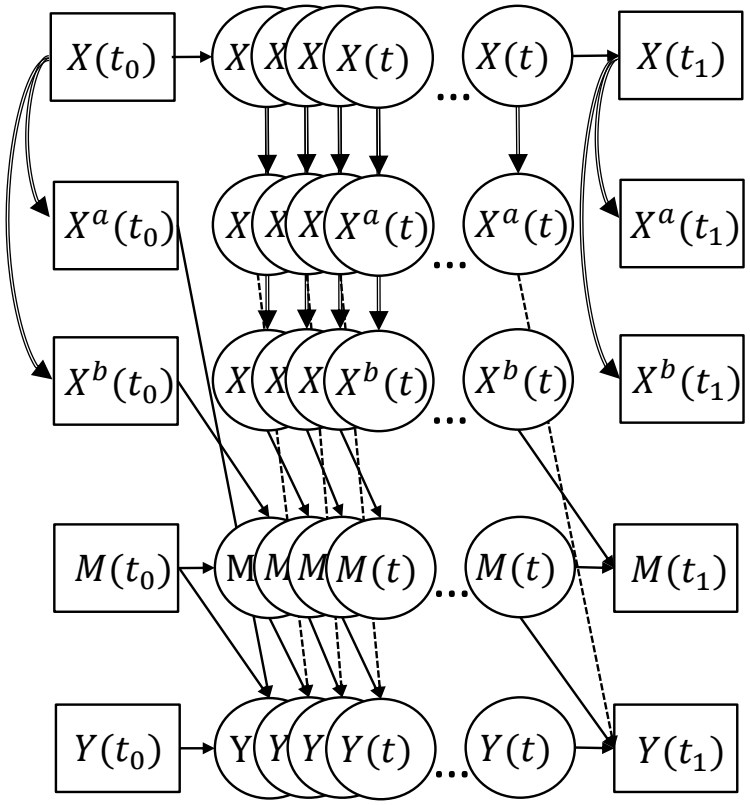


(c) $aCDE(2,9)$

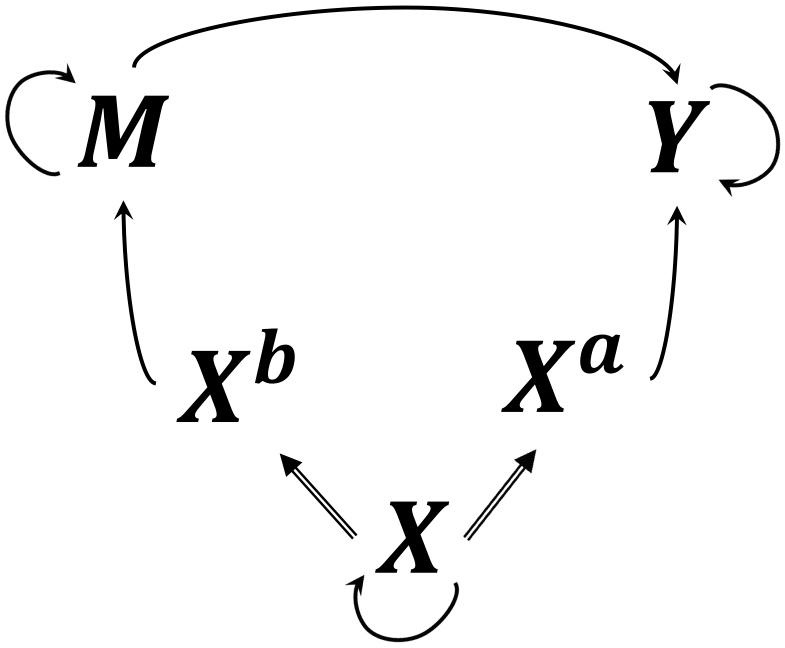


(d) $psDE(2) = iCDE(2)$

Figure 5. Trajectories of the multivariate process $\boldsymbol{Z}(t)$ showing the interventions described by different direct effects.



(a) Extended CT-VAR(1) SEM representation



(b) Extended Local Independence Graph

Figure 6. Two representations of an extended CT causal structure, for a mediation model with lower-triangular drift matrix

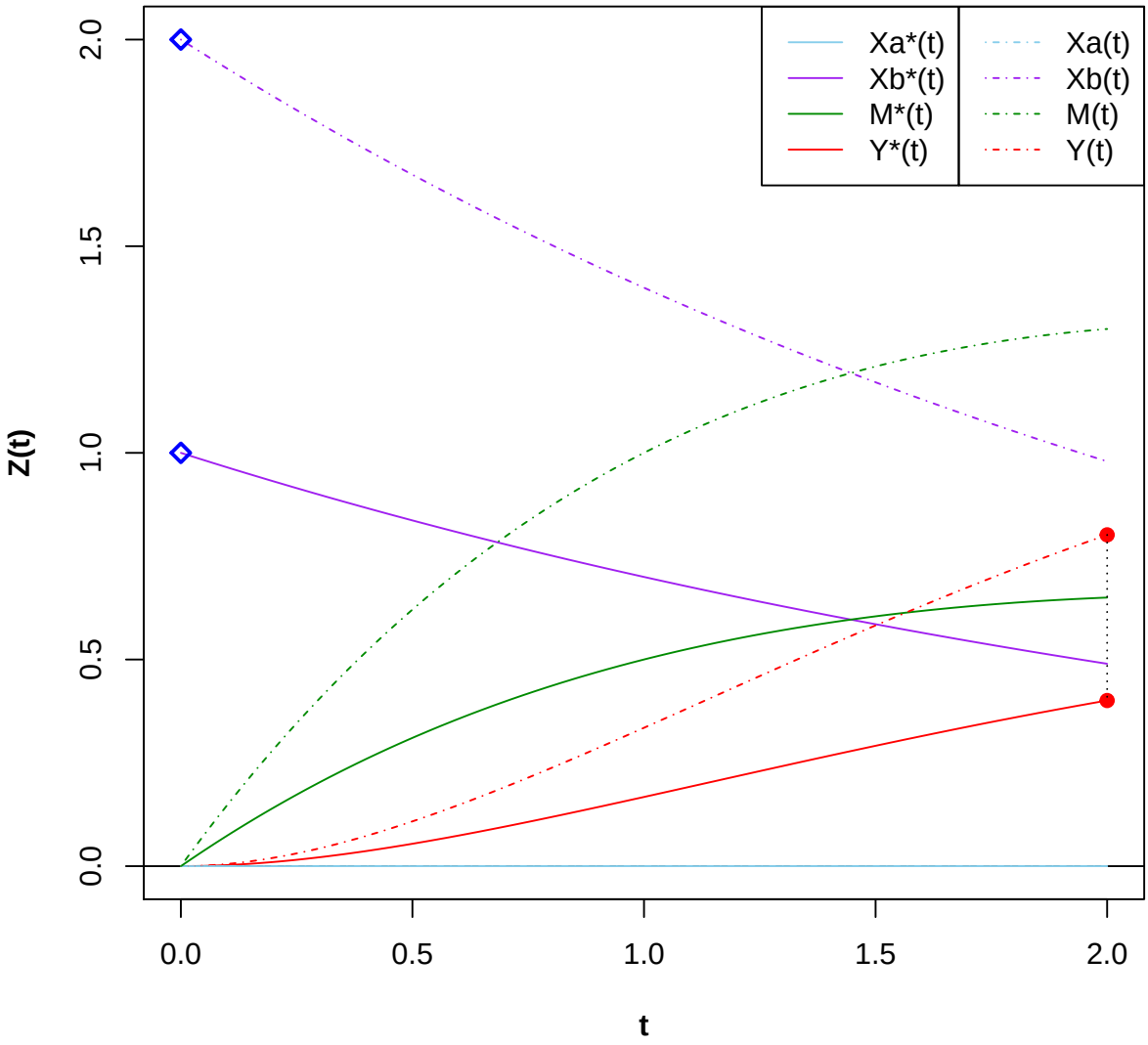


Figure 7. Trajectories of the extended multivariate system showing the interventions described by the interval Natural Indirect Effect $iNIE(\Delta t)$.

Appendix A

Proofs

In this section the equivalence of the path-tracing direct effect from the CT literature, described by Aalen et al. (2012, 2018) and Deboeck & Preacher (2016), denoted $pDE(\Delta t)$, and the variable-setting intervention, $iCDE(\Delta t)$, is proved. After the introduction of preliminary notation, we define the path-tracing direct effect for a general system using matrix algebra. The operation of intervening to set a variable-value at q intermediate moments in time over an interval Δt is expressed in matrix algebra form and this is used to define the acute controlled direct effect, $aCDE(\Delta t, q)$. The limit of this intervention as q tends to infinity is then defined as the $iCDE(\Delta t)$, and it is shown that this is equivalent to the path-tracing direct effect $pDE(\Delta t)$.

Preliminaries

Take it that we have a multivariate time-varying stochastic process, $\boldsymbol{Z}$, with the dimensions denoted $Z_1, \dots Z_V$, and with dynamics described by a first-order SDE or equivalently CT-VAR(1). As such we have a $V \times V$ drift matrix $\boldsymbol{A}$, which relates the $V \times 1$ column vectors $\frac{d\boldsymbol{Z}(t)}{dt}$ and $\boldsymbol{Z}(t)$. The elements of $\boldsymbol{Z}(t)$ are the positions or values of the process in each dimension at a particular time t, that is $Z_1(t), \dots Z_V(t)$, while the elements of $\frac{d\boldsymbol{Z}(t)}{dt}$ denote the rates of change in each dimension at a particular time t. Furthermore we can denote the *history* of some process Z_i from time point t to time point $t + \Delta t$ as $\bar{Z}_i(t, t + \Delta t)$; the history represents the values that $Z_i(s)$ takes on at each point in time s, for $t \leq s \leq t + \Delta t$.

In the derivation of direct effects through path-tracing and variable-setting operations, we will make use of a transformation matrix $\boldsymbol{S}$. Given that we treat the process Z_j as the mediator, we define $\boldsymbol{S}$ as a $V \times V$ matrix with zeros as off diagonal elements, a zero for the jth diagonal element, and ones as the other diagonal elements $\boldsymbol{S}_{ll}$, $l \in \{1 \dots V\} \backslash j$. Pre-multiplying a matrix by $\boldsymbol{S}$ produces a matrix with zeros for the jth row, and post-multiplying produces a matrix with zeros on the jth column. In the general

case, for variable setting operations, we will need to define some constant to which the putative mediating variable is set. For this purpose we will make use of the $V \times 1$ column vector $\boldsymbol{c}$, where the jth row contains the constant c, and all other elements are zero. Note that $\boldsymbol{SS} = \boldsymbol{S}$ and $\boldsymbol{Sc} = \boldsymbol{0}$

Path-tracing Direct Effect in matrix form

We can define the CT path-tracing direct effect in matrix notation. To find the direct effect from process Z_i to process Z_k, relative to process Z_j, over an interval Δt, given as $\phi^D_{ki}(\Delta t)$, we can take the matrix exponential of the effects matrix in which all of the paths in and out of the mediator Z_j are "switched off". We can express this altered drift matrix as $\boldsymbol{SAS}$. Pre and post-multiplying $\boldsymbol{A}$ by $\boldsymbol{S}$ sets the jth row and column entries of $\boldsymbol{A}$ equal to zero. For example, in a trivariate model in which we want to find the path-tracing direct effect of $Z_1(t)$ on $Z_3(t + \Delta t)$ relative to $\bar{Z}_2(t, t + \Delta t)$, we would have

$$\boldsymbol{SAS} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & a_{13} \\ 0 & 0 & 0 \\ a_{31} & 0 & a_{33} \end{bmatrix}$$

As such we can define the CT path-tracing direct effect of $Z_i(t)$ on $Z_k(t + \Delta t)$ relative to $\bar{Z}_j(t, t + \Delta t)$, denoted $pDE(\Delta t)_{ik.\bar{j}}$ as the element in the kth row and ith column (denoted $\langle k, i \rangle$) of the matrix exponential of this altered matrix,

$$pDE(\Delta t)_{ik.\bar{j}} = \boldsymbol{e}^{\boldsymbol{SAS}\Delta t} \langle k, i \rangle (z_i - z_i^*) \tag{22}$$

where, for completeness, this direct is defined relative to a $(z_i - z_i^*)$ unit increase of $Z_i(t)$.

Acute Variable-setting operations: the $aCDE(\Delta t, q)_{ik.j}$

Now that we have a matrix formulation for the path-based direct effect, we can investigate whether we can also formulate an equivalent variable-setting intervention. We can write the expected value of our process at time point $t + \Delta t$, given the values of the

process at some initial time-point t as

$$E\Big[\boldsymbol{Z}(t+\Delta t)|\boldsymbol{Z}(t)=\boldsymbol{z}(t)\Big]=\boldsymbol{e}^{\boldsymbol{A}\Delta t}\boldsymbol{z}(t)$$

where $\boldsymbol{z}(t)$ represents a $V\times 1$ column vector of values of $\boldsymbol{Z}(t)$, Suppose that we wish to split up our time interval Δt into $q+1$ discrete steps. As such, we can write

$$E\Big[\boldsymbol{Z}(t+\Delta t)|\boldsymbol{Z}(t)=\boldsymbol{z}(t)\Big]=\Big(\boldsymbol{e}^{\boldsymbol{A}\Delta t/(q+1)}\Big)^{q+1}\boldsymbol{z}(t)$$

Suppose that we we wish to set the value of some purported mediating variable $Z_j(t)$ to some constant c. We could express the expected value of $\boldsymbol{Z}(t+\Delta t)$ given this condition as

$$E\Big[\boldsymbol{Z}(t+\Delta t)|Z_j(t)=c\Big]=\Big(\boldsymbol{e}^{\boldsymbol{A}\Delta t/q+1}\Big)^{q+1}\Big(\boldsymbol{S}\boldsymbol{z}(t)+\boldsymbol{c}\Big)$$

where $\boldsymbol{c}$ is a $V\times 1$ column vector, with jth element c and zeros elsewhere. Let us now express the expected value that the variables $\boldsymbol{Z}$ would take on some short time interval after t, given as $t+\Delta t/(q+1)$, given the intervention on $Z_j(t)$ above, and an equivalent intervention on the value of $Z_j(t+\Delta t/(q+1))$. We can express this as

$$E\Big[\boldsymbol{Z}(t+\frac{\Delta t}{q+1})|Z_j(t)=Z_j(t+\frac{\Delta t}{q+1})=c\Big]=\boldsymbol{S}\Big(\boldsymbol{e}^{\boldsymbol{A}\frac{\Delta t}{q+1}}\Big)\Big(\boldsymbol{S}\boldsymbol{z}(t)+\boldsymbol{c}\Big)+\boldsymbol{c}$$

Repeating this procedure, we can find an expression for the expected value of $\boldsymbol{Z}(t+\Delta t)$ given that i) we split the time-interval Δt up into $q+1$ steps, ii) we set the initial value of $Z_j(t)$ equal to a constant c, and iii) at each $q+1$ step we set the value Z_j, that is, $\{Z_j(t), Z_j(t+\frac{\Delta t}{q+1}), Z_j(t+\frac{2\Delta t}{q+1}),\ldots,Z_j(t+\Delta t)\}$, equal to some constant c. We can express this as

$$E\Big[\boldsymbol{Z}(t+\Delta t)|Z_j(t)=\cdots=Z_j(t+\Delta t)=c\Big]=\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\frac{\Delta t}{q+1}}\Big)^{q+1}\boldsymbol{S}\boldsymbol{z}(t)+\sum_{n=0}^{q+1}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\frac{\Delta t}{q+1}}\Big)^{n}\boldsymbol{c} \quad (23)$$

We can use Equation (23) to define the acute controlled direct effect $aCDE(\Delta t, q)$: that is, the change in expected value of $Z_k(t + \Delta t)$, given an intervention to set the value of $Z_i(t)$ from x to x^*, while holding the value that Z_j takes on at time at q intermediate time points in the interval t to $t + \Delta t$ to a constant c.This is given as

$$\begin{aligned} aCDE(\Delta t, q)_{ik.j} = & E\Big[Z_k(t + \Delta t)|Z_i(t) = x, Z_j(t) = \cdots = Z_j(t + \Delta t) = c\Big] \\ & - E\Big[Z_k(t + \Delta t)|Z_i(t) = x^*, Z_j(t) = \cdots = Z_j(t + \Delta t) = c\Big] \end{aligned}$$

which simplifies to

$$\begin{aligned} aCDE(\Delta t, q)_{ik.j} &= \left(\left(\boldsymbol{S}e^{\boldsymbol{A}\frac{\Delta t}{q+1}}\right)^{q+1}\boldsymbol{S}\Big(\boldsymbol{z}(t) - \boldsymbol{z}^*(t)\Big)\right)\langle k, i\rangle \\ &= (\boldsymbol{S}e^{\boldsymbol{A}\frac{\Delta t}{q+1}})^{q+1}\langle k, i\rangle(z_i - z_i^*) \end{aligned} \tag{24}$$

where $\boldsymbol{z}(t)$ is a $V \times 1$ vector of values for $\boldsymbol{Z}(t)$, with ith element z_i, and $\boldsymbol{z}^*(t)$ is an identical column vector but with ith element z_i^*. Note that the second term on the left-hand side of Equation (23) cancels out.

Note that in the main text, a slightly different formulation of the $aCDE(\Delta t, q)_{ik.j}$ is given, in which the interventions on $Z_j(t)$ and $Z_j(t + \Delta t)$ are omitted. Both formulations are equivalent, as $Z_k(t + \Delta t)$ is independent of $Z_j(t + \Delta t)$. Furthermore, as can be seen in the Equation (24), the value of the $aCDE(\Delta t, q)_{ik.j}$ does not depend on $Z_j(t)$. This different formulation is given here to make clearer the relationship between the $aCDE(\Delta t, q)_{ik.j}$ and $iCDE(\Delta t)_{ik.j}$ defined below.

Interval variable-setting operations: the $iCDE(\Delta t)$

We first derive an expression for the expected value of $\boldsymbol{Z}(t + \Delta t)$ given that we intervene to set the value of Z_j to a constant *at every possible moment in time* in the interval Δt. This would be the limit of the expression in Equation (23) as q tends to infinity. As such we can say that a set of interventions to set

$Z_j(t), Z_j(t+\Delta t), Z_j(t+2\Delta t/q), \ldots, Z_j(t+\Delta t) = c$ is equal to setting $\bar{Z}_j(\Delta t) = c$.

Taking $q+1 = r$, we can write the limit of Equation (23) as

$$E\Big[\boldsymbol{Z}(t+\Delta t)|\bar{Z}_j(t+\Delta t) = c\Big] = \lim_{r\to\infty}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\frac{\Delta t}{r}}\Big)^r \boldsymbol{S}\boldsymbol{Z}(t) + \lim_{r\to\infty}\sum_{n=0}^{r}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\frac{\Delta t}{r}}\Big)^n \boldsymbol{c} \tag{25}$$

Taking the first term on the right hand-side of Equation (25), we can simplify

$$\begin{aligned}
\lim_{r\to\infty}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\Delta t/r}\Big)^r \boldsymbol{S} &= \lim_{r\to\infty}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\Delta t/r}\boldsymbol{S}\Big)^r \\
&= \lim_{r\to\infty}\left(\boldsymbol{S}\left(\sum_{d=0}^{\infty}\frac{(\boldsymbol{A}\Delta t/r)^d}{d!}\right)\boldsymbol{S}\right)^r \\
&= \lim_{r\to\infty}\left(\boldsymbol{S}\left(\boldsymbol{I} + \frac{\boldsymbol{A}\Delta t}{r} + \frac{\boldsymbol{A}^2\Delta t^2}{2r^2} + ...\right)\boldsymbol{S}\right)^r \\
&= \lim_{r\to\infty}\left(\lim_{r\to\infty}\left(\boldsymbol{S}\boldsymbol{I}\boldsymbol{S} + \frac{\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}\Delta t}{r} + \frac{\boldsymbol{S}\boldsymbol{A}^2\boldsymbol{S}\Delta t^2}{2r^2} + ...\right)\right)^r
\end{aligned}$$

As r approaches infinity, $1/r^2$ tends to zero quicker than $1/r$. As such we can write

$$\lim_{r\to\infty}(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\Delta t/r})^r\boldsymbol{S} \to \lim_{r\to\infty}\left(\boldsymbol{S}\boldsymbol{I}\boldsymbol{S} + \frac{\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}\Delta t}{r}\right)^r$$

Noting that $\boldsymbol{S}\boldsymbol{I}\boldsymbol{S} = \boldsymbol{S}\boldsymbol{I} = \boldsymbol{I}\boldsymbol{S}$, we can re-write this expression

$$\begin{aligned}
\lim_{r\to\infty}\left(\boldsymbol{S}\boldsymbol{I}\boldsymbol{S} + \frac{\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}\Delta t}{r}\right)^r &= \lim_{r\to\infty}\left(\boldsymbol{I} + \frac{\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}\Delta t}{r}\right)^r\boldsymbol{S} \\
&= \boldsymbol{e}^{\boldsymbol{S}\boldsymbol{A}\boldsymbol{S}\Delta t}\boldsymbol{S}
\end{aligned} \tag{26}$$

Taking the second term on the right-hand side of Equation (25),we can simplify

$$\begin{aligned}
\lim_{r\to\infty}\sum_{j=0}^{r}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\Delta t/r}\Big)^j\boldsymbol{c} &= \sum_{j=0}^{\infty}\lim_{r\to\infty}\Big(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}\Delta t/r}\Big)^j\boldsymbol{c} \\
&\to \sum_{j=0}^{\infty}\lim_{r\to\infty}\Big(\boldsymbol{S}\boldsymbol{I} + \frac{\boldsymbol{S}\boldsymbol{A}\Delta t}{r}\Big)^j\boldsymbol{c}
\end{aligned} \tag{27}$$

where, similar to above, $\boldsymbol{e}^{\boldsymbol{A}\Delta t/r}$ converges on the first two elements of its power series expansion, $\boldsymbol{I} + \frac{\boldsymbol{A}\Delta t}{r}$ as r tends towards infinity. Using Equations (26) and (26), we can now express the expected value of $Z_k(t + \Delta t)$ given interventions to set $\bar{Z}_j(t + \Delta t)$ to a constant c as

$$E\Big[\boldsymbol{Z}(t + \Delta t)|\bar{Z}_j(t + \Delta t) = c\Big] = \boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\boldsymbol{S}\boldsymbol{z}(t) + \sum_{j=0}^{\infty} \lim_{q+1\to\infty} \Big(\boldsymbol{S}\boldsymbol{I} + \frac{\boldsymbol{S}\boldsymbol{A}\Delta t}{q}\Big)^j \boldsymbol{c} \tag{28}$$

which we can now use to define the interval controlled direct effect.

The $iCDE_{ik.\bar{j}}(\Delta t)$ is defined as the change in expected value of $Z_k(t + \Delta t)$, given an intervention to set the value of $Z_i(t)$ from z_i^* to z_i, while also intervening to keep the value of the mediator Z_j at every moment in time the interval Δt to a constant value, $\bar{Z}_j(t + \Delta t) = c$.

$$\begin{aligned} iCDE(\Delta t)_{ik.\bar{j}} =& E\Big[Z_k(t + \Delta t)|Z_i(t) = z_i, \bar{Z}_j(t + \Delta t) = c\Big] \\ &- E\Big[Z_k(t + \Delta t)|Z_i(t) = z_i^*, \bar{Z}_j(t + \Delta t) = c\Big]. \end{aligned} \tag{29}$$

Using $\boldsymbol{z}(t)$ and $\boldsymbol{z}^*(t)$ as defined above, we can use Equation (28) to write

$$\begin{aligned} iCDE(\Delta t)_{ik.\bar{j}} &= \Big(\boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\boldsymbol{S}\Big(\boldsymbol{z}(t) - \boldsymbol{z}^*(t)\Big)\Big)\langle k, i\rangle \\ &= \Big(\boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\Big(\boldsymbol{z}(t) - \boldsymbol{z}^*(t)\Big)\Big)\langle k, i\rangle \\ &= \boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\langle k, i\rangle(z_i - z_i^*) \end{aligned} \tag{30}$$

which is exactly equal to our expression for the path-tracing direct effect in Equation (22). That is

$$iCDE(\Delta t)_{ik.\bar{j}} = pDE(\Delta t)_{ik.j} \tag{31}$$

showing that the path-tracing direct effect suggested by previous authors has an equivalent interpretation as a variable-setting intervention.

On the equivalences and non-equivalences of process trajectories

Here we have defined both the path-tracing and variable-setting operations as the difference between a pair of expected values of $Z_k(t + \Delta t)$. However, while we have shown that these effects are equal, that is, the differences between the relevant expected values are equal, in general the expected values themselves will not be equivalent. Comparing the path-tracing and variable-setting expected values in Equations (**??**) and (28), it is clear that

$$\boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\boldsymbol{z}(t) \neq \boldsymbol{e}^{\boldsymbol{SAS}\Delta t}\boldsymbol{S}\boldsymbol{z}(t) + \sum_{j=0}^{\infty} \lim_{q+1\to\infty} \Big(\boldsymbol{SI} + \frac{\boldsymbol{SA}\Delta t}{q}\Big)^{j}\boldsymbol{c}.$$

except in the case when $\boldsymbol{S}\boldsymbol{z}(t) = \boldsymbol{z}(t)$ and when $c = 0$. In substantive terms, this means that the path-tracing and variable-setting operations are equivalent when $c = 0$, the mediator is set to its equilibrium value. When this is is not the case, these values will differ according to the infinite sum term.

While a full discussion of this term is outside of the scope of this paper, whenever the absolute values of $\boldsymbol{SI} + \boldsymbol{SA}$ all lie within the unit circle, this sum converges on $\Big(\boldsymbol{I} - \boldsymbol{S} + \boldsymbol{SA}\Big)^{-1}\boldsymbol{c}$, as $\Delta t \to \infty$. As such, this term represents a vector of new long-run equilibirum positions for our processes, resulting from the effect of setting the mediator to a non-equilibrium value. This is analagous to case of a time-varying predictor variable resulting in a new equilibrium position, described by Hamaker (2005) for a DT model.

Appendix B

Extended Graphs and CT Indirect effects

In this section we describe the creation of an extended local independence graph and associated drift matrix, which enables for the definition and calculation of natural indirect effects in CT systems, $iNIE(\Delta t)$. This indirect effect can be defined as an $aCDE(\Delta t)$ in the extended LIG, where we calculate the direct effect of $Z_i(t)$ on $Z_k(t+\Delta t)$ relative to the mediator $\bar{Z}_i^a(t, t+\Delta t)$.

Creating an Indistinguishable Extended Local Independence Graph

Take it that we again have a V dimensional system, and we are interested in the indirect effect of process i on k through all other variables in our system $\boldsymbol{V}\backslash\{i,k\}$. That is, we want to calculate the indirect effect of changing the value of $Z_i(t)$ on the value of $Z_k(t+\Delta t)$ only through $\bar{Z}_l(t, t+\Delta t)$ for all $l \in \boldsymbol{V}\backslash\{i,k\}$. In order to define these indirect effects in an interventionist setting for a CT-VAR(1) process, we must first define an extended local independence graph and associated drift matrix. This extended drift matrix must have the characteristic that 1) it decomposes Z_i into a process which influences Z_k but not Z_l, given as Z_i^a, and a process which influences Z_l but not Z_k, given as Z_i^b, and 2) it is indistinguishable from the unextended LIG and drift matrix in the case where $Z_i^a = Z_i^b$.

For these criteria to be met, we must define the extended local independence graph such that Z_i fully determines Z_i^a and Z_i^b, and all reciprocal influences from Z_l and Z_k act on Z_i^a and Z_i^b through Z_i. In terms of the the extended drift matrix, we must construct this such that 1) all pathways (i.e. local dependencies) between the elements of Z_l, and pathways between Z_l and Z_k remain the same, 2) the outgoing pathway from process Z_i to Z_k is assigned only to Z_i^a, 3) there is no outgoing pathways from Z_i^a to any other process, 4) the outgoing pathways from Z_i to Z_l are assigned to Z_i^b, 5) the outgoing path from Z_i^b to Z_k is zero, 6) there is no pathways between Z_i^a and Z_i^b and 7) all incoming pathways to Z_i in the original drift matrix are assigned to both Z_i^a and Z_i^b.

To illustrate this, let us take a trivariate system with a full drift matrix

$$\boldsymbol{A}^{=}\begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \tag{32}$$

with local independence graph shown in Figure B1a. Given that we wish to find the indirect effect of $Z_1(t)$ on $Z_3(t+\Delta t)$ through $\bar{Z}_2(t+\Delta t)$, the corresponding drift matrix for the relevant extended LIG shown in Figure B1b is given as

$$\boldsymbol{A'} = \begin{bmatrix} A_{11} & 0 & A_{12} & A_{13} \\ 0 & A_{11} & A_{12} & A_{13} \\ 0 & A_{21} & A_{22} & A_{23} \\ A_{31} & 0 & A_{32} & A_{33} \end{bmatrix} \tag{33}$$

where the first and second column relate to Z_1^a and Z_1^b respectively. This extended drift matrix has the property that,taking the matrix exponential of the drift matrix multiplied by a vector of starting values in which $Z_1^a(t) = Z_1^b(t)$, results in 1) the same expected values for $Z_2(t+\Delta t)$ and $Z_3(t+\Delta t)$ as with the original drift matrix, 2) the same expected value for $Z_1^a(t+\Delta t)$ and $Z_1^b(t+\Delta t)$, and 3) these expected values for Z_1^a and Z_1^b are the same as for $Z_1(t+\Delta t)$ using the original drift matrix and LIG.

We can show that the LIG in Figure B1b, and the associated drift matrix of the extended LIG $\boldsymbol{A'}$ represents an indistinguishable description of observational data as the unextended graph (Figure B1a) and drift matrix, with a numerical example. Take it that we have a drift matrix

$$\boldsymbol{A}^{=}\begin{bmatrix} -.589 & .012 & .115 \\ .854 & -.406 & 1.017 \\ .199 & .026 & -.520 \end{bmatrix} \tag{34}$$

and furthermore let us define a vector of values for our system at time piont $t = 0$, $\boldsymbol{z}(0)$, taking for instance $z_1(0) = z_2(0) = z_3(0) = 1$. We can further define an equivalent set of values for ou extended system as $\boldsymbol{z'}(0)$, with $z_1^a(0) = z_1^b(0) = z_2(t) = z_3(0) = 1$. We can thus calculate

$$E[\boldsymbol{Z}(1)|\boldsymbol{Z}(0) = \boldsymbol{z}(0)] = \boldsymbol{e}^{\boldsymbol{A}}\boldsymbol{z}(0)$$

$$\begin{bmatrix} E[Z_1(1)] \\ E[Z_2(1)] \\ E[Z_3(1)] \end{bmatrix} = \begin{bmatrix} .644 \\ 1.942 \\ .750 \end{bmatrix}$$

where we can see that an equivalent vector of expected values is obtained using $\boldsymbol{A'}$

$$\boldsymbol{e}^{\boldsymbol{A'}}\boldsymbol{z'}(0) = \begin{bmatrix} E[Z_1^a(1)] \\ E[Z_1^b(1)] \\ E[Z_2(1)] \\ E[Z_3(1)] \end{bmatrix} = \begin{bmatrix} .644 \\ .644 \\ 1.942 \\ .750 \end{bmatrix}.$$

As can be seen from above, the extended drift matrix results in the exact same expected values for $Z_2(1)$ and $Z_3(1)$ as the unextended drift matrix, when $z_1^a(0) = z_1^b(0)$. Furthermore, the extended graph gives the same expected values for $Z_1^a(1)$ and $Z_1^b(1)$, as the unextended graph does for $E[Z_1(1)]$: this should be expected given the deterministic relationships from $Z_1(t)$, to $Z_1^a(t)$ and $Z_1^b(t)$ as denoted in the extended LIG.

Path-tracing indirect effect

The calculation of the path-tracing indirect effect defined by Deboeck & Preacher (2016) can be generalized to the case of any size or form of drift matrix by taking the extended DAG approach. We define the path-tracing indirect effect of $Z_i(t)$ on $Z_k(t + \Delta t)$ as

$$pIE(\Delta t)_{ik.\bar{j}} = \boldsymbol{e}^{\boldsymbol{SA'S}\Delta t}\langle k, i\rangle(z_i - z_i^*) \tag{35}$$

where $\boldsymbol{A}'$ is constructed as described above, and $\boldsymbol{S}$ has a zero on the Z_i^a diagonal, ones on the other diagonals, and zeros elsewhere. As such, the path-tracing indirect effect can be seen as the path-tracing *direct effect* of $Z_i^b(t)$ on $Z_k(t+\Delta t)$, relative to $\bar{Z}_i^a(t+\Delta t)$ (see Equation (22)).

Variable-setting indirect effect

We can define the interval natural indirect effect, $iNIE(\Delta t)$, as the difference in expected value of $Z_k(t+\Delta t)$, given an intervention to set $Z_i(t)$ (and equivalently $Z_i^b(t)$), from z_i^* to z_i, and an intervention to set $\bar{Z}_i^a(t+\Delta t) = c$, given as

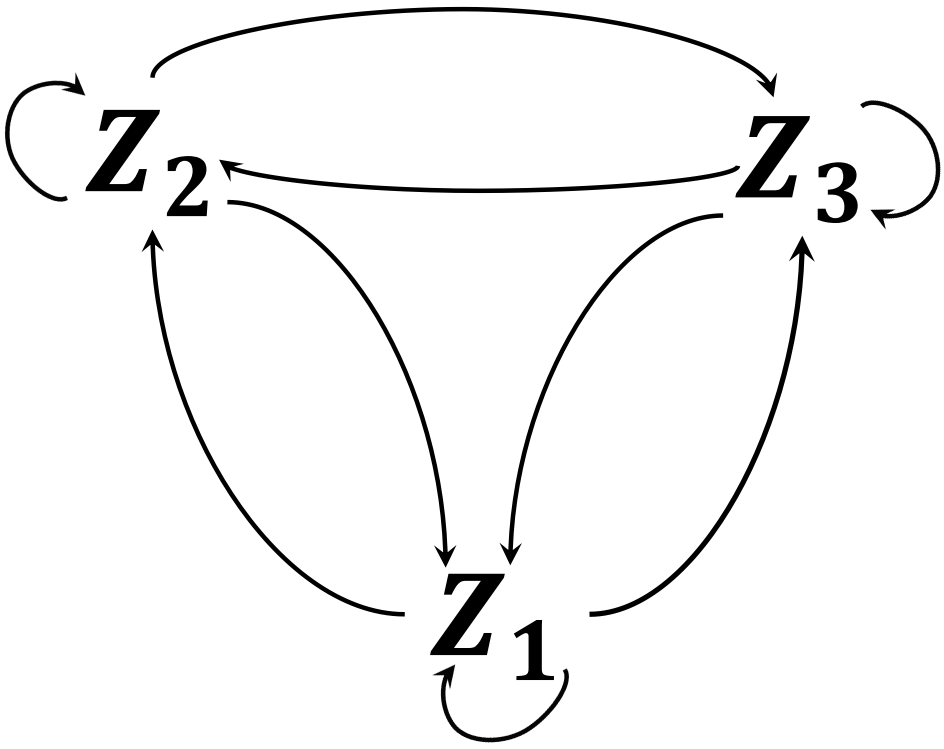
$$\begin{aligned} iNIE(\Delta t)_{ik.\bar{i^a}} =& E[Z_k(t+\Delta t)|Z_i(t) = z_i, \bar{Z}^a(t+\Delta t) = c] \\ & - E[Z_k(t+\Delta t)|Z_i(t) = z_i^*, \bar{Z}^a(t+\Delta t) = c] \end{aligned} \tag{36}$$

where, comparing this to Equation (29) we can see that this can be seen as the $iCDE(\Delta t)$ of $Z_i(t)$ on $Z_k(t+\Delta t)$ relative to the mediator Z_i^a.

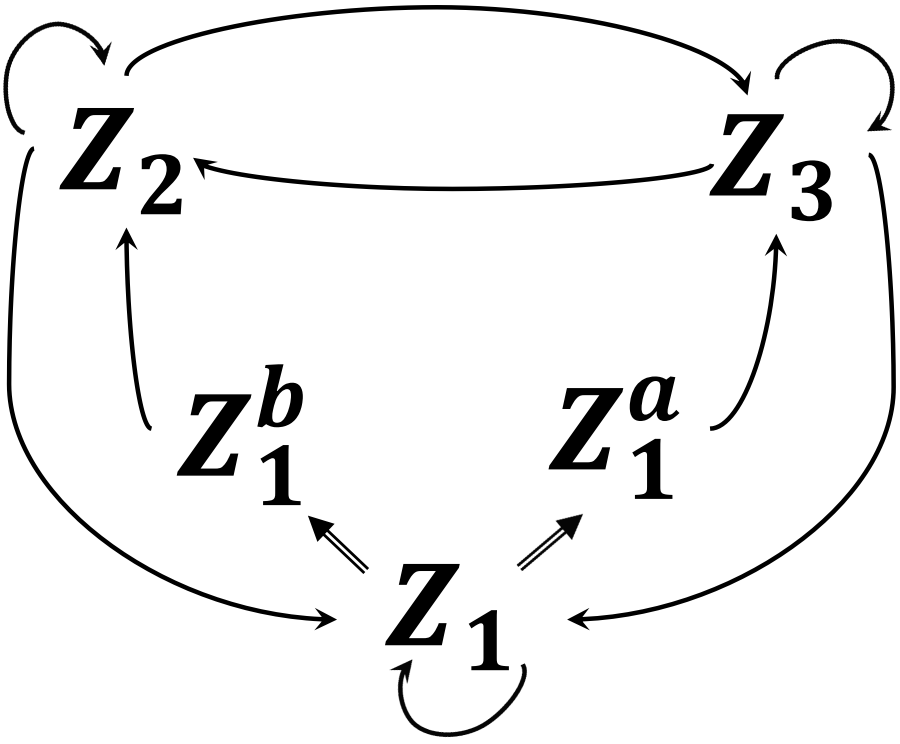
As such, to show that the variable-setting indirect effect $iNIE(\Delta t)$ is equal to the path-tracing indirect effect $pIE(\Delta t)$, it suffices to show that,

$$\lim_{r\to\infty}(\boldsymbol{S}\boldsymbol{e}^{\boldsymbol{A}'\Delta t/r}\boldsymbol{S})^r = \boldsymbol{e}^{\boldsymbol{S}\boldsymbol{A}'\boldsymbol{S}\Delta t}\boldsymbol{S} \tag{37}$$

for $\boldsymbol{S}$ as defined in Appendix A. The proof of this statement follows immediately from the proof of the direct effects in Appendix A, showing that the path-tracing indirect effect defined by previous authors in fact describes a variable-setting intervention in an extended LIG, $pIE(\Delta t)_{ik.\bar{j}} = iNIE(\Delta t)_{ik.\bar{i^a}}$.



(a) Local Independence Graph



(b) Extended Local Independence Graph

Figure B1. Two local independence graphs for a trivariate system with full drift matrix

References

Aalen, O. (1987). Dynamic modelling and causality. *Scandinavian Actuarial Journal*, *1987*(3-4), 177–190.

Aalen, O., Gran, J., Røysland, K., Stensrud, M., & Strohmaier, S. (2018). Feedback and mediation in causal inference illustrated by stochastic process models. *Scandinavian Journal of Statistics*, *45*(1), 62–86.

Aalen, O., Røysland, K., Gran, J., Kouyos, R., & Lange, T. (2016). Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical methods in medical research*, *25*(5), 2294–2314.

Aalen, O., Røysland, K., Gran, J., & Ledergerber, B. (2012). Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *175*(4), 831–861.

Angraini, Y., Toharudin, T., Folmer, H., & Oud, J. H. (2014). The relationships between individualism, nationalism, ethnocentrism, and authoritarianism in flanders: A continuous time-structural equation modeling approach. *Multivariate Behavioral Research*, *49*(1), 41–53.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Boker, S. M. (2002). Consequences of continuity: The hunt for intrinsic properties within parameters of dynamics in psychological processes. *Multivariate Behavioral Research*, *37*(3), 405–422.

Boker, S. M., Deboeck, P., Edler, C., & Keel, P. (2010). Generalized local linear approximation of derivatives from time series. In S. Chow & E. Ferrar (Eds.), *Statistical*

methods for modeling human dynamics: An interdisciplinary dialogue (p. 179-212). Boca Raton, FL: Taylor & Francis.

Boker, S. M., Montpetit, M. A., Hunter, M. D., & Bergeman, C. S. (2010). Modeling resilience with differential equations. In P. Molenaar & K. Newell (Eds.), *Learning and development: Individual pathways of change* (p. 183-206). Washington, DC: American Psychological Association.

Boker, S. M., Neale, M., & Rausch, J. (2004). Latent differential equation modeling with multivariate multi-occasion indicators. In O. H. van Montfort K. & A. Satorra (Eds.), *Recent developments on structural equation models* (pp. 151–174). Dordrecht, the Netherlands: Kluwer Academic.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological methodology*, 37–69.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In *Handbook of causal analysis for social research* (pp. 301–328). Springer.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of clinical psychology*, *64*(9), 1089–1108.

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, *9*, 91–121.

Bringmann, L., Lemmens, L., Huibers, M., Borsboom, D., & Tuerlinckx, F. (2015). Revealing the dynamic network structure of the beck depression inventory-ii. *Psychological medicine*, *45*(4), 747–757.

Bringmann, L., Pe, M., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., ... Kuppens, P. (2016). Assessing temporal emotion dynamics using networks. *Assessment*, *23*(4), 425–435.

Bringmann, L., Vissers, N., Wichers, M., Geschwind, N., Kuppens, P., Peeters, ... Tuerlinckx, F. (2013). A network approach to psychopathology: New insights into clinical longitudinal data. *PLoS ONE*, *8*, e60188, 1-13.

Chow, S., Ou, L., Ciptadi, A., Prince, E. B., You, D., Hunter, M. D., ... Messinger, D. S. (2018). Representing sudden shifts in intensive dyadic interaction data using differential equation models with regime switching. *Psychometrika*, 1–35.

Chow, S., Ram, N., Boker, S., Fujita, F., Clore, G., & Nesselroade, J. (2005). Capturing weekly fluctuation in emotion using a latent differential structural approach. *Emotion*, *5*(2), 208–225.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 558–557.

Dawid, A. P. (2010). Beware of the dag! In *Causality: Objectives and assessment* (pp. 59–86).

Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous time mediation analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 61–75.

Didelez, V. (2007). Graphical models for composable finite markov processes. *Scandinavian Journal of Statistics*, *34*(1), 169–185.

Didelez, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *70*(1), 245–264.

Driver, C. C., Oud, J. H. L., & Voelkle, M. C. (2017). Continuous time structural equation modeling with R package ctsem. *Journal of Statistical Software*, *77*(5), 1–35. doi: 10.18637/jss.v077.i05

Driver, C. C., & Voelkle, M. C. (2017). Introduction to hierarchical continuous time dynamic modelling with ctsem. *R package Vignette. Available online at: https://cran. r-project. org/web/packages/ctsem/index. html*.

Eichler, M., & Didelez, V. (2012). Causal reasoning in graphical time series models. *arXiv preprint arXiv:1206.5246*.

Epskamp, S., Maris, G. K., Waldorp, L. J., & Borsboom, D. (in press). Network psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *Handbook of psychometrics.* New York: Wiley.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development*, *58*, 80–92.

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods & Research*, *33*(3), 404–416.

Hamaker, E. L., & Dolan, C. V. (2009). Idiographic data analysis: Quantitative methodsâĂŤfrom simple to advanced. In *Dynamic process methodology in the social and developmental sciences* (pp. 191–216). Springer.

Hamaker, E. L., Kuiper, R., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102-116. doi: 10.1037/a0038889

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hartmann, J. A., Wichers, M., Menne-Lothmann, C., Kramer, I., Viechtbauer, W., Peeters, F., ... others (2015). Experience sampling-based personalized feedback and positive affect: a randomized controlled trial in depressed patients. *PLoS One*, *10*(6), e0128095.

Hernán, M. A. (2018). The c-word: Scientific euphemisms do not improve causal inference from observational data. *American journal of public health*(0), e1–e4.

Hernan, M. A., & Robins, J. M. (2013). *Causal inference*. Boca Raton, FL.

Kramer, I., Simons, C. J., Hartmann, J. A., Menne-Lothmann, C., Viechtbauer, W., Peeters, F., ... others (2014). A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. *World Psychiatry*, *13*(1), 68–77.

MacKinnon, D. (2008). *Introduction to statistical mediation analysis*. New York: Erlbaum.

McNally, R. J., Robinaugh, D. J., Wu, G. W., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(6), 836–849.

Mohr, D. C., Schueller, S. M., Montague, E., Burns, M. N., & Rashidi, P. (2014). The behavioral intervention technology model: an integrated conceptual and technological framework for ehealth and mhealth interventions. *Journal of medical Internet research*, *16*(6).

Naslund, J. A., Marsch, L. A., McHugo, G. J., & Bartels, S. J. (2015). Emerging mhealth and ehealth interventions for serious mental illness: a review of the literature. *Journal of mental health*, *24*(5), 321–332.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2009). A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, *74*, 395-418.

Oravecz, Z., Tuerlinckx, F., & Vandekerckhove, J. (2011). A hierarchical latent stochastic difference equation model for affective dynamics. *Psychological Methods*, *16*, 468–490.

Ou, L., Hunter, M. D., & Chow, S. (2017). WhatâĂŹs for dynr: A package for linear and nonlinear dynamic modeling in r. *Journal of Statistical Software*.

Oud, J., & Delsing, M. J. M. H. (2010). Continuous time modeling of panel data by means of SEM. In K. van Montefort, J. Oud, & A. Satorra (Eds.), *Longitudinal research with latent variables* (pp. 201–244). New York, NY: Springer.

Pearl, J. (2001). Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence* (pp. 411–420).

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. (2012). The causal mediation formulaâĂŤa guide to the assessment of pathways and mechanisms. *Prevention Science*, *13*(4), 426–436.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological methods*, *19*(4), 459.

Reichardt, C. S. (2011). Commentary: Are three waves of data sufficient for assessing mediation? *Multivariate Behavioral Research*, *46*(5), 842–851.

Robins, J. M. (2003). Semantics of causal dag models and the identification of direct and indirect effects. *Highly struc-tured stochastic systems*, 70–81.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, 143–155.

Robins, J. M., & Richardson, T. S. (2010). Alternative graphical causal models and the identification of direct effects. *Causality and psychopathology: Finding the determinants of disorders and their cures*, 103–158.

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*(2), 245.

Røysland, K., et al. (2012). Counterfactual analyses with graphical models based on local independence. *The Annals of Statistics*, *40*(4), 2162–2194.

Sarker, H., Hovsepian, K., Chatterjee, S., Nahum-Shani, I., Murphy, S. A., Spring, B., ... Kumar, S. (2017). From markers to interventions: The case of just-in-time stress intervention. In *Mobile health* (pp. 411–433). Springer.

Sarker, H., Sharmin, M., Ali, A. A., Rahman, M. M., Bari, R., Hossain, S. M., & Kumar, S. (2014). Assessing the availability of users to engage in just-in-time intervention in the natural environment. In *Proceedings of the 2014 acm international joint conference on pervasive and ubiquitous computing* (pp. 909–920).

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data.

Sokol, A., & Hansen, N. R. (2014). Causal interpretation of stochastic differential equations. *Electron. J. Probab*, *19*(100), 1–24.

Steele, J. S., & Ferrer, E. (2011). Latent differential equation modeling of self-regulatory and coregulatory affective processes. *Multivariate Behavioral Research*, *46*(6), 956–984.

Toharudin, T., Oud, J. H., & Billiet, J. B. (2008). Assessing the relationships between nationalism, ethnocentrism, and individualism in Flanders using Bergstrom's approximate discrete model. *Statistica Neerlandica*, *62*(1), 83–103.

VanderWeele, T. J. (2015). *Explanation in causal inference: methods for mediation and interaction*. New York,NY: Oxford University Press.

VanderWeele, T. J., & Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(3), 917–938.

Vansteelandt, S. (2007). On confounding, prediction and efficiency in the analysis of longitudinal and cross-sectional clustered data. *Scandinavian journal of statistics*, *34*(3), 478–498.

Voelkle, M., Oud, J. H. L., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. *Psychological Methods*, *17*, 176-192.

Wright, S. (1934). The method of path coefficients. *The annals of mathematical statistics*, *5*(3), 161–215.