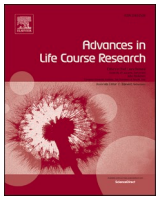
Advances in Life Course Research 60 (2024) 100617

Contents lists available at ScienceDirect

Advances in Life Course Research

journal homepage: www.elsevier.com/locate/alcr



Estimating causal effects from panel data with dynamic multivariate panel models

Jouni Helske [a,b,*], Santtu Tikka [b]

[a] INVEST Research Flagship Centre, University of Turku, Finland
[b] Department of Mathematics and Statistics, University of Jyväskylä, Finland

ARTICLE INFO

Keywords:
Bayesian methods
Causal inference
Markov models
Intervention
Panel data
Prediction

ABSTRACT

Panel data are ubiquitous in scientific fields such as social sciences. Various modeling approaches have been presented for observational causal inference based on such data. Existing approaches typically impose restrictive assumptions on the data-generating process such as Gaussian responses or time-invariant effects, or they can only consider short-term causal effects. To surmount these restrictions, we present the dynamic multivariate panel model (DMPM) that supports time-varying, time-invariant, and individual-specific effects, multiple responses across a wide variety of distributions, and arbitrary dependency structures of lagged responses of any order. We formally demonstrate how DMPM facilitates causal inference within the structural causal modeling framework and we take a Bayesian approach for the estimation of the posterior distributions of the model parameters and causal effects of interest. We demonstrate the use of DMPM by applying the approach to both real and synthetic data.

1. Introduction

Longitudinal panel data that consists of multiple individuals measured over multiple time points are common in various fields such as sociology, econometrics, and psychology. For example, analyzing individual-level life course data is valuable for assessing important life transitions, such as to employment or parenthood, and how policy reforms and interventions can impact the lives of individuals and the stability of their families. As conducting a randomized control trial is often infeasible in the social sciences, causal inference must be based on observational studies. Several approaches have been developed to analyze such data, tailored to specific research questions, characteristics of the available data, and information about the underlying causal graph or data-generating process.

There are two general approaches to observational causal inference: The potential outcomes framework (Rosenbaum & Rubin, 1983) and structural causal models (SCM, Pearl, 2009). While potential outcomes are more popular in social sciences (Imbens, 2020, 2024) here we follow the SCM approach, which we will discuss in more detail in Section 2. One of the key components of the SCM framework is directed acyclic graphs (DAG) which are used to represent the assumed causal relationships between observed and unobserved variables relevant to the research question (Lauritzen, 1996; Spirtes et al., 2001). This DAG can be then used to assess whether the available data are sufficient for answering causal queries of interest without any parametric assumptions. If the answer is yes, i.e., the causal effect of interest is identifiable (possibly after further assumptions or data collection), we can then proceed with suitable statistical modeling of the data to estimate the causal effects. In general, DAGs provide a convenient visual method for explicitly stating and presenting the general assumptions regarding the data-generating process, and they have long been used also in the context of structural equation models (SEM, Blalock, 1964; Kline, 2011) and general Bayesian modeling (Spiegelhalter, 1998).

Our interest is in assessing the causal effects of an intervention on some variable x_t at time t on a variable y_{t+k}, $k \geq 0$ from panel data with N units (individuals) and T time points. Typically, $k \leq T - t$, meaning that we consider effects of interventions only within the observed time period 1, …, T, although forecasting beyond T is also possible assuming that we have data on potential exogenous variables of the system beyond T. Both x_t and y_{t+k} can be vector-valued variables with mixed types (e.g., continuous and categorical), and they can exhibit cross-lagged dependencies such that y_t and x_t both depend on the past values of y_t and x_t (i.e., $y_{t-\ell}$ and $x_{t-\ell}$ for some $\ell > 0$). For example, x_t could be the size of the social network of an individual and y_t their level of community

* Correspondence to: INVEST Research Flagship Centre, University of Turku, Turku, Finland.
E-mail address: jouni.helske@utu.fi (J. Helske).

https://doi.org/10.1016/j.alcr.2024.100617
Received 12 October 2023; Received in revised form 9 April 2024; Accepted 7 May 2024
Available online 10 May 2024

participation, and we could be interested in the causal effect of x_t on y_t, …, y_{t+k}, but also how intervening on y_t affects future values y_{t+1}, y_{t+2}, … and x_{t+1}, x_{t+2}, …. In many panel data models, the causal inference is based on the estimated model parameters (regression/path coefficients) which, under suitable assumptions, can be interpreted causally, for example as the effect of x_t on y_{t+1}. We do not directly assume that any model parameter has a causal interpretation, but such interpretations may arise in specific circumstances, as we will show in Section 2. Instead, following the SCM framework, we focus on the distributions of response variables after the intervention has been carried out, i.e., interventional distributions, to identify and estimate causal effects of interest.

1.1. Long-term and time-varying effects

Throughout the paper, we consider short-term and long-term causal effects of a variable x_t on y_{t+k}, which we define as follows: when $0 \leq k \leq 1$, the causal effect is short-term (i.e., a contemporaneous or a one-step-ahead causal effect) and when $k \geq 1$, the causal effect is long-term. Note that these long-term effects are distinct from the long-run coefficients obtained from autoregressive distributed lag and error-correction models (Shamsollahi et al., 2022; Thombs, 2022) that describe the changes in the long-run equilibrium of the system, which corresponds to a long-term effect of x_t on y_{t+k} with $k \to \infty$. To facilitate our approach for estimating short-term and long-term causal effects, we assume that the structure of the causal graph (DAG) is time-invariant, i.e., the presence or absence of nodes and the edges between them does not depend on time.

In general, it is nontrivial to assess the long-term effect of say x_t on y_{t+k} directly as a function of the model parameters, especially in non-Gaussian or non-stationary settings and when the interest is not only in the average causal effect (ACE) but the whole distribution of y_{t+k} after the intervention. Interest in long-term effects can also necessitate the expansion of univariate models to multivariate responses. For example, in the cross-lagged panel model formulation of (Allison et al., 2017) with two response variables y_t and x_t, only the equation related to y_t is directly estimated. However, due to the cross-lagged dependencies between y_t and x_t, we also need to model how x_t depends on y_t in order to estimate the effect of x_t on y_{t+k} for $k > 1$ at time t. Thus, it is necessary to jointly model all relevant endogenous variables of the assumed underlying causal graph. In the context of vector autoregression (VAR) models (Holtz-Eakin et al., 1988; Sims, 1980), impulse response functions are often employed for assessing the long-term effects of external shocks. In a more general setting, probabilistic simulations offer a natural way to study such long-term effects in a multivariate cross-lagged setting, where manipulating one variable can lead to complex feedback dynamics.

Even if the interest is mostly in the short-term effects of some variables on the outcome variables(s), these effects may change over time (i.e., depend on t), for example, because the effect of a policy change accumulates or diminishes over time, or due to unmeasured changes in the individuals' common environment (shared context). Even if it is deemed unlikely that a variable has such a time-varying effect, it can still be beneficial to allow time-variability in the model formulation, as it can serve as a diagnostic check for unobserved time-varying confounding. More specifically, if domain knowledge indicates that the effect of a specific covariate on a response should be constant in time, but the model parameter estimates indicate the opposite, then there may be reason to believe that an unobserved time-varying confounder is present. These considerations are especially relevant in scenarios where the time period under study is moderate or long, e.g., when studying how changes in family relationships such as the loss of a parent affect the life satisfaction and mental well-being during the life course trajectories from childhood to elderhood. Such events can have both immediate effects and long-term effects, the magnitude of which likely depends on the age when the event is experienced.

When the time-varying effect itself does not have a strong serial dependency and the number of time points is relatively small, a simple dummy-coded time variable in the model could be reasonable in theory (as suggested for example by Wooldridge, 2010), but quickly leads to computational and overfitting issues as the number of time points increases, so alternative approaches have to be considered. In multilevel modeling, time-specific random effects may be used to capture time-varying effects. Alternatively, when time-varying coefficients cannot be assumed to be exchangeable as in multilevel models but exhibit some pattern, an approach that takes into account dependency between adjacent coefficients can be more appropriate. For example, specifically for panel data, Sun et al. (2009) used kernel smoothing to estimate time-varying effects, while Harvey (1978) proposed a state space modeling (SSM, Durbin & Koopman, 2012) approach to deal with time-varying effects. The SSM approach has also been used in the context of generalized linear models (Durbin & Koopman, 2012; Harvey & Phillips, 1982; Helske, 2022). While the SSM approaches are highly flexible, they are often computationally demanding due to the assumed latent processes for the regression coefficients and an analytically intractable marginal likelihood when the observations are non-Gaussian (Andrieu et al., 2010; Vihola et al., 2020). Methods based on the varying coefficients models (Eubank et al., 2004; Haslbeck et al., 2021; Hastie & Tibshirani, 1993) can be computationally more convenient, especially when the varying coefficients are based on splines with suitable penalties (Lang & Brezger, 2004; Wood, 2020).

1.2. Earlier approaches

Traditional methods to study causal effects in panel data settings in social sciences and especially in economics include variants and extensions of fixed effect and random effect regression models which can also include, for example, instrumental variables and propensity score weighting and matching (for a general overview see (Wooldridge, 2010), and (Barban et al., 2020) for an application of matching for causal inference in life course research). Especially in psychology, cross-lagged panel models and their various extensions are often employed (Allison et al., 2017; Asparouhov et al., 2018; Bollen & Brand, 2010; Hamaker et al., 2015; Mulder & Hamaker, 2021; Zyphur et al., 2020) and typically formulated as SEMs (see also Pakpahan et al., 2017, for a perspective to life course research). Often T, the number of time points, is assumed to be small or moderate, say $T = 4$ or $T = 10$, whereas N, the number of individuals can be large. Except for dynamic SEM and related variants (Asparouhov et al., 2018), SEM approaches are mostly suitable for cases with relatively small T due to the computational complexity of their classical wide format formulation (Asparouhov et al., 2018). In contrast, our main interest is in scenarios with relatively large T, say tens, hundreds, or even thousands of time points (while N can still be small or large). For example, Helske et al. (2018) used a subsample of the German National Educational Panel Survey (Blossfeld et al., 2011) with $N = 1731$ and $T = 434$. Hudde & Jacob (2023) advocate the use of monthly data for more fine-grained life course trajectories in comparison to more commonly used yearly data which might miss some of the short-term effects of various life events (e.g., divorce).

In life course research, variations of Markovian models are gaining more interest for studying transitions in life course trajectories (Liao et al., 2022; Piccarreta & Studer, 2019), sometimes as complementary tools for non-probabilistic sequence analysis approaches (Helske et al., 2018, 2023; Pennoni & Piccarreta, 2017; Scott et al., 2024). Instead of directly modeling the dependency between consecutive observations, researchers sometimes choose to employ hidden Markov models (HMM, Bartolucci et al., 2013; MacDonald & Zucchini, 1997) and their extensions. In HMMs, it is assumed that an observation y_t depends on some underlying time-varying latent state z_t which has a discrete and finite state space (in contrast to SSMs where the state space is continuous). The latent states follow a Markov process, whereas the observations are

conditionally independent given the latent states. These models are highly flexible and encompass various special cases such as latent class models and mixture HMMs (Vermunt et al., 2008). Also, when extending SEM to non-Gaussian responses, the cross-lagged dependencies of different variables are defined on the latent level instead of the observational level (Asparouhov and Muthén, 2021; Bollen, 1989; Finney & DiStefano, 2013; Muthén, 1984), which can be then cast as an SSM. While their origins are in predictive tasks such as speech recognition (Baum & Petrie, 1966) and control theory (Kalman, 1960), especially HMMs and their variations have been popular in social sciences as exploratory and descriptive tools (Liao et al., 2022). For example, Helske et al. (2018) used HMMs to describe and visualize underlying latent structures of work and family trajectories jointly. Basic HMMs can also be extended in various ways, for example, Bartolucci et al. (2007) considered multiple observation sequences per individual, whereas Altman (2007) introduced a general mixed HMM approach which incorporates covariates and random effects to model both the observational and latent state distributions.

In typical HMM applications, causality is generally addressed only implicitly; if the model aligns with the data-generating process, the estimated effects can be interpreted causally. Some of the exceptions are (Bartolucci et al., 2016) which use the concept of potential latent variables and where the initial latent state probabilities and state transitions are assumed to depend on a discrete treatment variable with two or more categories. Bartolucci et al. (2023) consider a scenario where a binary treatment affects the initial, transition or emission probabilities of an HMM with data on pre- and post-treatment outcomes, similar to difference-in-differences methods (Callaway & Sant'Anna, 2021).

The aforementioned models are typically estimated using least squares or a maximum likelihood (ML) approach (e.g., Allison et al., 2017; Bollen & Brand, 2010; Hamaker et al., 2015; Wooldridge, 2010; Zyphur et al., 2020). An exception in the SEM framework is the dynamic SEM (Asparouhov et al., 2018) which is based on a Bayesian approach, while general multilevel models and SSMs are also often treated as Bayesian models (Bürkner, 2017; Gelman et al., 2013; Helske & Vihola, 2021; Triantafyllopoulos, 2021). On the other hand, HMMs and other discrete latent variable models are more often based on ML. In discrete latent variable models, the likelihood is typically multimodal (Frühwirth-Schnatter, 2006), which is especially problematic for Bayesian estimation which generally does not focus on finding a single global maximum, but the full posterior distribution.

In the ML approaches, the point estimates of the parameters are used to form plug-in estimates of some functions of interest. Especially for non-linear or non-Gaussian model components and long-term causal effects, relying on point estimates obtained from the ML method can severely underestimate the uncertainty and bias of the causal estimates as in forecasting problems (Chatfield, 2000). While bootstrap methods can offer a reasonable approach for uncertainty assessment in some scenarios, it can be problematic for serially dependent data and complex models with individual-specific effects (see e.g., (Morris, 2002 and discussion in (Helske, 2015, Chapter 4.1)). In contrast, Bayesian approaches provide automatic uncertainty quantification for any estimates of interest such as ACE, and enable a natural way to incorporate domain knowledge such as information from previous studies into the modeling workflow via the prior distribution, which can also be used to control model sparsity and induce desirable correlation structures (Gelman et al., 2013; Oganisian & Roy, 2021). However, while Bayesian methods can enable the estimation of complex models that are infeasible to estimate with ML, the Bayesian paradigm naturally has its own set of challenges, such as proper choice of priors and slow convergence rates and other computational issues with high-dimensional parameters or covariates (Gelman et al., 2013; Li et al., 2023).

1.3. The proposed approach

In this paper, we introduce the dynamic multivariate panel model (DMPM) for causal inference and general Bayesian modeling in the context of panel data. DMPMs can jointly estimate models consisting of multiple responses following various distributions, with time-invariant, time-varying, and individual-specific effects. When combined with the formal causal inference theory of structural causal models, DMPMs can be used to simulate both short-term and long-term counterfactual predictions under various scenarios of interest. We illustrate our approach using simulated and real data modeled with the R (R Core Team, 2024) package dynamite (Tikka & Helske, 2023) that provides an interface for fitting DMPMs. The codes to reproduce all analyses and the dynamite package can be found on GitHub (https://github.com/helske/dmpm).

Given the generality of DMPMs, they share similarities with various established modeling techniques. For example, classic VAR models can be regarded as a specific instance of a Gaussian DMPM by treating the response variables as a multivariate Gaussian response variable, predicted by its lagged values. Moreover, DMPMs provide support for non-Gaussian responses and random effects in a manner akin to generalized linear mixed models. DMPM also extends basic Markov models with explanatory variables, time-varying effects, and joint modeling of multiple response variables with mixed distributions. Multiple responses consisting of mixed distributions have also been considered earlier in the HMM context for example by Bartolucci et al. (2013). Time-varying effects defined via splines have been studied earlier for example in a VAR setting by Haslbeck et al. (2021), but instead of estimating the smoothness penalization for the splines via cross-validation, DMPM models this via a random walk prior where the corresponding standard deviation is estimated jointly with other model parameters.

The paper is structured as follows. Section 2 introduces SCMs and our general causal inference workflow for panel data combined with Bayesian inference. Section 3 provides formal details of DMPMs and discusses their estimation. Section 4 demonstrates the use of DMPMs in practice via examples using both real and synthetic data. Section 5 concludes the paper with a discussion.

2. Causal inference for panel data

We consider the structural causal modeling framework of Pearl (2009). In this framework, a variable X is viewed as being the direct cause of another variable Y, if Y is a function of X. An SCM consists of a set of observed variables, a set of unobserved background variables, a set of functions that defines the causal relationships between all variables, and a joint probability distribution over the background variables. This distribution also induces a joint probability distribution over the observed variables through the functional relationships of the model. Finally, the functional relationships induce the causal diagram.

The SCM framework enables the study of interventions which are actions that change the functional structure of the model. Formally, interventions are operationalized by the do$(\cdot)$ operator. For instance, an intervention that forces a variable X to take the value x irrespective of the value it would have attained otherwise is denoted by do$(X = x)$. Such an intervention targeting a causal model essentially changes the function that defines X into a new function that outputs the constant value x. The distribution of a variable Y of interest under this intervention is called an interventional distribution and is denoted by $p(Y = y|\text{do}(X = x))$ or just $p(y|\text{do}(x))$ for short. Whenever it is possible to represent the interventional distribution using only the observed probability distribution of the SCM, we say that the interventional distribution is identifiable.

The term "causal effect" refers to different quantities in literature. Here we use the term to refer to the full interventional distribution $p(y|\text{do}(x))$, from which quantities such as the expected value of the interventional distribution $\text{E}(y|\text{do}(x))$ and ACEs such as

$$\text{E}(y|\text{do}(X = a)) - \text{E}(y|\text{do}(X = b)),$$

where a and b are two possible values of X, can be derived.

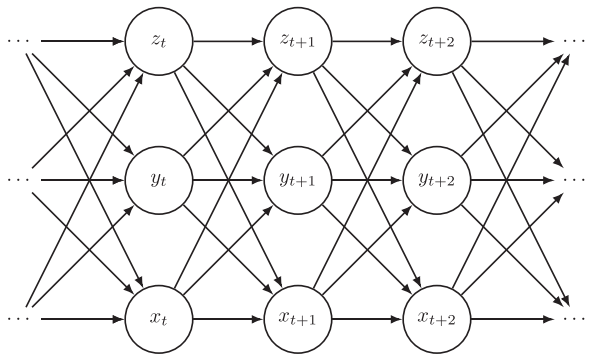


Fig. 1. A causal graph of three responses y, x, and z, with arrows corresponding to the assumed direct causal effects. A cross-section at times t, $t+1$, and $t+2$ is shown, and only some of the arrows and nodes are fully depicted for readability.

Our choice of the SCM framework is motivated by several factors. First, a graphical representation of the system under study via the causal graph easily communicates all causal assumptions about the variables of interest, without having to explicitly specify conditional independence properties of counterfactual quantities such as those encapsulated by the ignorability and exchangeability assumptions in the potential outcomes framework (Hernán and Robins, 2020; Rosenbaum & Rubin, 1983). A graphical representation can be understood even by those not well-versed in causal inference theory and jargon. Second, as a consequence of the graphical representation, we do not need to make any additional assumptions regarding the identifiability of causal effects, as the sufficient and necessary conditions of causal effect identifiability are encoded in the causal diagram. This holds for a wide variety of scenarios in terms of the available data including purely observational data (Shpitser & Pearl, 2006a,b), a combination of experimental and observational data (Kivva et al., 2022; Lee et al., 2020a), and data collected from heterogeneous domains (Lee et al., 2020b; Tikka et al., 2021). Thus, even though we focus on observational settings, the approach we present in this section could be adapted to scenarios where some experimental data is also available. Third, we gain access to full interventional distributions instead of just point estimates which in turn allows us to quantify the uncertainty of causal effects and compute other quantities of interest, such as probabilities of specific events under the intervention, for instance. This feature also couples seamlessly with a Bayesian approach, as we will demonstrate.

The general causal inference pipeline in the SCM framework proceeds as follows. First, the causal model is constructed based on a combination of expert knowledge and causal discovery methods. Next, the causal effects of interest are identified by using do-calculus (Pearl, 1995), a graphical criterion, such as the backdoor or the frontdoor criterion (Pearl, 2009), or an identifiability algorithm such as the ID algorithm (Shpitser & Pearl, 2006b). If the effects of interest are identifiable, then the next step is estimation, where the appropriate statistical modeling tools should be leveraged. Next, we describe this general process in greater detail in the context of panel data and long-term causal effects. Our approach shares similarities with the methods employed by Bijlsma & Wilson (2019) and Nisén et al. (2022), who utilized ML-based generalized linear models with backdoor adjustment and bootstrapping to estimate the causal determinants of fertility and the consequences of delayed parenthood. Here we present a theoretical framework for SCM-based causal inference using panel data that is more broadly applicable. Furthermore, the Bayesian DMPM we present in Section 3 offers greater flexibility in modeling and estimating causal effects in the context of panel data.

2.1. Estimating temporal causal effects

We distinguish two types of long-term causal effects of interest: the effect of an intervention at a single time point on a future time point $p(y_{t+k}|\text{do}(X_t = x))$ which we refer to as an atomic intervention and the effect of interventions at multiple sequential time points $p(y_{t+k}|\text{do}(X_t = x_1, \ldots, X_{t+k} = x_k))$ which we refer to as a recurring intervention. Naturally, it is possible to distinguish other types of causal effects of interest and the value assignments of the interventions could also be based on the values of other variables in the system such as under dynamic treatment regimes (Murphy, 2003). However, in this paper, we focus on atomic and recurring interventions for simplicity. Note that these two types of *interventions* are different, albeit related, concepts than the short-term and long-term *effects*.

We consider the causal graph in Fig. 1 where we have three types of response variables: nodes y_t represent the response variable(s) of interest, x_t the intervention variable(s), and z_t additional covariate(s) (this decomposition is done only to simplify the exposition and is in general not necessary). In the figure, each edge corresponds to the regressive and auto-regressive (either time-invariant or time-varying) effects of the previous time point on the next. For example, x_t could be geographical mobility that influences individual's educational attainment y_t and employment status z_t at time t. Education level, in turn, impacts employment opportunities, or an individual might relocate to pursue further education. Finally, lack of employment can encourage individuals to seek further education or to move in order to improve their employment opportunities. While x_t, y_t, and z_t could also depend on some time-invariant unobserved variables and observed strictly exogenous variables, we omit them from the graph to keep it more readable. We also do not draw any model parameters in the figure, as we have not made any parametric assumptions about the distributions of the variables at this point. In this example, there are no contemporaneous dependencies between the responses, and the highest-order lag dependency is of first-order. While we focus on this specific graph for simplicity, the following methods generalize to other graphs with a similar structure. For example, there could be more or fewer variables of interest, or some of the edges might not exist in the causal graph. While each node can represent multiple variables, here we consider them to be univariate without loss of generality (Tikka et al., 2023). Our interest is in estimating the causal effect of x_t on y_{t+k}, but other effects could be estimated analogously. In particular, if the causal graph is time-invariant in the sense that the graph structure does not depend on t, then all the effects of interventions at time t on any response variables y_t, x_t, or z_t at future time point(s) are identifiable and estimable using the following backdoor adjustment approach.

By using the backdoor adjustment, we find a non-parametric identifying formula for the interventional distribution:

$$p(y_{t+k}|\text{do}(x_t)) = \sum_{y_t, z_t} p(y_{t+k}|y_t, x_t, z_t)p(y_t, z_t). \tag{1}$$

This result is easy to see from Fig. 1 because at any time t, the observations other than the variable being intervened on block all backdoor paths to any responses at time $t+k$. The summations in the formula should be understood as integrals when the corresponding variable is continuous. Equation (1) shows that in order to estimate $p(y_{t+k}|\text{do}(x_t))$, we need to model the distributions $p(y_{t+k}|y_t, x_t, z_t)$ and $p(y_t, z_t)$. Fortunately, we can avoid estimating the marginal distribution $p(y_t, z_t)$, because simply computing the average of $p(y_{t+k}|y_t, x_t, z_t)$ over the individuals serves as a Monte Carlo approximation of $p(y_{t+k}|\text{do}(x_t))$ (Hernán and Robins, 2020). However, in some instances, it may be of interest to also model $p(y_t, z_t)$ explicitly. Note that in cases where we have data only on one individual with no external knowledge of $p(y_t, z_t)$, we can still estimate conditional interventional distributions $p(y_{t+k}|\text{do}(x_t), y_t, z_t)$.

In principle, we could construct a statistical model directly for the distribution $p(y_{t+k}|y_t, x_t, z_t)$, but say we are also interested in $p(y_{t+k}|\text{do}$

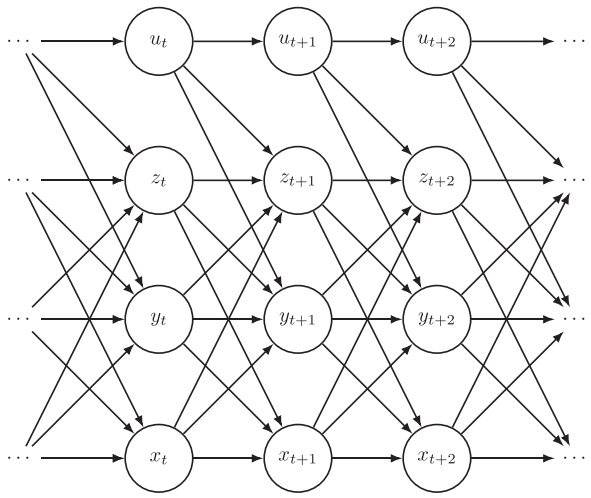


Fig. 2. A causal graph of three responses y, x, z, and unobserved time-varying confounder u, with arrows corresponding to the assumed direct causal effects. A cross-section at times t, $t+1$, and $t+2$ is shown.

(x_{t-1})), or some other intervention at a time other than t. The backdoor criterion still applies, but now we would need a new model for $p(y_{t+k}|y_{t-1}, x_{t-1}, z_{t-1})$, and similarly for other potentially interesting causal effects. Fortunately, we can rewrite the expression (1) such that we only have to model the relationship of observations at time t given the previous time point(s). Let $w_t = (y_t, x_t, z_t)$. Then the first term of equation (1) can be written as

$$p(y_{t+k}|w_t) = \sum_{w_{t+1},\ldots,w_{t+k-1},} p(y_{t+k}|w_{t+k-1})p(w_{t+k-1}|w_{t+k-2})\cdots p(w_{t+1}|w_t). \quad (2)$$

We can simulate the distributions on the right-hand side of equation (2) directly by constructing a statistical model. For fixed model parameters, denoted by $\boldsymbol{\theta}$ (we will henceforth use a bold font to denote vectors), we can simulate new values of y_{t+k} for each of the N individuals in the data by first simulating w_{t+1} using the observed values of y_t, z_t, and the value of x_t fixed by the intervention, and then repeating the same process for each time point from $t+2$ to $t+k$, leading to samples from $p(y_{t+k}|\text{do}(x_t), \boldsymbol{\theta})$ which takes into account the variability due to $p(y_t, z_t)$ and the trajectories $w_t, \ldots, w_{t+k}$. Repeating this over all M posterior samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M$ gives us a sample from the posterior of the post-interventional distribution $p(y_{t+k}|\text{do}(x_t))$.

For the ACE

$$\text{E}(y_{t+k}|\text{do}(x_t = a)) - \text{E}(y_{t+k}|\text{do}(x_t = b)),$$

contrasting two values a and b, we compute the expected values by replacing the sampling from $p(y_{t+k}|w_{t+k-1}, \boldsymbol{\theta})$ in the earlier approach with the computation of $\text{E}(y_{t+k}|w_{t+k-1}, \boldsymbol{\theta})$. Then, marginalizing over the distribution $p(y_t, z_t)$ we obtain an estimate for $\text{E}(y_{t+k}|\text{do}(x_t), \boldsymbol{\theta})$. By repeating the same procedure for both values a and b, and by subtracting the former estimate from the latter, we obtain the ACE conditional on the model parameters $\boldsymbol{\theta}$. The posterior distribution of the ACE can then be obtained by repeating this computation over the posterior samples of the model parameters, from which we can compute for example the posterior mean of the ACE as

$$\frac{1}{M}\sum_{j=1}^{M} \text{E}(y_{t+k}|\text{do}(x_t = a), \boldsymbol{\theta}_j) - \text{E}(y_{t+k}|\text{do}(x_t = b), \boldsymbol{\theta}_j),$$

where M is the number of posterior samples and $\boldsymbol{\theta}_j$ is the jth posterior sample. It is straightforward to extend this approach to a situation where the model contains higher-order lagged response values as covariates or contemporaneous dependencies between responses.

As for the posterior distribution of $p(y_{t+k}|\text{do}(x_t))$, the posteriors of ACE estimates for $k > 1$ contain not only the variation due to the parameter uncertainty but also the due to the simulation of a finite number of trajectories $w_t, \ldots, w_{t+k}$. This latter Monte Carlo variation can be reduced by sampling multiple trajectories per individual.

Finally, we consider the case in Fig. 2, which has the same structure as Fig. 1 but with the addition of an unobserved time-varying confounder denoted by u, which is common to each individual.

For simplicity, assume that y_t is linear-Gaussian and

$$E(y_{t,i}|\boldsymbol{\theta}) = \alpha + \beta_y y_{t-1,i} + \beta_x x_{t-1,i} + \beta_z z_{t-1,i} + \beta_{uz} u_{t-1} z_{t-1,i},$$

i.e., the unobserved confounder u acts as an effect modifier for z. Now let $\delta_t = \beta_z + \beta_{uz} u_{t-1}$ so that $E(y_{t,i}|\boldsymbol{\theta}) = \alpha + \beta_y y_{t-1,i} + \delta_t z_{t-1,i}$, i.e., we have a linear model with a time-varying coefficient δ_t. Consequently, with a suitable assumption for the functional form of δ_t (e.g., a spline), we can still estimate the causal effect of x on y even if our causal graph contains unobserved time-varying confounders. Naturally, the causal effect of z on y (β_z) is not identifiable without further assumptions. Also, in theory, the values of u could vary by individual, and we would still obtain an identifiable causal effect of x, but in such a case the time-varying effects δ_t should also be individual-specific, which in practice could lead to considerably higher computational burden. In general, when unobserved confounders are present (either time-invariant or time-varying), the identifiability of causal effects of interest has to be assessed based on the causal assumptions, i.e., the causal graph (Shpitser & Pearl, 2006b). With parametric assumptions, such as linear causal relationships, additional causal effects can sometimes be identified, for example by using instrumental variables (Imbens & Angrist, 1994; Pearl, 2009) or more advanced methods such as auxiliary cutsets (Kumor et al., 2020).

2.2. Incompatibility of long-term effects and latent Markov models

Compared to Markovian models operating on the observational level, latent level models such as HMMs and SSMs can provide

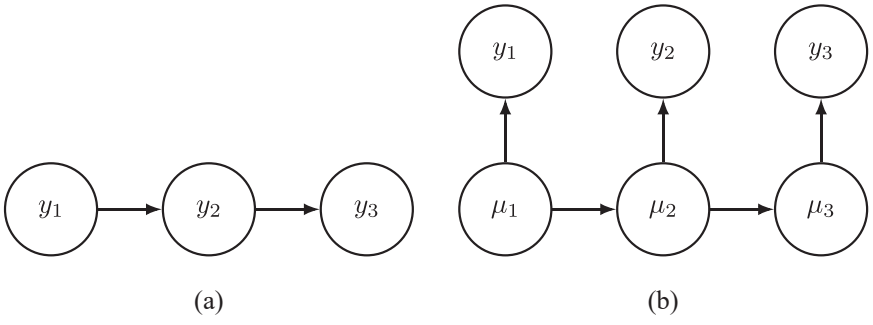


Fig. 3. Example causal graphs illustrating a direct causal chain between variables of interest (a) and a latent process that generates the observations (b). Intervening on the observations y_t in (b) has no effect on future observations. (a) A causal graph where y_1 is a direct cause of y_2, and y_2 is a direct cause of y_3. (b) A causal graph where the variables y_1, y_2, and y_3 do not form a causal chain but are instead realizations of a latent process μ.

additional flexibility in modeling complex processes, and they allow more straightforward handling of missing response variables (Durbin & Koopman, 2012; MacDonald & Zucchini, 1997). Fig. 3 illustrates the difference between a Markov model (Fig. 3a) and a hidden Markov model (Fig. 3b) for three time points $t = 1, 2, 3$.

The conditional independence assumption of observations in basic HMMs and SSMs is problematic in terms of causal inference, especially when the interest is in estimating long-term causal effects. In general, estimating such effects involves simulating new observations given the observations at the current time point; a dependency which is broken in Fig. 3b. In the graph of Fig. 3a, there is a direct causal chain between the variables, and manipulating (i.e., intervening) y_1 affects the interventional distributions of y_2 and y_3 (the future values). In the graph of Fig. 3b, however, an intervention on y_1 does not affect the future values of y (y_2 and y_3); the intervention should target the unobserved latent variable μ_1. While this can be less of a problem with Gaussian observations, when it can be reasonable to assume that the observed variables y_t are noisy versions of the latent variables μ, and we could in theory intervene on μ, such an interpretation does not generalize to other observational distributions such as categorical and binomial distributions if the latent level is assumed to be Gaussian as in SEMs and typical SSMs. For example, for the binomial distribution with a logit link, the latent level corresponds to a probability (on the logit scale), and it can be challenging to conceptualize an intervention on this probability. Intervening on the latent probability is also very different than intervening on the observed quantities as the former manipulates the underlying system (e.g., the probability that a rare virus is present in a community), whereas the latter manipulates the realization from this system (the virus causes hospitalizations). Using an intervention on a latent level process as an indirect way of estimating the causality on the observational level can be problematic also in practice. Consider a DAG with a structure as in Fig. 3b with binary observations $y_t \sim$ Bernoulli(logit$^{-1}(\mu_t)$) and a latent random walk process $\mu_t = \mu_{t-1} + \epsilon_t$, $\epsilon_t \sim$ N(0, 1). Now, estimating the causal effect $E(y_{t+k}|\text{do}(y_t = 1))$ indirectly by intervening on latent $\mu_t = \text{logit}(P(y_t = 1))$ corresponds to $E(y_{t+k}|\text{do}(\mu_t = \infty))$, but fixing $\mu_t = \infty$ leads to $\mu_{t+1} = \cdots = \mu_{t+k} = \infty$ and thus $E(y_{t+k}|\text{do}(\mu_t = \infty)) = 1$ for all $k \geq 0$ with probability 1.

While Markov models, where the observations directly depend on past observations, are more suitable than standard HMMs for long-term causal inference, it is naturally possible to extend HMMs so that the observation equation contains lagged responses as predictors, which corresponds to a graph similar to the one in Fig. 3b with additional edges from the previous observation y_1 to the current observation y_2 and from y_2 to y_3. Such double chain Markov models are considered for example by Berchtold (1999). These models combine some of the strengths of both traditional Markov and hidden Markov models, but they can be computationally very demanding in panel data settings like HMMs, as we need to estimate separate a hidden state trajectory for each individual with a typically multimodal likelihood function.

3. The dynamic multivariate panel model

We consider data consisting of T time points and N individuals where we have measured $C \geq 1$ response variables from each individual. In other words, the response variables of individual i are $\boldsymbol{y}_{t,i} = (y_{1,t,i}, \ldots, y_{c,t,i}, \ldots, y_{C,t,i})$, $t = 1, \ldots, T$, $i = 1, \ldots, N$. The response variables can be continuous or discrete. When referring to the random variable itself, we simply drop the index related to the individual and write y_t, or $y_{c,t}$ when referring to a specific response. Each element of $\boldsymbol{y}_{t,i}$ can depend on other responses at the same time point t, the past values of the responses $\boldsymbol{y}_{t-\ell,i}$, $\ell = 1, \ldots$ (with potentially different sets of past values for different responses), as well as additional covariates $\boldsymbol{x}_{t,i}$ which are treated as non-stochastic (strictly exogenous) in the subsequent modeling. The covariates $\boldsymbol{x}_{t,i}$ can be either time-varying or time-invariant. The first L time points are treated as fixed data, where L is the largest order of lag dependence in the model. For instance, $L = 1$ for a model where a first-order lag dependence is the highest.

We assume that the joint distribution of $\boldsymbol{y}_{t,i}$ given past observations and current covariates factorizes as follows for all $t = L + 1, \ldots, T$ and $i = 1, \ldots, N$

$$p_t(\boldsymbol{y}_{t,i}|\boldsymbol{y}_{1:t-1,i}, \boldsymbol{x}_{t,i}) = \prod_{c=1}^{C} p_{c,t}(y_{c,t,i}|\boldsymbol{y}_{\pi(c),t,i}, \boldsymbol{y}_{1:t-1,i}, \boldsymbol{x}_{t,i}), \tag{3}$$

where $\boldsymbol{y}_{1:t-1,i}$ denotes the past values of all responses $(\boldsymbol{y}_{1,i}, \ldots, \boldsymbol{y}_{t-1,i})$, π is an ordering of the responses, and $\boldsymbol{y}_{\pi(c),t,i}$ denotes the responses before response number c in this ordering. This ordering captures the contemporaneous dependency structure of the model and rules out cyclic dependencies. Because our model inherently includes a temporal order, it is not sensible to allow cyclic structures. Also, assumptions of contemporaneous feedback loops are rarely true in practice but originate from a temporal view of the problem that is too simplified (Murray & Kunicki, 2022). The conditional distributions $p_{t,c}$ can be different for different response variables, and the parameters of these distributions can depend on t, i.e., we allow the dynamics of our system to evolve over time.

Given suitable link functions depending on our distributional assumptions, we define the linear predictors $\boldsymbol{\eta_{c,t}}$ for all responses $\boldsymbol{y_{c,t}}$ with the following general form:

$$\eta_{c,t,i} = \alpha_{c,t} + \boldsymbol{u}_{c,t,i}^\top \boldsymbol{\beta}_c + \boldsymbol{w}_{c,t,i}^\top \boldsymbol{\delta}_{c,t} + \boldsymbol{z}_{c,t,i}^\top \boldsymbol{\nu}_{c,i} + \boldsymbol{\lambda}_{c,i}^\top \boldsymbol{\psi}_{c,t} \tag{4}$$

where $\alpha_{c,t}$ is the common, possibly time-varying, intercept term. The vectors $\boldsymbol{u}_{c,t,i}$, $\boldsymbol{w}_{c,t,i}$, and $\boldsymbol{z}_{c,t,i}$ define the covariates corresponding to the vectors of time-invariant effects $\boldsymbol{\beta}_c$, time-varying effects $\boldsymbol{\delta}_{c,t}$, and individual-specific random effects $\boldsymbol{\nu}_{c,i}$, respectively. As in multilevel models, $\boldsymbol{z}_{c,t,i}$ can and often does contain the same covariates as in $\boldsymbol{u}_{c,t,i}$ or $\boldsymbol{w}_{c,t,i}$, so that the individual-specific effects $\boldsymbol{\nu}_{c,i}$ are deviations from the overall effect $\boldsymbol{\beta}_c$ or $\boldsymbol{\delta}_{c,t}$. In general covariates in $\boldsymbol{u}_{c,t,i}$ and $\boldsymbol{w}_{c,t,i}$ should be unique to ensure the identifiability of the corresponding effects, although it is possible to separately estimate for example the average level of the effect and the temporal deviation from this with suitable priors or constraints (for example by using sum-to-zero constraints on $\sum_{t=1}^{T} \boldsymbol{\delta}_{c,t}$). Note that the covariates in the vectors $\boldsymbol{u}_{c,t,i}$, $\boldsymbol{w}_{c,t,i}$, and $\boldsymbol{z}_{c,t,i}$ may contain values of other response variables $\boldsymbol{y}_{\pi(c),t,i}$ that appear before $y_{c,t,i}$ in the ordering π, values of lagged response variables $\boldsymbol{y}_{1:t-1,i}$, and values of other non-stochastic covariates $\boldsymbol{x}_{t,i}$. These can also include for example individual-specific means for Mundlak's approach (Mundlak, 1978) which solves the critique of the random effects by the proponents of the fixed effect models (Bell et al., 2018, 2019). The random effects $\boldsymbol{\nu}_{1,i}, \ldots, \boldsymbol{\nu}_{C,i}$ are assumed to follow a zero-mean Gaussian distribution, and they can be correlated both within and between response variables. We will use the symbols σ and ρ to denote the standard deviations and correlations of random effects, respectively.

The final term $\boldsymbol{\lambda}_{c,i}^\top \boldsymbol{\psi}_{c,t}$ is an alternative to a common time-varying intercept $\alpha_{c,t}$, consisting of latent individual loadings $\boldsymbol{\lambda}_{c,i}$ and a latent dynamic factor $\boldsymbol{\psi}_{c,t}$, which can be correlated across response variables similarly as the common correlated effects discussed by Thombs (2022), for example. In a single response model, this term can be regarded as a special case of dynamic latent factor models (Bai & Wang, 2015; Geweke, 1977) with only a single factor, except that we treat $\boldsymbol{\psi}_{c,t}$ as a spline instead of a stationary autoregressive process. The purpose of this term is to model how individuals respond differently to some unobserved, common process $\boldsymbol{\psi}_{c,t}$. For example, our response variable could be consumer spending, affected by the changing environment modeled by $\boldsymbol{\psi}_{c,t}$ that is common to the individuals. Here $\boldsymbol{\psi}_{c,t}$ could reflect the latent overall economic status of the country (instead of a specific measure such as inflation rate), and $\boldsymbol{\lambda}_{c,i}$ would then describe how the individuals react differently to these latent changes in the economy. In a model with multiple responses, the latent factors $\boldsymbol{\psi}_{c,t}$ can be allowed to correlate with each other. While coefficients $\boldsymbol{\delta}_{c,t}$ and latent factors $\boldsymbol{\psi}_{c,t}$ can capture common time-varying confounding, we assume that the

underlying causal graph is time-invariant in the sense that there are no individual-specific unobserved variables affecting only some distinct time points.

We define the time-varying coefficients $\boldsymbol{\delta}_{c,t}$ (and similarly for the time-varying intercept $\boldsymbol{\alpha}_{c,t}$ and latent factor $\boldsymbol{\psi}_{c,t}$) using Bayesian P-splines (penalized B-splines) (Eilers & Marx, 1996; Lang & Brezger, 2004) as

$$\delta_{c,t,k} = \boldsymbol{b}_t^\top \boldsymbol{\omega}_{c,k}, \quad k = 1, \ldots, K,$$

where K is the number of covariates, $\boldsymbol{b}_t$ is a vector of B-spline values at time t obtained from a cubic B-spline basis that we assume has been constructed with equally spaced knots on the time interval from $L+1$ to T with D degrees of freedom, and $\boldsymbol{\omega}_{c,k}$ is a vector of corresponding spline coefficients. In general, the degrees of freedom, i.e., the number of B-splines D used for constructing the splines for the study period $L+1, \ldots, T$ can be chosen freely, but too large D can result in overfitting. To mitigate this, we define a random walk prior (Lang & Brezger, 2004) for $\boldsymbol{\omega}_{c,k}$ as

$$\begin{aligned} \omega_{c,k,1} &\sim p(\omega_{c,k,1}), \\ \omega_{c,k,d} &\sim \mathrm{N}\left(\omega_{c,k,d-1}, \tau_{c,k}^2\right), \quad d = 2, \ldots, D \end{aligned}$$

with a user-defined prior $p(\omega_{c,k,1})$ on the first coefficient, which due to the structure of b_1 corresponds to the prior on $\delta_{c,L+1,k}$. Here, the parameter $\tau_{c,k}$ controls the smoothness of the spline curves, which can be estimated along with other parameters of the model. Thus, the exact value of D is typically not too important, however, it is possible to use for example cross-validation approaches such as leave-one-out or leave-future-out cross-validation (Bürkner et al., 2020; Vehtari et al., 2016) for choosing the value of D.

Conforming with the structural causal modeling, we consider interventions on responses and non-stochastic covariates but not on the model parameters. For instance, we can consider the causal effects of responses on other responses, or the causal effects of non-stochastic predictors on responses. Importantly, we note that we do not directly imbue the model parameters of a DMPM with causal interpretations, but instead consider causal effects through interventions and interventional distributions. Whether and how the model parameters can be interpreted causally depends on the specifics of the causal graph and the model (Westreich & Greenland, 2013). Analogously to general Bayesian analysis, where it is straightforward to perform posterior inference for a quantity of interest which can be a complex function of the estimated posterior distribution of the parameters (Gelman et al., 2013), the estimated full interventional distribution provides flexibility in answering causal queries in an interpretable way, tailored to the specific research question.

While DMPMs are multivariate models by definition, we have thus far only considered the scenario where each response is related to a single linear predictor. We can easily extend the definition of the linear predictor in equation (4) to also accommodate multivariate distributions such as the multivariate Gaussian distribution and other distributions that are typically modeled with multiple components, such as the categorical distribution with a linear predictor for each category, excluding the reference, via the softmax link. Formally, we can accomplish this by simply replacing the index c with c, s in equation (4) where s denotes the index of the dimension, $s = 1, \ldots, S(c)$, and $S(c)$ is the number of dimensions of response c.

3.1. Estimation

We estimate the unknown parameters of our model using a Bayesian approach by combining the likelihood implied by model (3) with priors on the time-invariant regression coefficients $\boldsymbol{\beta}_c$, intercept α_c (or $\alpha_{c,L+1}$ in case the intercept is assumed to be time-varying), time-varying coefficients $\boldsymbol{\delta}_{c,L+1}$, the random walk standard deviation parameters $\boldsymbol{\tau}$, the standard deviations of the latent loadings $\lambda_{c,i}$ and of the individual-specific random effects $\nu_{c,i}$. For correlated random effects, we assume that their covariance matrix has been parametrized using the Cholesky factorization, and the prior is specified for the Cholesky factor (Lewandowski et al., 2009). In addition, there may be distribution-specific parameters, i.e., parameters that are not related to the linear predictors $\boldsymbol{\eta}_{c,t}$ of equation (4), but those that parametrize the density functions $p_{c,t}$, such as the standard deviation of a Gaussian distribution or the dispersion parameter of a negative binomial distribution. The choice of priors is not always a straightforward task, but general recommendations have been presented that are suitable for a wide range of models and data (Gelman et al., 2013). Assuming that the priors are not specified so that they exclude the true parameter values, the Bayesian estimation of DMPM provides asymptotically consistent parameter estimates under similar regularity conditions as in the case of ML estimation (see, e.g., Van der Vaart, 1998).

The joint posterior distribution of the model parameters can be estimated via Markov chain Monte Carlo (MCMC) methods, for example with the R package dynamite which uses computationally scalable MCMC algorithms provided by the probabilistic programming language Stan (Stan Development Team, 2022) for the posterior sampling. By using the posterior samples obtained via MCMC, we can evaluate the expected values of arbitrary functions of the model parameters and the data. These estimates do not suffer from bias caused by the potential non-linearity of the functions of interest and naturally account for parameter uncertainty by marginalizing over the posterior samples. This is particularly crucial in the context of causal inference when computing ACEs, as they are defined as the expected value of a function of a joint distribution (Pearl, 2009).

4. Examples

Because the DMPM is very general, it is not feasible to fully explore all variations of distributions and effects. Instead, we illustrate some of the core capabilities of the DMPM using three examples. In the first example, we use simulated data where the ground truth is known, whereas in the latter examples, we use openly available data for reproducibility purposes. While there are many survey and register data sources of large and complex panel data that could be interesting to analyze using DMPMs, these datasets are not easily accessible. The real data examples are not fully realistic due to the lack of some key variables in these open data, and as such we do not claim that the subject-specific conclusions are necessarily valid. All analyses were carried out in R using the R packages dynamite, dplyr(Wickham et al., 2023), and ggplot2(Wickham, 2016). The code for reproducing all examples is provided as supplementary material on GitHub. The appendix contains further details of the examples. All provided posterior intervals are equal-tailed for ease of interpretation.

4.1. Bivariate Gaussian Model

As the first example, we consider data generated from the following model

$$\begin{aligned} y_{1,t,i} &= 0.6y_{1,t-1,i} + 0.4y_{2,t-1,i} + x_{t,i} + \epsilon_{1,t,i} \\ y_{2,t,i} &= -0.1y_{1,t-1,i} + 0.9y_{2,t-1,i} + 0.4x_{t,i} + \epsilon_{2,t,i} \\ \epsilon_{1,t,i} &\sim \mathrm{N}(0, 0.4^2) \\ \epsilon_{2,t,i} &\sim \mathrm{N}(0, 0.4^2) \end{aligned} \tag{5}$$

where $t = 1, \ldots, 100$, $i = 1, \ldots, 500$, and $x_{t,i}$ is a known binary variable where value 1 acts as an indicator for some intervention. Note that in a general DMPM, the variable x_t does not need to be binary; its values can vary per individual and depend on the past values of $y_{1,t}$ and $y_{2,t}$ (in which case the variable x_t should be modeled jointly with $y_{1,t}$ and $y_{2,t}$). We consider two simple scenarios: in the first scenario $x_{t,i}$ is targeted by an atomic intervention with $x_{t,i} = 1$ for $t = 81$ and zero otherwise for

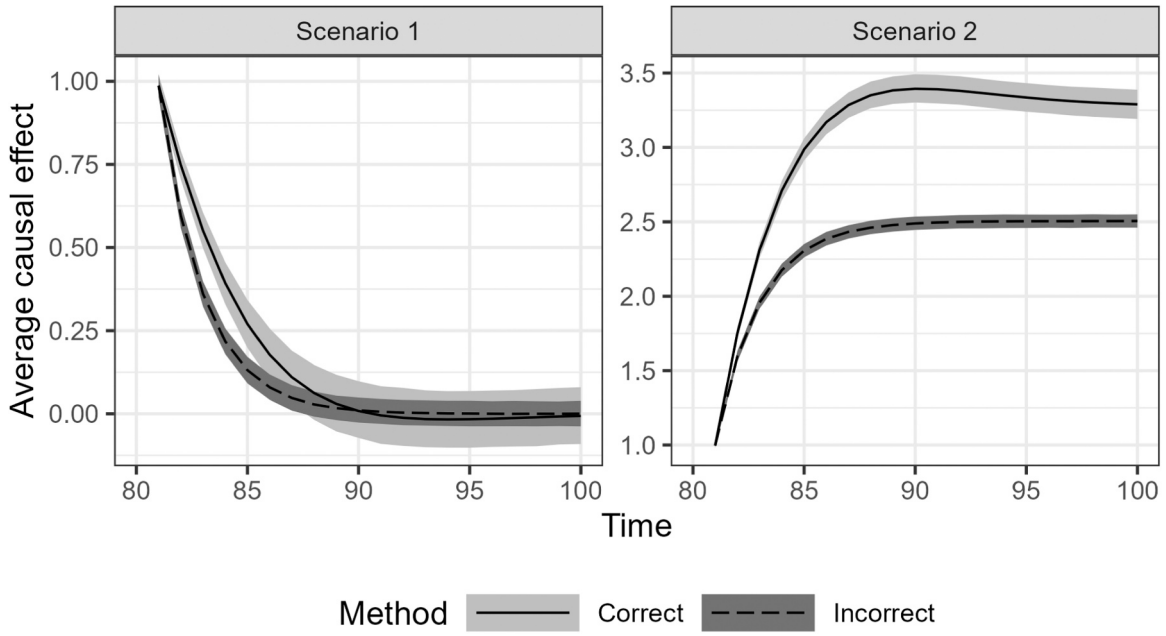


Fig. 4. Posterior means and 95% marginal quantiles of $\mathrm{ACE}_1(t, k)$ in the left panel and $\mathrm{ACE}_2(t, k)$ in the right panel, based on simulation of both y_1 and y_2 (correct method, solid line) and only y_2 (incorrect method, dashed line) in the bivariate Gaussian model example. In scenario 1, $x_{t,i}$ is targeted by an atomic intervention. In scenario 2, $x_{t,i}$ is targeted by a recurring intervention.

each individual. In the second scenario, the intervention assigns $x_{i,t} = 1$ if $t > 80$ and zero otherwise, corresponding to a recurring intervention. In both scenarios, the interest is in assessing the effect of the intervention on $y_{1,t}$, $t = 81, \ldots, 100$ compared to a counterfactual setting where no intervention was made ($x_{t,i} = 0$ for all t).

Using model (5), we simulated trajectories of $y_{1,t}$ and $y_{2,t}$ for $N = 500$ individuals under both scenarios, where we first estimated model (5) and then, using the posterior samples of the model parameters, simulated new realizations of $y_{1,t}$ and $y_{2,t}$ for the post-intervention period using both the observed values of x_t and for the counterfactual case of $x_t = 0$ for all t. We also repeated this posterior predictive simulation where only $y_{1,t}$ was simulated and $y_{2,t}$ was kept fixed to its observed values, ignoring the fact that according to our data-generating process, both the covariate x_t and the past value of the response $y_{1,t-1}$ affect $y_{2,t}$.

The ACEs of the two interventions at time $t + k$ are

$$\begin{aligned}\mathrm{ACE}_1(t,k) &= \mathrm{E}(y_{1,t+k}|\mathrm{do}(x_t = 1, x_{t+1} = 0, \ldots, x_{t+k} = 0)) - \mathrm{E}(y_{1,t+k}|\mathrm{do}(x_t = 0, \ldots, x_{t+k} = 0)),\\ \mathrm{ACE}_2(t,k) &= \mathrm{E}(y_{1,t+k}|\mathrm{do}(x_t = 1, \ldots, x_{t+k} = 1)) - \mathrm{E}(y_{1,t+k}|\mathrm{do}(x_t = 0, \ldots, x_{t+k} = 0)).\end{aligned}$$

Fig. 4 shows the posterior estimates of the mean and 95% marginal quantiles of $\mathrm{ACE}_1(k)$ and $\mathrm{ACE}_2(k)$ based on the correct counterfactual simulation of both $y_{1,t}$ and $y_{2,t}$, and the incorrect simulation of only $y_{1,t}$. As expected, the simulation of only $y_{1,t}$ leads to incorrect causal effect estimates, more specifically underestimation in this case. The error is especially large in the case of the recurring intervention, while the error diminishes for the atomic intervention as the effect of x_{81} on $y_{1,t}$ and $y_{2,t}$ decreases with increasing t.

From the left panel of the figure, we can observe that the ACE of the atomic intervention eventually levels off to zero in the first scenario. This occurs because even though the coefficient of x_t for $y_{1,t}$ is 1, the coefficients of $y_{1,t-1}$ and $y_{2,t-1}$ are less than 1. Thus, when we initially intervene by setting x_t to 1 at time $t = 81$, the ACE is exactly 1, and the effect of this intervention diminishes over time because x_t was set to 0 for the latter time points, and the initial intervention will affect y_1 only through its past values (and the past values of y_2). The same phenomenon occurs even if we do not model y_2. In the right panel depicting the second scenario with the recurring intervention, the importance of the correct modeling approach becomes evident. Instead of propagating the initial intervention that forces x_t to have the value 1 only at the first time point, we keep applying this intervention at each time point. As the intervention is applied constantly, and because x_t affects y_1 both directly and via past values of y_1 and y_2, we must correctly model both responses. We can also compute the true ACE using equation (5) by solving the recurrence relation for $y_{1,t}$ to obtain $\lim_{k\to\infty} \mathrm{ACE}_2(t, k) = 13/4$ (see Appendix A for details), which the simulated ACE is converging towards when modeling both responses as can be seen from the figure.

4.2. Employment trajectories

As the second example, we study a subset of the data from the Swiss Household Panel (Tillmann et al., 2016) that is available in the R package march(Maitre & Emery, 2020). The data consist of the employment trajectories of 845 individuals (421 women and 424 men). There is a vast literature on the determinants of employment. For example, Sianesi (2004) studied the short-term and long-term causal effects of labor market programs on employment using propensity score matching in the potential outcomes framework. The effects of previous employment experiences on the current employment status have been found to be significant for example by Heckman (1981) who discusses the importance of modeling the state dependency and heterogeneity of individuals using the labor force participation of women as an empirical application. Given our limited data, we do not aim to contribute to the literature of labor economics itself but merely illustrate how our

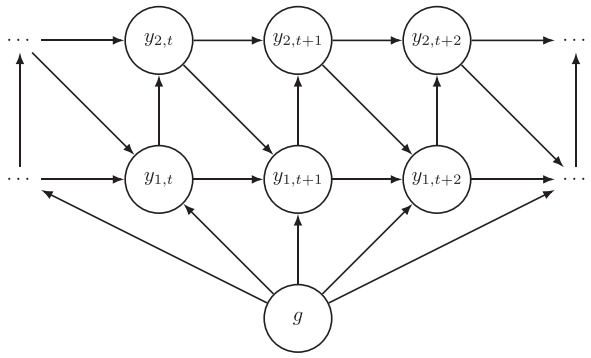


Fig. 5. A cross-section of the causal graph of the employment example at times t, $t+1$, and $t+2$, where $y_{1,t}$ is the employment status at time t, g is the gender, and $y_{2,t}$ is variable indicating whether the person has ever worked full-time up to and including time t.

proposed approach can be used to estimate long-term causal effects in a slightly more complex setting than in Section 4.1.

The employment status was observed every two years from the age of 20 until the age of 44 (i.e., at 13 time points), and it is coded as being either 1 (full-time employee) or 0 (other). We are interested in how an intervention on employment at the age of 30 affects the future employment status. More specifically, we compare two interventions: forcing everyone to work full time vs. forcing everyone to not work full time at the age of 30. This is an example of an atomic intervention that is carried out only at a single time point. After the intervention, the employment status is once again defined by the model equations at future time points. Our assumed causal graph is presented in Fig. 5. In addition to depending on the binary gender variable g (where $0 =$ woman and $1 =$ man, as coded in the original data) and the previous employment status $y_{1,t}$, we assume that the employment $y_{1,t+1}$ depends on the full work history via a binary variable $y_{2,t}$, which has the value 0 if the person has never worked full-time before age $t+1$ and 1 otherwise. In other words, the variable $y_{2,t}$ describes whether the person has ever worked full-time and it is defined by its previous value $y_{2,t-1}$ and the current employment status $y_{1,t}$. The graph in Fig. 5 is naturally a simplification of the true causal relationships related to employment, as we ignore the potential effects of education and differences between industries, for instance.

We model the probability of full-time employment as follows:

$$\begin{aligned} y_{1,t,i} \mid \eta_{t,i} &\sim \text{Bernoulli}\left(\text{logit}^{-1}(\eta_{t,i})\right) \\ \eta_{t,i} &= \begin{array}{l}(\beta_1 y_{2,t-1,i} + \delta_{t,1} + \delta_{t,2} y_{1,t-1,i}) g_i \\ +(\beta_2 y_{2,t-1,i} + \delta_{t,3} + \delta_{t,4} y_{1,t-1,i})(1-g_i)\end{array} \\ y_{2,t,i} &= (1-y_{2,t-1,i}) y_{1,t,i} + y_{2,t-1,i}, \end{aligned}$$

i.e., for both genders, we assume time-invariant effects for the work-history $y_{2,t}$, time-varying intercept, and time-varying effects of previous work status. We define the time-varying effects $\delta_{t,j}$, $j = 1, \ldots, 4$, as splines with $D = 6$ degrees of freedom. The resulting estimates were not sensitive to reasonable choices of D (see Appendix C).

While our main interest is in the ACE, given the graph in Fig. 5, we can interpret the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ as conditional causal effects on the logit scale. The mean of the time-invariant coefficients $\boldsymbol{\beta}$ of the work history were estimated as 0.07 for women, with a 95% posterior interval (-0.38, 0.52), and 0.58 (0.14, 1.01) for men, indicating that having at least some full-time work experience has a positive effect for men, but no effect for women. Fig. 6 shows the estimated posterior means and the 95% posterior intervals for the time-varying coefficients for men and women. The direction of change in these time-varying effects seems plausible: the coefficients corresponding to the gender-specific intercepts indicate decreases in the probabilities of transitioning to full-time employment if the person was not employed at the previous time point. For men, this probability starts to decrease after the age of 26, whereas for women the decrease starts already at the age of 22 (or earlier), but levels off around age 32. On the other hand, the probability of being employed conditional on being employed at the previous time point increases considerably over time for both men and women.

For estimating the ACE for ages 32, …, 44 given the intervention at the age of 30 (full-time employee vs. not), we simulate trajectories of $y_{1,t}$ and $y_{2,t}$. Fig. 7 shows the posterior means and the 95% posterior intervals for the ACE

$$P(y_{1,t} = 1 \,|\, \text{do}(y_{1,30} = 1)) - P(y_{1,t} = 1 \,|\, \text{do}(y_{1,30} = 0))$$

as well as the observed values of

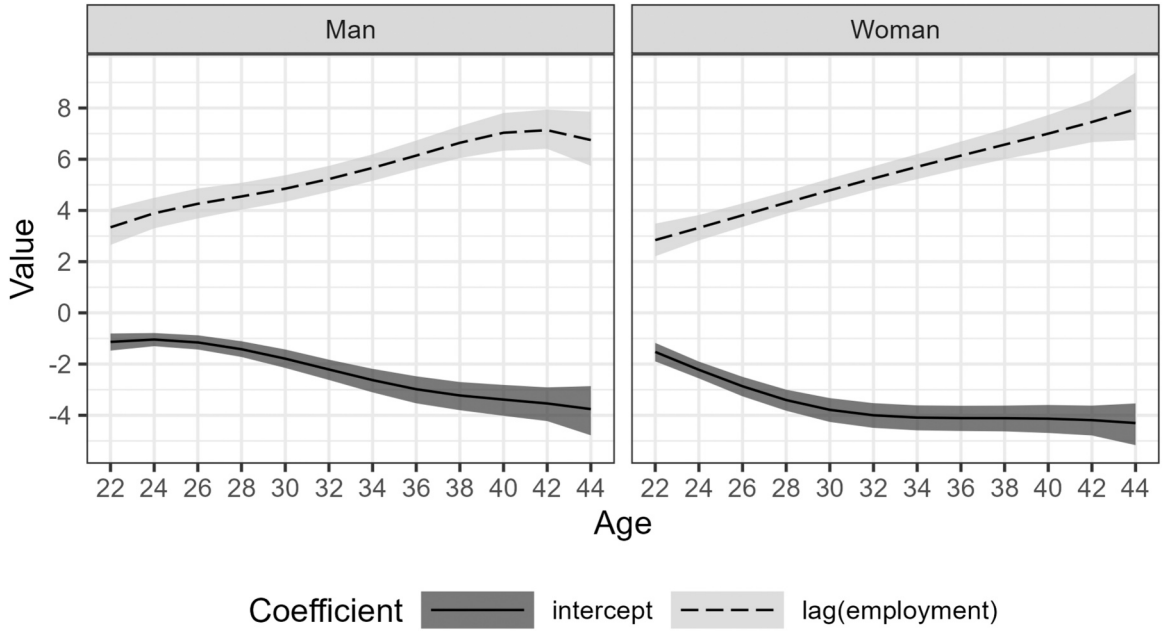


Fig. 6. Posterior means and 95% posterior intervals of the time-varying coefficients of the employment model for the intercept (solid line) and employment at the previous time point (dashed line). The left panel shows the coefficients for men and the right for women, respectively.

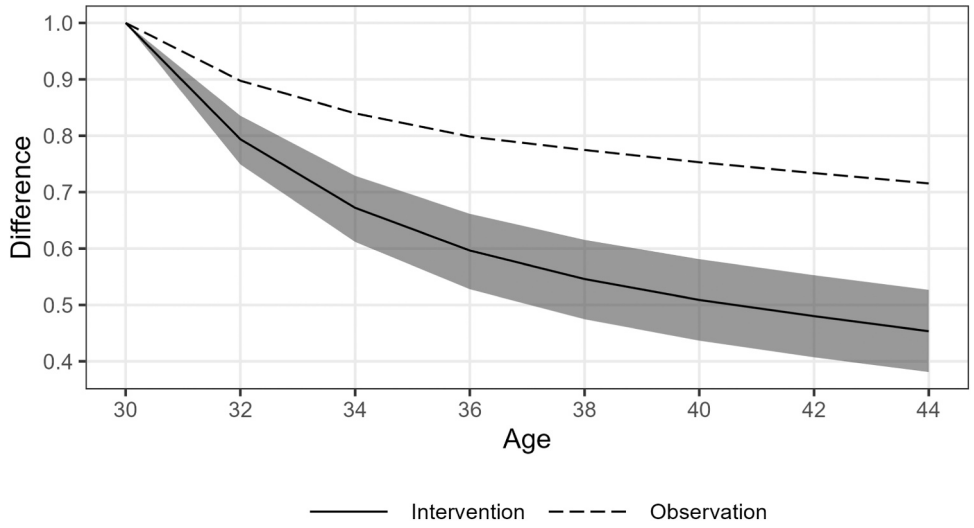


Fig. 7. Posterior means and 95% posterior intervals for $P(y_{1,t} = 1|\text{do}(y_{1,30} = 1)) - P(y_{1,t} = 1|\text{do}(y_{1,30} = 0))$ (solid line) and observed values of the difference of conditional probabilities $P(y_{1,t} = 1|y_{1,30} = 1) - P(y_{1,t} = 1|y_{1,30} = 0)$ (dashed line) for the employment model.

$$P(y_{1,t} = 1|y_{1,30} = 1) - P(y_{1,t} = 1|y_{1,30} = 0)$$

for $t = 32, \ldots, 44$, corresponding to the difference in full-time employment probabilities between those individuals who worked full-time at the age of 30 and those who did not. These observed *associations* differ considerably from the estimates of the ACE due to confounding and selection bias. Naturally, due to the simplified causal graph of this example, the ACE estimates are also probably subject to some bias as we do not adjust for all potentially important factors such as education.

4.3. Partnership Statuses

As the third illustration, we analyze partnership status sequence data; a subset of the German Family Panel Study (Brüderl et al., 2022), which is provided as supplementary material of (Raab & Struffolino, 2022, https://sa-book.github.io/). This data consists of information on the partnership statuses of 1834 individuals between the ages of 18 and 40 (22 time points, as we use the yearly granularity of the data). Our categorical response variable is the partnership status (single, living apart together, cohabiting, married). We also use data on the individuals' gender (coded as binary in the original data, with 0 = man and 1 = woman) and church attendance. Appendix D contains some descriptive summary statistics of the data. Church attendance is recorded as a time-invariant binary variable such that 0 corresponds to not having attended church and 1 corresponds to multiple church attendances per year. Naturally, church attendance as a proxy to religiousness is not time-invariant in reality, but here we assume that religiousness is a stable personality trait measured by our time-invariant church attendance (for discussion on religiousness and personality, see, e.g., Saucier, 2019 and Entringer et al., 2023). Here we illustrate how we could study the effect of religiousness on the transition probabilities between different partnership statuses. If we had more detailed yearly information on church attendance, we could model it along with the partnership status as a second response variable, with religiousness (affecting partnership status) defined by the previous five years of church attendance, for example. There are of course many other factors influencing the partnership trajectories (and religiousness) such as education and sosioeconomic background, which should be taken into account in proper studies of religiousness and partnership trajectories. There were 32 individuals who had missing church attendance information, which we removed from the analysis for simplicity. This is unlikely to affect the results due to the large size of the data sample.

We denote the partnership status by y_t and model the probability of being in a particular status s using the previous status, church attendance a, gender g, as well as with an individual-specific random intercept term. The individual-specific random effects capture the time-invariant heterogeneity between individuals, namely that individuals differ in their behavior of transitioning between different partnership statuses. The regression coefficients and random intercepts vary by category, and the random intercepts are modeled as correlated. Let $M = \{\text{LAT, COH, MAR}\}$. The model is

Table 1
Posterior means, standard deviations, and posterior intervals of the partnership model parameters for categories Living apart together (LAT), Cohabiting (COH), and Married (MAR). The coefficients of category Single (S) are fixed to zero for identifiability. Variables S, LAT, COH, and MAR refer to the coefficient of the category at the previous time point ($\beta_{k,s,1}$). Woman refers to the coefficient associated with women ($\beta_{s,3}$), while Church refers to the coefficient for individuals attending church multiple times a year ($\beta_{s,2}$).

Category	Variable	Parameter	Mean	SD	2.5%	97.5%
LAT	Woman	$\beta_{\text{LAT},3}$	0.25	0.05	0.15	0.35
COH	Woman	$\beta_{\text{COH},3}$	0.33	0.06	0.22	0.44
MAR	Woman	$\beta_{\text{MAR},3}$	0.29	0.07	0.16	0.42
LAT	Church	$\beta_{\text{LAT},2}$	0.18	0.05	0.08	0.29
COH	Church	$\beta_{\text{COH},2}$	− 0.11	0.06	− 0.22	0.01
MAR	Church	$\beta_{\text{MAR},2}$	0.67	0.07	0.53	0.80
LAT	S	$\beta_{\text{S,LAT},1}$	− 1.75	0.05	− 1.84	1.66
COH	S	$\beta_{\text{S,COH},1}$	− 2.90	0.06	− 3.02	2.78
MAR	S	$\beta_{\text{S,MAR},1}$	− 4.72	0.11	− 4.95	4.50
LAT	LAT	$\beta_{\text{LAT,LAT},1}$	1.53	0.06	1.42	1.64
COH	LAT	$\beta_{\text{LAT,COH},1}$	0.34	0.06	0.23	0.46
MAR	LAT	$\beta_{\text{LAT,MAR},1}$	− 1.33	0.09	− 1.50	1.16
LAT	COH	$\beta_{\text{COH,LAT},1}$	− 0.83	0.11	− 1.06	0.61
COH	COH	$\beta_{\text{COH,COH},1}$	2.99	0.08	2.84	3.14
MAR	COH	$\beta_{\text{COH,MAR},1}$	1.14	0.08	0.98	1.30
LAT	MAR	$\beta_{\text{MAR,LAT},1}$	− 0.29	0.14	− 0.57	0.01
COH	MAR	$\beta_{\text{MAR,COH},1}$	− 0.97	0.18	− 1.33	0.63
MAR	MAR	$\beta_{\text{MAR,MAR},1}$	4.48	0.10	4.28	4.68
LAT		$\sigma_{\nu,\text{LAT}}$	0.48	0.05	0.39	0.57
COH		$\sigma_{\nu,\text{COH}}$	0.33	0.07	0.18	0.45
MAR		$\sigma_{\nu,\text{MAR}}$	0.31	0.07	0.17	0.45
		$\rho_{\text{LAT,COH}}$	0.68	0.12	0.41	0.89
		$\rho_{\text{LAT,MAR}}$	0.83	0.11	0.56	0.98
		$\rho_{\text{COH,MAR}}$	0.34	0.24	− 0.19	0.76

Table 2
Posterior mean transition probabilities and their 95% posterior intervals for those who regularly attend church (top) and those who do not (bottom) for the partnership model.

			(a) Regularly attend church	
	S	LAT	COH	MAR
S	0.75 (0.73, 0.76)	0.19 (0.18, 0.21)	0.05 (0.04, 0.05)	0.02 (0.01, 0.02)
LAT	0.11 (0.10, 0.12)	0.66 (0.64, 0.68)	0.16 (0.15, 0.18)	0.06 (0.06, 0.07)
COH	0.03 (0.03, 0.04)	0.02 (0.02, 0.02)	0.70 (0.68, 0.73)	0.24 (0.22, 0.27)
MAR	0.01 (0.00, 0.01)	0.01 (0.00, 0.01)	0.00 (0.00, 0.00)	0.99 (0.99, 0.99)
			(b) Do not regularly attend church	
	S	LAT	COH	MAR
S	0.77 (0.76, 0.79)	0.17 (0.15, 0.18)	0.05 (0.05, 0.06)	0.01 (0.01, 0.01)
LAT	0.13 (0.12, 0.14)	0.63 (0.61, 0.65)	0.21 (0.19, 0.22)	0.04 (0.03, 0.04)
COH	0.04 (0.03, 0.04)	0.02 (0.02, 0.02)	0.81 (0.80, 0.83)	0.13 (0.12, 0.14)
MAR	0.01 (0.01, 0.01)	0.01 (0.01, 0.01)	0.00 (0.00, 0.01)	0.98 (0.97, 0.98)

$$
\begin{aligned}
y_{t,i}|\boldsymbol{\eta}_{t,i} &\sim \text{Categorical}\big(\text{softmax}(\boldsymbol{\eta}_{t,i})\big)\\
\eta_{\text{S},t,i} &= 0\\
\eta_{s,t,i} &= \sum_{k\in M}\beta_{k,s,1}I(y_{t-1,i}=k)+\beta_{s,2}a_i+\beta_{s,3}g_i+\nu_{s,i}, \quad s\in M\\
\nu_i|\boldsymbol{\sigma}_\nu,\mathbf{R} &\sim \text{N}\big(\mathbf{0},\boldsymbol{\sigma}_\nu^\top\mathbf{R}\boldsymbol{\sigma}_\nu\big)\\
\boldsymbol{\sigma}_\nu &= (\sigma_{\nu,\text{LAT}},\sigma_{\nu,\text{COH}},\sigma_{\nu,\text{MAR}})\\
\mathbf{R} &= \begin{bmatrix} 1 & \rho_{\text{LAT,COH}} & \rho_{\text{LAT,MAR}}\\ \rho_{\text{LAT,COH}} & 1 & \rho_{\text{COH,MAR}}\\ \rho_{\text{LAT,MAR}} & \rho_{\text{COH,MAR}} & 1 \end{bmatrix}
\end{aligned}
$$

where $\text{softmax}(\boldsymbol{x}) = \exp(\boldsymbol{x})/\sum_{k=1}^{K}\exp(x_k)$ for a vector $\boldsymbol{x} = (x_1, \ldots, x_K)$ and $I(\cdot)$ is an indicator function that has the value 1 if the argument is true and 0 otherwise. For each status $s \in M$, the parameters $\beta_{k,s,1}$ correspond to the effect of the status at the previous time point, $\beta_{s,2}$ is the effect of attendance, $\beta_{s,3}$ is the effect of gender, and $\nu_{s,i}$ is the individual-specific random effect. The standard deviation vector and correlation matrix of the random effects are $\boldsymbol{\sigma}_\nu$ and $\mathbf{R}$, respectively.

Table 1 shows the estimated regression coefficients for each category, as well as the standard deviations of the random effects (σ_ν) and their correlations ρ. Category Single (S) was set as the reference category for identifiability purposes, leading to a log odds interpretation of the coefficients. We see that church attendance increases the log odds of the Living apart together (LAT) status and especially the Married (MAR) status, and to some extent decreases the log odds of Cohabitation (COH). Women seem to have higher log odds for LAT, COH, and MAR. We could also similarly interpret the lagged effects of status, but instead of log odds interpretations, we study the corresponding transition probability matrices.

We estimate two marginal transition matrices as follows. For computing transitions from status s to each of the four statuses, we set the status of each individual to s and their Church variable to 0 (no attendance) and estimate their transition probabilities for the next time point. Averaging over these probabilities gives the expected transition probabilities for those who did not attend church. By repeating this for each starting status and again with the Church variable set to 1, we acquire the two transition matrices shown in Table 2.

From the transition probabilities, we see that those who are more religious (as measured by church attendance) tend to move faster to the Married status (higher probabilities in the MAR column) and they especially spend less time in the Cohabiting state (70% vs. 81%). There is also a slightly smaller chance of moving back to previous states, as can be seen from the smaller probabilities on the lower diagonals.

5. Discussion

We proposed the DMPM for modeling panel data applicable to both Gaussian and non-Gaussian data with potentially time-varying and individual-specific effects and showed how DMPMs can be used to estimate the full posterior distribution of long-term causal effects. This enables us to answer new causal questions related to panel data more realistically than before.

While DMPMs are highly general, they do have some limitations. The presented model does not support multiple levels of hierarchy in the random effects, for example, individuals nested in cities or countries. We did not consider explicit error terms in the linear predictor, which could be used to construct more complex correlation structures between different responses. We also did not consider individual-specific time-varying effects. We note that these limitations are mainly technical, but explicit error terms could also pose difficulties for identifiability and interpretation concerning causal inference. We leave these topics for possible directions for future research.

One of the strengths of latent variable approaches such as HMMs is the straightforward handling of missing response variables. As DMPMs operate on the observational level, dealing with missing data can be less trivial especially when missing variables are discrete, as they cannot be treated as unknown parameters in the gradient-based MCMC sampling (Neal, 2011), as is otherwise typical under the Bayesian paradigm. However, standard multiple imputation techniques (Van Buuren, 2018) combined with the Bayesian estimation of DMPMs is straightforward, although computationally intensive, as we can combine the posterior samples from multiple MCMC runs based on different imputed datasets (Gelman et al., 2013). Of course, a simple listwise deletion can also be a reasonable option under suitable missing data mechanisms (see, e.g., Van Buuren, 2018, sec. 2.7), especially when the missing data pattern is sparse.

While the fully Bayesian DMPM is compatible with the general Bayesian workflow (Gelman et al., 2020) and thus allows us to rely on the common Bayesian model evaluation and comparison methods such as posterior predictive checking and cross-validation, from a causal inference perspective we, following (Pearl, 1995), advocate the construction and critical assessment of the assumed causal graph as the first step. If this graph leads to identifiable causal effects that are compatible with the structure of the DMPM, then the tools provided by the R package dynamite can be used for defining, estimating, and obtaining predictions from DMPMs.

Funding

This work was funded by the Research Council of Finland (decision numbers 331817, 355153, and 345546).

CRediT authorship contribution statement

Jouni Helske: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Santtu Tikka:** Writing – review & editing,

Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declarations of interest

Declarations of interest: none.

Acknowledgments

This research was partially supported by the INVEST Research Flagship Centre.

Appendix A. ACE for the Bivariate Gaussian Model Example

We consider the model from Section 4.1:

$$\begin{aligned} y_{1,t} &= 0.6y_{1,t-1} + 0.4y_{2,t-1} + x_{t,i} + \epsilon_{1,t} \\ y_{2,t} &= -0.1y_{1,t-1} + 0.9y_{2,t-1} + 0.4x_t + \epsilon_{2,t} \\ \epsilon_{1,t} &\sim \mathrm{N}(0, 0.4^2) \\ \epsilon_{2,t} &\sim \mathrm{N}(0, 0.4^2). \end{aligned}$$

Our goal is to compute the following average causal effect when $k \to \infty$:

$$\mathrm{ACE}_2(t,k) = \mathrm{E}(y_{1,t+k} | \mathrm{do}(x_t = 1, \ldots, x_{t+k} = 1)) - \mathrm{E}(y_{1,t+k} | \mathrm{do}(x_t = 0, \ldots, x_{t+k} = 0)).$$

First, we solve the recurrence relation for $y_{1,t}$ under the intervention and take the expectation. We simplify the notation by dropping the do($\cdot$) terms and simply treat x_t as fixed. For the intervention do($x_t = 1$, …, $x_{t+k} = 1$), the model equations become

$$\begin{aligned} \mathrm{E}(y_{1,t}) &= 0.6\mathrm{E}(y_{1,t-1}) + 0.4\mathrm{E}(y_{2,t-1}) + 1 \\ \mathrm{E}(y_{2,t}) &= -0.1\mathrm{E}(y_{1,t-1}) + 0.9\mathrm{E}(y_{2,t-1}) + 0.4. \end{aligned}$$

To further simplify the notation, we denote $Y_t = \mathrm{E}(y_{1,t})$ and $Z_t = \mathrm{E}(y_{2,t})$. Thus we have

$$\begin{aligned} Y_t &= 0.6Y_{t-1} + 0.4Z_{t-1} + 1 \\ Z_t &= -0.1Y_{t-1} + 0.9Z_{t-1} + 0.4. \end{aligned}$$

Solving for Z_{t-1} in the first equation and incrementing the time index by one yields

$$2Z_t = 5Y_{t+1} - 3Y_t - 5.$$

Substituting this into the second equation for both Z_t and Z_{t-1} yields

$$\frac{5}{2}Y_{t+1} - \frac{3}{2}Y_t - \frac{5}{2} = -\frac{1}{10}Y_{t-1} + \frac{9}{10}\left(\frac{5}{2}Y_{t+1} - \frac{3}{2}Y_t - \frac{5}{2}\right) + \frac{2}{5}.$$

By solving for Y_{t+1} we obtain a second-order recurrence relation

$$Y_{t+1} = \frac{3}{2}Y_t - \frac{29}{50}Y_{t-1} + \frac{13}{50}.$$

This is a linear nonhomogeneous recurrence, and its steady-state value is

$$\frac{\frac{13}{50}}{1 - \frac{3}{2} + \frac{29}{50}} = \frac{13}{4}.$$

Thus we can convert the recurrence to a homogeneous form, whose characteristic polynomial $\lambda^2 - \frac{3}{2}\lambda + 29/50$ has roots that are all less than 1 in absolute value. Thus the recurrence converges to the steady state. For the intervention do($x_t = 0$, …, $x_{t+k} = 0$), it is easy to see that the expectation converges to 0. Thus we obtain the desired limit

$$\lim_{k\to\infty} \mathrm{ACE}_2(t,k) = \frac{13}{4} = 3.25.$$

This result can also be obtained by noting that the model is a stationary VAR.

B. Additional Details of the Bivariate Gaussian Model Example

In Section 4.1 we simulated two datasets from the model

$$\begin{aligned} y_{1,t,i} &= \alpha_1 + \beta_1 y_{1,t-1,i} + \beta_2 y_{2,t-1,i} + \beta_3 x_{t,i} + \epsilon_{1,t,i} \\ y_{2,t,i} &= \alpha_2 + \beta_4 y_{1,t-1,i} + \beta_5 y_{2,t-1,i} + \beta_6 x_{t,i} + \epsilon_{2,t,i} \\ \epsilon_{1,t,i} &\sim \mathrm{N}(0, \sigma_1^2) \\ \epsilon_{2,t,i} &\sim \mathrm{N}(0, \sigma_2^2) \end{aligned} \tag{6}$$

with the parameter values fixed to those in equation (5). Table 3 shows the estimated parameters in the first scenario and Table 4 in the second scenario (both data were generated using the same parameter values).

Table 3
Posterior means, standard deviations, posterior intervals, $\widehat{R}$ estimates, and effective sample sizes for the model estimated using the data in the first scenario of Section 4.1.

Parameter	Mean	SD	2.5%	97.5%	$\widehat{R}$	Bulk-ESS	Tail-ESS
α_1	0.00	0.00	0.00	0.00	1.00	6569	3210
α_2	0.00	0.00	− 0.01	0.00	1.00	8118	2912
β_1	0.60	0.00	0.60	0.61	1.00	5249	3160
β_2	0.39	0.00	0.39	0.40	1.00	5045	3347
β_3	0.99	0.02	0.95	1.02	1.00	3647	3217
β_4	− 0.10	0.00	− 0.10	− 0.09	1.00	5590	3188
β_5	0.90	0.00	0.89	0.90	1.00	5898	3281
β_6	0.39	0.02	0.35	0.42	1.00	3328	2441
σ_1	0.40	0.00	0.40	0.40	1.00	6932	3060
σ_2	0.40	0.00	0.40	0.40	1.00	6508	3153

Table 4
Posterior means, standard deviations, posterior intervals, $\widehat{R}$ estimates, and effective sample sizes for the model estimated using the data in the second scenario of Section 4.1.

Parameter	Mean	SD	2.5%	97.5%	$\widehat{R}$	Bulk-ESS	Tail-ESS
α_1	0.00	0.00	0.00	0.00	1.00	5342	2686
α_2	0.00	0.00	− 0.01	0.00	1.00	5939	3294
β_1	0.60	0.00	0.60	0.61	1.00	3361	3185
β_2	0.39	0.00	0.39	0.40	1.00	4777	3055
β_3	1.00	0.01	0.98	1.01	1.00	3027	2973
β_4	− 0.10	0.00	− 0.10	− 0.09	1.00	3482	2716
β_5	0.90	0.00	0.89	0.90	1.00	4555	2563
β_6	0.40	0.01	0.38	0.41	1.00	3474	2582
σ_1	0.40	0.00	0.40	0.40	1.00	5726	2956
σ_2	0.40	0.00	0.40	0.40	1.00	5929	2849

As expected, the posterior means are close to the true values specified in equation (5) that were used to generate the data. The models converge well as indicated by the convergence diagnostics: $\widehat{R}$ statistics are less than 1.01 and the effective sample sizes are sufficient (see Vehtari et al., 2021, for details on these diagnostics).

Computation time with four parallel chains, each with 2000 iterations (of which the first 1000 were discarded as warm-up) was about 13 and 18 s in the first and the second scenario, respectively, on a Windows laptop with a six-core Intel i3–1215 U processor and 16 GB of RAM.

C. Additional Details of the Employment Example

Table 5 shows the contingency table of 424 men and 421 women of the data of Section 4.2 in terms of whether they were ever full-time employed between the ages of 20 and 44. Table 6 shows the frequencies of the ages of first full-time employment for those who were employed full-time at some point during the study period.

Table 5
Contingency table of gender and full-time employment for the data of Section 4.2.

	Never worked full-time	Worked full-time
Woman	102	319
Man	35	389

Table 6
Frequency table of age of first full-time employment for the data of Section 4.2.

Age	20	22	24	26	28	30	32	34	36	38	40	42
Frequency	509	72	40	35	24	6	7	4	3	4	2	2

Table 7 shows the estimated time-invariant parameters and the corresponding convergence diagnostics of the employment model of Section 4.2. Parameters τ_k, $k = 1, \ldots, 4$ are the standard deviation parameters of the random-walk prior for the spline coefficients of the time-varying effect δ_k. Computation time with four parallel chains, each with 5000 iterations (of which the first 2500 were discarded as warm-up) was about 5 min on a Windows laptop with a six-core Intel i3–1215 U processor and 16 GB of RAM.

Table 7
Posterior means, standard deviations, posterior intervals, $\widehat{R}$ estimates, and effective sample sizes for the employment model of Section 4.2.

Parameter	Mean	SD	2.5%	97.5%	$\widehat{R}$	Bulk-ESS	Tail-ESS
β_1	0.58	0.22	0.13	1.01	1.00	12464	8182
β_2	0.07	0.23	− 0.38	0.52	1.00	11622	7958
τ_1	1.13	0.39	0.56	2.06	1.00	6954	7094
τ_2	1.32	0.41	0.68	2.26	1.00	7739	7218
τ_3	1.10	0.37	0.56	2.00	1.00	6898	7183
τ_4	1.31	0.38	0.73	2.22	1.00	8087	5691

We also tested how the choice of D, the degrees of freedom of the spline coefficients, affects the results. We estimated the same model as in the main text with $D = 4, 6, 8, 15$. While this choice has a small effect on the smoothness of the time-varying parameters, it had no meaningful effect on the causal effect estimates (see the figures in GitHub). We also compared the predictive accuracy of these models using the approximate leave-one-out cross-validation. No significant differences between the models were found, i.e., the differences in the estimated log predictive densities were less than the corresponding standard errors (see the GitHub codes for further details).

D. Additional Details of the Partnership Example

Table 8 shows the contingency table of 808 men and 1026 women of the data of Section 4.3 with respect to their church attendance.

Table 8
Contingency table of gender and church attendance for the data of Section 4.3.

	Does not attend	Attends
Man	528	280
Woman	591	435

The observed average transition probabilities conditional on the church attendance are shown in Table 9. The implied stationary distributions based on the observed and estimated transition probabilities in Table 2 are shown in the GitHub codes.

Table 9
Observed mean transition probabilities for those who regularly attend church (top) and those who do not (bottom) for the partnership data.

(a) Regularly attend church	S	LAT	COH	MAR
S	0.78	0.17	0.03	0.01
LAT	0.10	0.67	0.16	0.07
COH	0.03	0.02	0.74	0.21
MAR	0.00	0.00	0.00	0.99
(b) Do not regularly attend church	S	LAT	COH	MAR
S	0.80	0.14	0.05	0.01
LAT	0.12	0.66	0.19	0.03
COH	0.03	0.02	0.82	0.13
MAR	0.01	0.01	0.01	0.98

Table 10 shows the parameter estimates of the model of Section 4.3 together with the corresponding convergence diagnostics. Computation time with four parallel chains, each with 6000 iterations (of which the first 1000 were discarded as warm-up) was about 2 h and 50 min on a Windows laptop with a six-core Intel i3–1215 U processor and 16 GB of RAM.

Table 10
Posterior means, standard deviations, posterior intervals, $\widehat{R}$ estimates, and effective sample sizes for the partnership model of Section 4.3.

Category	Variable	Parameter	Mean	SD	2.5%	97.5%	$\widehat{R}$	Bulk-ESS	Tail-ESS
LAT	Woman	$\beta_{\text{LAT},3}$	0.25	0.05	0.15	0.35	1.00	14783	15520
COH	Woman	$\beta_{\text{COH},3}$	0.33	0.06	0.22	0.44	1.00	14576	15497
MAR	Woman	$\beta_{\text{MAR},3}$	0.29	0.07	0.16	0.42	1.00	16154	15395
LAT	Church	$\beta_{\text{LAT},2}$	0.18	0.05	0.08	0.29	1.00	16214	16764
COH	Church	$\beta_{\text{COH},2}$	− 0.11	0.06	− 0.22	0.01	1.00	18142	16425
MAR	Church	$\beta_{\text{MAR},2}$	0.67	0.07	0.53	0.80	1.00	16402	16307
LAT	S	$\beta_{\text{S,LAT},1}$	− 1.75	0.05	− 1.84	1.66	1.00	9884	12601
COH	S	$\beta_{\text{S,COH},1}$	− 2.90	0.06	− 3.02	2.78	1.00	13281	14252
MAR	S	$\beta_{\text{S,MAR},1}$	− 4.72	0.11	− 4.95	4.50	1.00	25252	14883
LAT	LAT	$\beta_{\text{LAT,LAT},1}$	1.53	0.06	1.42	1.64	1.00	10594	13534
COH	LAT	$\beta_{\text{LAT,COH},1}$	0.34	0.06	0.23	0.46	1.00	15584	14865
MAR	LAT	$\beta_{\text{LAT,MAR},1}$	− 1.33	0.09	− 1.50	1.16	1.00	14354	15981
LAT	COH	$\beta_{\text{COH,LAT},1}$	− 0.83	0.11	− 1.06	0.61	1.00	17607	15773
COH	COH	$\beta_{\text{COH,COH},1}$	2.99	0.08	2.84	3.14	1.00	9895	13480
MAR	COH	$\beta_{\text{COH,MAR},1}$	1.14	0.08	0.98	1.30	1.00	14761	13806
LAT	MAR	$\beta_{\text{MAR,LAT},1}$	− 0.29	0.14	− 0.57	0.01	1.00	20853	16412
COH	MAR	$\beta_{\text{MAR,COH},1}$	− 0.97	0.18	− 1.33	0.63	1.00	23131	16357
MAR	MAR	$\beta_{\text{MAR,MAR},1}$	4.48	0.10	4.28	4.68	1.00	16861	15874
LAT		$\sigma_{\nu,\text{LAT}}$	0.48	0.05	0.39	0.57	1.00	4619	8676
COH		$\sigma_{\nu,\text{COH}}$	0.33	0.07	0.18	0.45	1.00	2268	3146
MAR		$\sigma_{\nu,\text{MAR}}$	0.31	0.07	0.17	0.45	1.00	3956	5330
		$\rho_{\text{LAT,COH}}$	0.68	0.12	0.41	0.89	1.00	4312	5890
		$\rho_{\text{LAT,MAR}}$	0.83	0.11	0.56	0.98	1.00	5583	8641
		$\rho_{\text{COH,MAR}}$	0.34	0.24	− 0.19	0.76	1.00	4201	5538

We also tested a model with additional interaction term between church attendance and gender. These terms did not differ from zero and based on the approximate leave-one-out cross-validation this interaction model had somewhat worse predictive performance (see the GitHub codes for further details).

References

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum likelihood for cross-lagged panel models with fixed effects. *Socius*, 3. https://doi.org/10.1177/2378023117710578

Altman, R. M. (2007). Mixed hidden Markov models. *Journal of the American Statistical Association, 102*(477), 201–210. https://doi.org/10.1198/016214506000001086

Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72*(3), 269–342. https://doi.org/10.1111/j.1467-9868.2009.00736.x

Asparouhov, T., Hamaker, E. L., & Muthén, B. (2018). Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 359–388. https://doi.org/10.1080/10705511.2017.1406803

Asparouhov, T., & Muthén, B. (2021). Expanding the Bayesian structural equation, multilevel and mixture models to logit, negative-binomial, and nominal variables. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 622–637. https://doi.org/10.1080/10705511.2021.1878896

Bai, J., & Wang, P. (2015). Identification and Bayesian estimation of dynamic factor models. *Journal of Business & Economic Statistics, 33*(2), 221–240. https://doi.org/10.1080/07350015.2014.941467

Barban, N., Luna, X., de Lundholm, E., Svensson, I., & Billari, F. C. (2020). Causal effects of the timing of life-course events: Age at retirement and subsequent health. *Sociological Methods & Research, 49*(1), 216–249. https://doi.org/10.1177/0049124117729697

Bartolucci, F., Farcomeni, A., & Pennoni, F. (2013). Latent markov models for longitudinal data. *Chapman; Hall/CRC*. https://doi.org/10.1201/b13246

Bartolucci, F., Pennoni, F., & Francis, B. (2007). A latent Markov model for detecting patterns of criminal activity. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*(1), 115–132. https://doi.org/10.1111/j.1467-985X.2006.00440.x

Bartolucci, F., Pennoni, F., & Vittadini, G. (2016). Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. *Journal of Educational and Behavioral Statistics, 41*(2), 146–179. https://doi.org/10.3102/1076998615622234

Bartolucci, F., Pennoni, F., & Vittadini, G. (2023). A causal latent transition model with multivariate outcomes and unobserved heterogeneity: Application to human capital development. *Journal of Educational and Behavioral Statistics, 48*(4), 387–419. https://doi.org/10.3102/10769986221150033

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics, 37*(6), 1554–1563. https://doi.org/10.1214/aoms/1177699147

Bell, A., Fairbrother, M., & Jones, K. (2019). Fixed and random effects models: Making an informed choice. *Quality & Quantity, 53*, 1051–1074. https://doi.org/10.1007/s11135-018-0802-x

Bell, A., Jones, K., & Fairbrother, M. (2018). Understanding and misunderstanding group mean centering: A commentary on Kelley et al.'s dangerous practice. *Quality & Quantity, 52*, 2031–2036. https://doi.org/10.1007/s11135-017-0593-5

Berchtold, A. (1999). The double chain Markov model. *Communications in Statistics - Theory and Methods, 28*(11), 2569–2589. https://doi.org/10.1080/03610929908832439

Bijlsma, M. J., & Wilson, B. (2019). Modelling the socio-economic determinants of fertility: A mediation analysis using the parametric g-formula. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 183*(2), 493–513. https://doi.org/10.1111/rssa.12520

Blalock, H. M. (1964). *Causal inferences in nonexperimental research*. University of North Carolina Press,.

Blossfeld, H.-P., Rossbach, H.-G., & Von Maurice, J. (2011). *Education as a lifelong process: The German national educational panel study (NEPS)*. Springer,.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons,. https://doi.org/10.1002/9781118619179

Bollen, K. A., & Brand, J. E. (2010). A general panel model with random and fixed effects: A structural equations approach. *Social Forces, 89*(1), 1–34. https://doi.org/10.1353/sof.2010.0072

Brüderl, J., Drobnič, S., Hank, K., Neyer, F. J., Walper, S., Alt, P., Borschel, E., Bozoyan, C., Garrett, M., Geissler, S., Gonzalez Avilés, T., Gröpler, N., Hajek, K., Herzig, M., Lenke, R., Lorenz, R., Lutz, K., Peter, T., Preetz, R., … Wetzel, M.(2022)The German family panel (pairfam).10.4232/pairfam.5678.13.0.0.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., Gabry, J., & Vehtari, A. (2020). Approximate leave-future-out cross-validation for Bayesian time series models. *Journal of Statistical Computation and Simulation, 90*(14), 2499–2523. https://doi.org/10.1080/00949655.2020.1783262

Van Buuren, S. (2018). *Flexible imputation of missing data* (second edition.,). Chapman; Hall/CRC,. https://doi.org/10.1201/9780429492259

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chatfield, C. (2000). *Time-series forecasting*. Chapman; Hall/CRC,.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (2nd ed..,). Oxford University Press,.

Eilers, P. H. C., & Marx, B. D. (1996). Flexible smoothing with *B*-splines and penalties. *Statistical Science, 11*(2), 89–102. https://doi.org/10.1214/ss/1038425655

Entringer, T. M., Gebauer, J. E., & Kroeger, H. (2023). Big five personality and religiosity: Bidirectional cross-lagged effects and their moderation by culture. *Journal of Personality, 91*(3), 736–752. https://doi.org/10.1111/jopy.12770

Eubank, R. L., Huang, C., Maldonado, Y. M., Wang, N., Wang, S., & Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66*(3), 653–667. https://doi.org/10.1111/j.1467-9868.2004.B5595.x

Finney, S. J., & DiStefano, C. (2013). Non-normal and categorical data in structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed..,, pp. 439–492). Information Age Publishing.
Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer,.
Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed..,). Chapman; Hall/CRC,.
Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., and Modràk, M. (2020). Bayesian workflow.⟨https://arxiv.org/abs/2011.01808⟩.
Geweke, J. (1977). The dynamic factor analysis of economic time series. In D. J. Aigner, & A. S. Goldberger (Eds.), *Latent variables in socio-economic models*. North-Holland.
Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods, 20*(1), 102. https://doi.org/10.1037/a0038889
Harvey, A. C. (1978). The estimation of time-varying parameters from panel data. *Annales Dèlelőtt lainséé Xxxxxxxx, 30/31*, 203–226. https://doi.org/10.2307/20075291
Harvey, A. C., & Phillips, G. D. A. (1982). The estimation of regression models with time-varying parameters. In M. Deistler, E. Fürst, & G. Schwödiauer (Eds.), *Games, economic dynamics, and time series analysis* (pp. 306–321). Physica-Verlag HD. https://doi.org/10.1007/978-3-662-41533-7_18.
Haslbeck, J. M. B., Bringmann, L. F., & Waldorp, L. J. (2021). A tutorial on estimating time-varying vector autoregressive models. *Multivariate Behavioral Research, 56*(1), 120–149. https://doi.org/10.1080/00273171.2020.1743630
Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 55*(4), 757–796. https://doi.org/10.1111/j.2517-6161.1993.tb01939.x
Heckman, J. J. (1981). Heterogeneity and state dependence. In S. Rosen (Ed.), *Studies in labor markets* (pp. 91–140). University of Chicago Press.
Helske, J. (2015). Prediction and interpolation of time series by state space models. PhD Dissertation, University of Jyväskylä. Department of Mathematics and Statistics.⟨https://jyx.jyu.fi/handle/123456789/49043⟩.
Helske, J. (2022). Efficient Bayesian generalized linear models with time-varying coefficients: The walker package in R. *SoftwareX, 18*, Article 101016. https://doi.org/10.1016/j.softx.2022.101016
Helske, J., & Vihola, M. (2021). bssm: Bayesian inference of non-linear and non-Gaussian state space models in R. *The R Journal, 13*(2), 578–589. https://doi.org/10.32614/RJ-2021-103
Helske, S., Helske, J., & Eerola, M. (2018). Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data. In G. Ritschard, & M. Studer (Eds.), *Sequence analysis and related approaches: Innovative methods and applications* (pp. 185–200). Springer International Publishing. https://doi.org/10.1007/978-3-319-95420-2_11.
Helske, S., Keski-Säntti, M., Kivelä, J., Juutinen, A., Kääriälä, A., Gissler, M., Merikukka, M., & Lallukka, T. (2023). Predicting the stability of early employment with its timing and childhood social and health-related predictors: A mixture Markov model approach. *Longitudinal and Life Course Studies, 14*(1), 73–104. https://doi.org/10.1332/175795921×16609201864155
Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Boca Raton: Chapman & Hall/CRC,.
Holtz-Eakin, D., Newey, W., & Rosen, H. S. (1988). Estimating vector autoregressions with panel data. *Econometrica, 56*(6), 1371–1395. https://doi.org/10.2307/1913103
Hudde, A., & Jacob, M. (2023). Thereas more in the data! Using month-specific information to estimate changes before and after major life events. *Sociological Science, 10*, 830–856. https://doi.org/10.15195/v10.a29
Imbens, G. W. (2020). Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics. *Journal of Economic Literature, 58*(4), 1129–1179. https://doi.org/10.1257/jel.20191597
Imbens, G. W. (2024). Causal inference in the social sciences. *Annual Review of Statistics and Its Application*. https://doi.org/10.1146/annurev-statistics-033121-114601
Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica, 62*(2), 467–475. ⟨http://www.jstor.org/stable/2951620⟩.
Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering, Series D, 82*, 35–45. https://doi.org/10.1115/1.3662552
Kivva, Y., Mokhtarian, E., Etesami, J., & Kiyavash, N. (2022). Revisiting the general identifiability problem. Vol. 180. In J. Cussens, & K. Zhang (Eds.), *Proceedings of the 38th conference on uncertainty in artificial intelligence* (pp. 1022–1030). PMLR. Vol. 180 ⟨https://proceedings.mlr.press/v180/kivva22a.html⟩. Vol. 180.
Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford Press,.
Kumor, D., Cinelli, C., & Bareinboim, E. (2020). Efficient identification in linear structural causal models with auxiliary cutsets. Vol. 119. In H. D. III, & A. Singh (Eds.), *Proceedings of the 37th international conference on machine learning* (pp. 5501–5510). PMLR. Vol. 119 ⟨https://proceedings.mlr.press/v119/kumor20a.html⟩. Vol. 119.
Lang, S., & Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics, 13*(1), 183–212. https://doi.org/10.1198/1061860043010
Lauritzen, S. L. (1996). *Graphical models*. Clarendon Press.
Lee, S., Correa, J. D., & Bareinboim, E. (2020a). General identifiability with arbitrary surrogate experiments. Vol. 115. In R. P. Adams, & V. Gogate (Eds.), *Proceedings of the 35th uncertainty in artificial intelligence conference* (pp. 389–398). PMLR. Vol. 115 ⟨https://proceedings.mlr.press/v115/lee20b.html⟩. Vol. 115.
Lee, S., Correa, J., & Bareinboim, E. (2020b). General transportability — synthesizing observations and experiments from heterogeneous domains. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(06), 10210–10217. https://doi.org/10.1609/aaai.v34i06.6582
Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis, 100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008
Li, F., Ding, P., & Mealli, F. (2023). Bayesian causal inference: A critical review. *Philosophical Transactions of the Royal Society A, 381*. https://doi.org/10.1098/rsta.2022.0153
Liao, T. F., Bolano, D., Brzinsky-Fay, C., Cornwell, B., Fasang, A. E., Helske, S., Piccarreta, R., Raab, M., Ritschard, G., Struffolino, E., & Studer, M. (2022). Sequence analysis: Its past, present, and future. *Social Science Research, 107*, Article 102772. https://doi.org/10.1016/j.ssresearch.2022.102772
MacDonald, I. L., & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series*. Vol. 110. CRC Press,.
Morris, J. S. (2002). The BLUPs are not "best" when it comes to bootstrapping. *Statistics & Probability Letters, 56*(4), 425–430. https://doi.org/10.1016/S0167-7152(02)00041-X
Mulder, J. D., & Hamaker, E. L. (2021). Three extensions of the random intercept cross-lagged panel model. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(4), 638–648. https://doi.org/10.1080/10705511.2020.1784738
Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica, 46*(1), 69–85. https://doi.org/10.2307/1907490
Murphy, S. A. (2003). Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 65*(2), 331–355.
Murray, E. J., & Kunicki, Z. (2022). *As the wheel turns: Causal inference for feedback loops and bidirectional effects*. OSF Preprints,. https://doi.org/10.31219/osf.io/9em5q
Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132.
Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov chain Monte Carlo*. Chapman; Hall/CRC. https://doi.org/10.1201/b10905.
Nisén, J., Bijlsma, M. J., Martikainen, P., Wilson, B., & Myrskylä, M. (2022). The gendered impacts of delayed parenthood: A dynamic analysis of young adulthood. *Advances in Life Course Research, 53*, Article 100496. https://doi.org/10.1016/j.alcr.2022.100496
Oganisian, A., & Roy, J. A. (2021). A practical introduction to Bayesian estimation of causal effects: Parametric and nonparametric approaches. *Statistics in Medicine, 40*(2), 518–551. https://doi.org/10.1002/sim.8761
Pakpahan, E., Hoffmann, R., & Kröger, H. (2017P). Statistical methods for causal analysis in life course research: An illustration of a cross-lagged structural equation model, a latent growth model, and an autoregressive latent trajectories model. *International Journal of Social Research Methodology, 20*(1), 1–19. https://doi.org/10.1080/13645579.2015.1091641
Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika, 82*(4), 669–688. https://doi.org/10.1093/biomet/82.4.669
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed..,). Cambridge University Press.
Maitre, O., Emery, K. (2020). march: Markov chains. R package version 3.3.2.⟨https://CRAN.R-project.org/package=march⟩.
Pennoni, F., and Piccarreta, R. (2017). Dynamic sequential analysis of careers.CLADAG 2017 11th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society, 1–6.
Piccarreta, R., & Studer, M. (2019). Holistic analysis of the life course: Methodological challenges and new perspectives. *Advances in Life Course Research, 41*, Article 100251. https://doi.org/10.1016/j.alcr.2018.10.004
R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing,. ⟨https://www.R-project.org/⟩.
Raab, M., & Struffolino, E. (2022). *Sequence analysis*. SAGE.
Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41
Saucier, G. (2019). Is religiousness a form of variation in personality, or in culture, or neither? Conceptual issues and empirical indications. *Archive for the Psychology of Religion, 41*(3), 216–223. https://doi.org/10.1177/0084672419894677
Scott, M. A., LeGoff, J. M., & Gauthier, J. A. (2024). History matters: The statistical modelling of the life course. *Quality & Quantity, 58*, 445–469. https://doi.org/10.1007/s11135-023-01648-1
Shamsollahi, A., Zyphur, M. J., & Ozkok, O. (2022). Long-run effects in dynamic systems: New tools for cross-lagged panel models. *Organizational Research Methods, 25*(3), 435–458. https://doi.org/10.1177/1094428121993228
Shpitser, I., and Pearl, J. (2006a). Identification of conditional interventional distributions.Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence, 437–444.
Shpitser, I., & Pearl, J. (2006b). Identification of joint interventional distributions in recursive semi-Markovian causal models. *Proceedings of the 21st National Conference on Artificial Intelligence, 2*, 1219–1226.
Sianesi, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *The Review of Economics and Statistics, 86*(1), 133–155. https://doi.org/10.1162/003465304323023723
Sims, C. A. (1980). Macroeconomics and reality. *Econometrica, 48*(1), 1–48. https://doi.org/10.2307/1912017
Spiegelhalter, D. J. (1998). Bayesian graphical modelling: A case-study in monitoring health outcomes. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 47*(1), 115–133. https://doi.org/10.1111/1467-9876.00101
Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (2nd ed..,). MIT Press,. https://doi.org/10.7551/mitpress/1754.001.0001
Stan Development Team. (2022). The Stan C++ library.⟨https://mc-stan.org/⟩.

Sun, Y., Carroll, R. J., & Li, D. (2009). Semiparametric estimation of fixed-effects panel data varying coefficient models. *Nonparametric econometric methods*. Emerald Group Publishing Limited,.

Thombs, R. P. (2022). A guide to analyzing large *N*, large *T* panel data. *Socius, 8*, Article 23780231221117645. https://doi.org/10.1177/23780231221117645

Tikka, S., Helske, J., & Karvanen, J. (2023). Clustering and structural robustness in causal diagrams. *Journal of Machine Learning Research, 24*(195), 1–32. ⟨http://jmlr.org/papers/v24/21-1322.html⟩.

Tikka, S., Hyttinen, A., & Karvanen, J. (2021). Causal effect identification from multiple incomplete data sources: A general search-based approach. *Journal of Statistical Software, 99*(5), 1–40. https://doi.org/10.18637/jss.v099.i05

Tillmann, R., Voorpostel, M., Kuhn, U., Lebert, F., Ryser, V.-A., Lipps, O., Wernli, B., & Antal, E. (2016). The Swiss household panel study: Observing social change since 1999. *Longitudinal and Life Course Studies, 7*(1), 64–78.

Triantafyllopoulos, K. (2021). Bayesian inference of state space models: Kalman filtering and beyond. *Springer Texts in Statistics*. Springer International Publishing,. https://doi.org/10.1007/978-3-030-76124-0

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press,. https://doi.org/10.1017/CBO9780511802256

Vehtari, A., Gelman, A., & Gabry, J. (2016). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC (with discussion). *Bayesian Analysis, 16*(2), 667–718. https://doi.org/10.1214/20-BA1221

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, 373–385. https://doi.org/10.1037/a0014851

Vihola, M., Helske, J., & Franks, J. (2020). Importance sampling type estimators based on approximate marginal Markov chain Monte Carlo. *Scandinavian Journal of Statistics, 47*(4), 1339–1376. https://doi.org/10.1111/sjos.12492

Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology, 177*(4), 292–298.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag,. ⟨https://ggplot2.tidyverse.org⟩.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). dplyr: A grammar of data manipulation.R package version 1.1.4.

Wood, S. N. (2020). Inference and computation with generalized additive models and their extensions. *TEST, 29*(2), 307–339. https://doi.org/10.1007/s11749-020-00711-5

Tikka, S., & Helske, J. (2023). dynamite: An R Package for Dynamic Multivariate Panel Models. *arXiv preprint arXiv:2302.01607*.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., Hamaker, E. L., Shamsollahi, A., Pierides, D. C., Koval, P., & Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods, 23*(4), 651–687. https://doi.org/10.1177/1094428119847278