Research Report

Pepijn Vink (6100252)[1]

ABSTRACT
Lorem Ipsum…

1. Introduction

1.1. *Causal Models and Backdoor Paths*

Include: Introduction to types of DAG structures, backdoor paths, and terminology

1.2. *Confounding in Cross-Lagged Panel Models*

Include: Introduce an example SCM (and possibly the CLPM)

When taking the effect of x_4 on x_5 as an example, confounding of this effect happens through several paths. First, there are paths through the x_t variables and y_t variables. An example is the $x_4 \leftarrow y_3 \rightarrow y_4 \rightarrow y_5$ path, but also $x_4 \leftarrow x_3 \rightarrow y_4 \rightarrow y_5$ and other paths that go back more. Secondly, there is confounding due to the time-specific confounders U_{it}. The $x_4 \rightarrow y_5$ effect would then be confounded by the $x_4 \leftarrow U_4 \rightarrow y_4 \rightarrow y_5$ path (and no other paths). Lastly, there is confounding through the time-invariant confounders C. Confounding on the $x_4 \rightarrow y_5$ path would, for example be a $x_4 \leftarrow C_1 \rightarrow y_5$ path, but also a $x_4 \leftarrow x_3 \leftarrow C_1 \rightarrow y_5$ and a $x_4 \leftarrow x_3 \leftarrow C_1 \rightarrow y_4 \rightarrow y_5$ path as well as many more. These types of confounding would result in biased estimates if the coefficient of regressing y_5 on only x_4 were used as an estimate.

The first type of confounding, confounding through previous instances of x and y, can be resolved by controlling for these previous instances in the regression equation. This is

what the classic cross-lagged panel model does. This also resolves some of the confounding by C that runs through previous timepoints of x and y, so that the $x_4 \rightarrow y_5$ path is only confounded by $x_4 \leftarrow C \rightarrow y_5$ paths. The second type of confounding, through time-specific confounders U_t, is resolved by allowing residuals of x_t and y_t to be correlated. The third type of confounding, that through time-invariant confounders C, can sometimes be resolved through the inclusion of a latent factor, which is the approach that the random intercept cross-lagged panel model (RI-CLPM) and the dynamic panel model (DPM) take.

1.3. *Comparison of RI-CLPM and DPM*

i.e. residual based and observed based. conceptual interpretations

1.4. *RI-CLPM and DPM as Data Generating Mechanisms*

Although the RI-CLPM and the DPM have some similarities when it comes to analyses, when interpreting them as causal models each of them has different implications regarding the data generating mechanism.

The RI-CLPM implies that a person's deviation from their mean at $t-1$ has a causal effect on their deviation from their mean at timepoint t. It implies that the trait-like factor for a time varying variable x is only related to the vector of a time varying variable y through the covariance with its random intercept. This can reflect, for example, between person differences in the tendency to give high responses. It may thus represent a type of measurement error that is systematic within a person, but random between persons.

The DPM implies that observed scores at timepoint $t-1$ have a direct causal effect on observed scores at timepoint t. Furthermore, in the DPM, the latent factor for x is allowed to have an effect on observed scores for y through the observed scores on x (and in part through observed scores in earlier lags of y), rather than only through its covariance with the latent factor for y.

1.5. *Confounding in the RI-CLPM and DPM*

It is known that both the RI-CLPM and the DPM, in specific cases, control for unobserved time-invariant confounders. Usami et al. (2019), for example, show that when the effects of the confounders are stable over time, as well as the lagged effects, the confounder does not affect the estimated of the lagged effects when using the RI-CLPM. The latent factor in the DPM, however, also controls for time-invariant confounders with time-invariant effects when lagged effects are not constant over time (Murayama & Gfrörer, 2022). Furthermore, estimates will still be biased when the effects of the confounders are unstable over time.

Therefore, the simplest forms of both the RI-CLPM and the DPM will not yield unbiased estimates of the causal effects if there are unobserved time-invariant confounders with time-varying effects. The thesis will explore ways to address these types of confounding in the RI-CLPM using causal inference techniques such as propensity score

weighting (Brown et al., 2021; e.g. Vansteelandt & Daniel, 2014). However, before this question is answered, the extent to which unobserved time-invariant confounders affect the estimates of the RI-CLPM and the DPM should be evaluated. Thus, this research report will explore different types of time-invariant confounders in panel data (i.e. with time-stable effects versus time-varying effects) and assess their effects on estimates of the RI-CLPM and DPM as well as modified versions of these models with freed factor loadings.

2. Methods

2.1. *Causal Models*

For all simulations, the following data generating mechanism will be simulated:

$$
\begin{aligned}
\begin{bmatrix} x_{it} \\ y_{it} \end{bmatrix} &= \Phi \begin{bmatrix} x_{i,t-1} \\ y_{i,t-1} \end{bmatrix} + \mathbf{B}_{ct} \begin{bmatrix} C_{1i} \\ C_{2i} \end{bmatrix} + U_{it} + \begin{bmatrix} \epsilon_{xit} \\ \epsilon_{yit} \end{bmatrix} \\
&= \begin{bmatrix} \phi_{xx} & \phi_{xy} \\ \phi_{yx} & \phi_{yy} \end{bmatrix} \begin{bmatrix} x_{i,t-1} \\ y_{i,t-1} \end{bmatrix} + \begin{bmatrix} \beta_{xc_1t} & \beta_{xc_2t} \\ \beta_{yc_1t} & \beta_{yc_2t} \end{bmatrix} \begin{bmatrix} C_{1i} \\ C_{2i} \end{bmatrix} + U_{it} + \begin{bmatrix} \epsilon_{xit} \\ \epsilon_{yit} \end{bmatrix}
\end{aligned} \tag{1}
$$

for $t = 2, ..., T$ and

$$
\begin{bmatrix} x_{i1} \\ y_{i1} \end{bmatrix} = \begin{bmatrix} \beta_{xc_11} & \beta_{xc_21} \\ \beta_{yc_11} & \beta_{yc_21} \end{bmatrix} \begin{bmatrix} C_{1i} \\ C_{2i} \end{bmatrix} + U_{i1} + \begin{bmatrix} \epsilon_{xi1} \\ \epsilon_{yi1} \end{bmatrix} .
$$

Here, for person i at timepoint t, $\begin{bmatrix} x_{it} & y_{it} \end{bmatrix}^T$ is the data vector, Φ is the matrix of lagged effects, $\mathbf{B}_{ct}$ is the matrix of effects of the time-invariant confounders, $\begin{bmatrix} C_{1i} & C_{2i} \end{bmatrix}^T$ is the vector of confounder values, U_{it} is the value of the time-specific confounder, and $\begin{bmatrix} \epsilon_{xit} & \epsilon_{yit} \end{bmatrix}^T$ is the vector of residuals.

In addition,

$$
\begin{bmatrix} \epsilon_{xit} \\ \epsilon_{yit} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_x & 0 \\ 0 & \psi_y \end{bmatrix} \right),
$$

$$
\begin{bmatrix} C_{1i} \\ C_{2i} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{C_1} & 0 \\ 0 & \psi_{C_2} \end{bmatrix} \right),
$$

$$
U_{it} \sim \mathcal{N} \left(0, \psi_u \right),
$$

3

for $i = 1, ..., N$.

This SCM is similar to the DPM as it does not have an explicit decomposition of within and between effects and it is thus an observation based model, rather than a residual based model. One way that it differs from the DPM, other than having an observed confounder instead of a latent factor, is that the DPM includes covariances between the residuals at each timepoint. Because our model is specified as a DAG, two-headed arrows are not included, and this is thus expressed as a latent confounder between the residuals at each timepoint. It can be shown that when these the effect of these confounders on the residuals are 1, this specification using a confounder u_t with $\sigma^2_{u_t} = \psi_u$ is equivalent to specifying a covariance between the residuals where $\sigma_{x_t y_t} = \psi_u$. Furthermore, the variance of the residuals at each timepoint, ψ_x and ψ_y are equal to the residual variances in the DPM, minus the variance of the unique factor u_t at that timepoint.

The specific SCMs that will be simulated are described below.

2.1.1. One confounder, time-stable effects. (Model 1)

When there is only one confounder C, and its effects are invariant over time, Equation 1 reduces to:

$$\begin{bmatrix} x_{it} \\ y_{it} \end{bmatrix} = \begin{bmatrix} \phi_{xx} & \phi_{xy} \\ \phi_{yx} & \phi_{yy} \end{bmatrix} \begin{bmatrix} x_{i,t-1} \\ y_{i,t-1} \end{bmatrix} + \begin{bmatrix} \beta_{xc} \\ \beta_{yc} \end{bmatrix} C_i + U_{it} + \begin{bmatrix} \epsilon_{xit} \\ \epsilon_{yit} \end{bmatrix}.$$

The following choices were made for the parameters for the simulation: $\Phi = \begin{bmatrix} \phi_{xx} & \phi_{xy} \\ \phi_{yx} & \phi_{yy} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.15 \\ 0.1 & 0.3 \end{bmatrix}$, $\beta_c = \begin{bmatrix} 0.10 \\ 0.12 \end{bmatrix}$, $\psi_u = 0.3$, $\psi_x = \psi_y = 1 - \psi_u = 0.4$, and $\psi_C = 5$.

2.1.2. One confounder, time-stable effect for x, time-varying effect for y (Model 2)

The following choices were made for the parameters for the simulation: $\Phi = \begin{bmatrix} \phi_{xx} & \phi_{xy} \\ \phi_{yx} & \phi_{yy} \end{bmatrix} = \begin{bmatrix} 0.2 & 0.15 \\ 0.1 & 0.3 \end{bmatrix}$, $\beta_{ct} = \begin{bmatrix} 0.10 \\ 0.12 \end{bmatrix}$ for $t \leq 47$, $\beta_{ct} = \begin{bmatrix} 0.10 \\ 0.20 \end{bmatrix}$ for $t \geq 48$, $\psi_u = 0.3$, $\psi_x = \psi_y = 1 - \psi_u = 0.4$, and $\psi_C = 5$. Thus, until $t = 48$, this model is equivalent to the previous one. However, at $t = 48$, the effect of C on y increases (and afterwards remains stable).

2.1.3. Two confounders. Time-stable effects on x, time-varying effects on y. (Model 3)

Include: At $t \leq 47$, $\boldsymbol{B}_{ct} = \begin{bmatrix} 0.1 & 0.08 \\ 0.12 & 0.15 \end{bmatrix}$. At $t \geq 48$, $\boldsymbol{B}_{ct} = \begin{bmatrix} 0.1 & 0.08 \\ 0.15 & 0.12 \end{bmatrix}$.

Mogelijk verschillende sample momenten vergelijken. Dus wat als het bij $t = 1$ verandert, wat als bij $t = 3$ en wat als bij $t = 5$

2.2. *Model Evaluation*

I will simulate 1000 datasets according to each SCM, each with T = 50, and N = 500. The first 45 timepoints will then be discarded, as these are used to ensure convergence towards a trend. The last 5 timepoints will be kept and analyzed.

Model 1 and Model 2 will be analyzed using a version of the RI-CLPM and the DPM with only one factor. In addition to this, Model 2 will also be analyzed using an alternative of the RI-CLPM with free factor loadings. This is not necessary for the DPM, as a DPM with 1 latent factor and free loadings is equivalent to a version of the RI-CLPM with 1 factor and free loadings (Lüdtke & Robitzsch, 2021). Model 3 will be analyzed using the RI-CLPM and the DPM, as well as versions of these with free factor loadings.

For all models, at all iterations, the cross-lagged effect of x_4 on y_5 will be extracted and evaluated. Specifically, the biases and MSEs corresponding with the analysis techniques for the different causal models will be reported with their corresponding Monte Carlo SEs, as well as the coverage rate of the 95% confidence interval (see Morris et al., 2019).

3. Results

Maybe do model 1, result 1, model 2, result 2, model 3, result 3 instead (instead of Models, Results)

4. Discussion

References

Brown, D. W., Greene, T. J., Swartz, M. D., Wilkinson, A. V., & DeSantis, S. M. (2021). Propensity score stratification methods for continuous treatments. *Statistics in Medicine*, *40*(5), 1189–1203. https://doi.org/10.1002/sim.8835

Lüdtke, O., & Robitzsch, A. (2021). *A Critique of the Random Intercept Cross-Lagged Panel Model* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6f85c

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Murayama, K., & Gfrörer, T. (2022). *Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for choosing a statistical model from a causal inference perspective* [Preprint]. PsyArXiv.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*(5), 637–657. https://doi.org/10.1037/met0000210

Vansteelandt, S., & Daniel, R. m. (2014). On regression adjustment for the propensity score. *Statistics in Medicine*, *33*(23), 4053–4072. https://doi.org/10.1002/sim.6207