One Step Toward Causality: Unobserved Time-Invariant Confounding in Cross-Lagged Panel Models

Pepijn Vink (6100252)

Utrecht University

MSBBSS Research Report

Supervisors: Ellen Hamaker, Jeroen Mulder

22 Dec 2023

blank words

One Step Toward Causality: Unobserved Time-Invariant Confounding in Cross-Lagged Panel Models

Introduction

Cross-lagged panel designs, which are based on measuring the same individuals on the same variables for two or more occasions, are commonly used in psychological research to investigate relationships between two or more variables over time. In the last few years, the random intercept cross-lagged panel model (RI-CLPM, Hamaker et al., 2015; Mulder & Hamaker, 2021) has become a popular method for analyzing cross-lagged relationships in panel data. The RI-CLPM, shown in Figure 1a, is an extension of the cross-lagged panel model (CLPM). The model is used to decompose observed variables into stable between person differences and temporal within person components, and the dynamics that are usually most of interest are modeled using these within person components. Specifically, the interest is generally in the effects of the variables on each other at later timepoints: the cross-lagged effects. These cross-lagged parameters are often interpreted as causal effects, albeit implicitly. However, one assumption that such causal interpretations rely on is that both time-varying and time-invariant confounders are adequately controlled for.

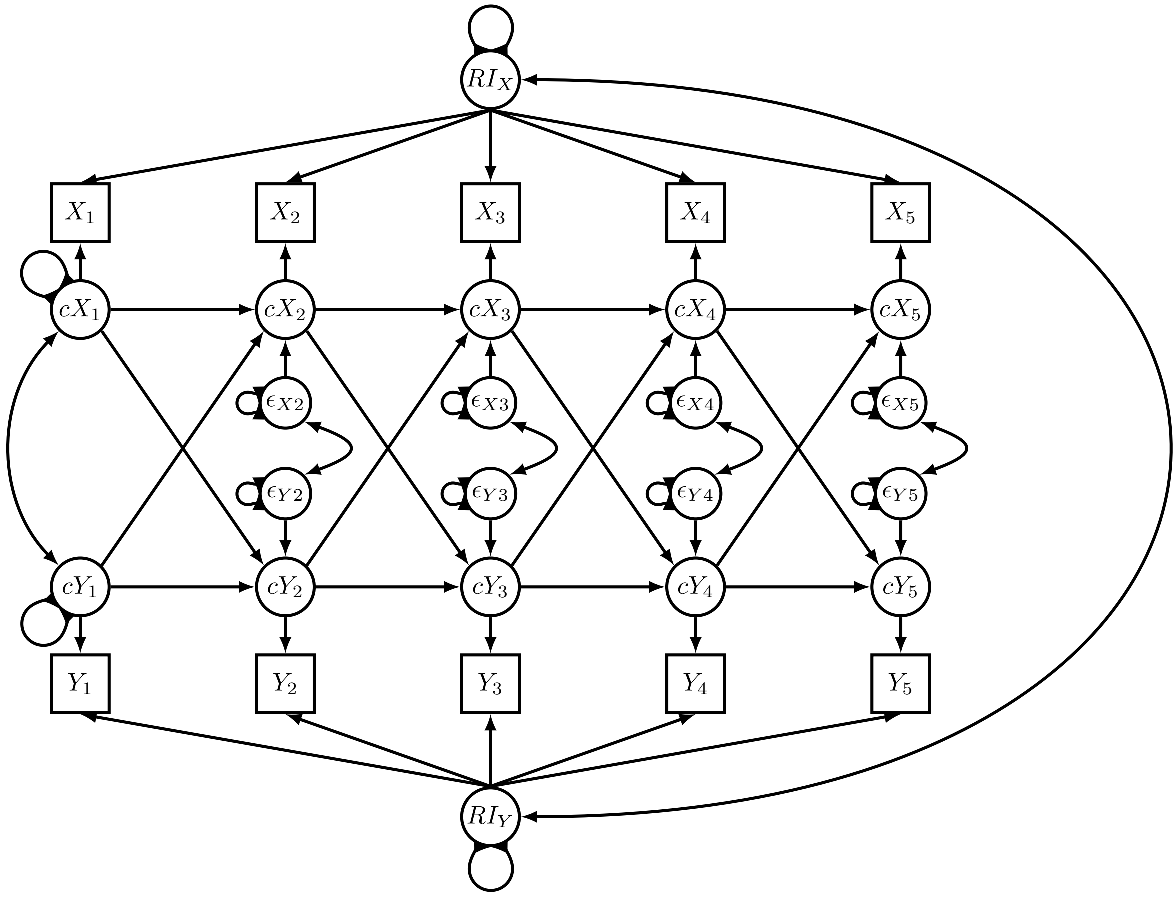
Usami et al. (2019) show that when certain assumptions are met, the random intercept in the RI-CLPM controls for unobserved heterogeneity. This requires, in particular, that the confounders are time-invariant, and that the effects of the confounders on the variables of interest are stable over time. However, it has not yet been studied how the RI-CLPM performs when the effects of unobserved time-invariant confounders are of a time-varying nature. Murayama and Gfrörer (2022) show that the Dynamic Panel Model (DPM), which is shown in Figure 1b, may be an alternative in this case. This model is similar to the RI-CLPM due to the inclusion of lagged effects as well as a latent component, but the dynamics are now situated between the observed variables, rather than between the within-person components. While some prefer this model because it is less restrictive and it is claimed to better control for unobserved confounding (Allison et al., 2017), others would regard the absence of an explicit within-between decomposition as a downside (e.g. Hoffman & Hall, 2024). Furthermore, it is unknown how either of these models perform when the underlying causal model is more complex, for example when multiple confounders with time-varying effects are involved.

Therefore, the effect of time-invariant confounders on both the RI-CLPM and the DPM should be explored further. The current research report will evaluate the extent to which unobserved time-invariant confounders affect the estimates of the RI-CLPM and DPM. We will assess whether the estimates of the RI-CLPM and the DPM give a good indication of underlying causal effects when multiple unobserved time-invariant confounders are involved. This will serve as a preliminary exploration for the thesis, which will address the use of causal inference methods such as propensity score adjustment and inverse probability weighting (e.g. Brown et al., 2021; Vansteelandt & Daniel, 2014) to control for observed time-invariant confounders in the RI-CLPM.

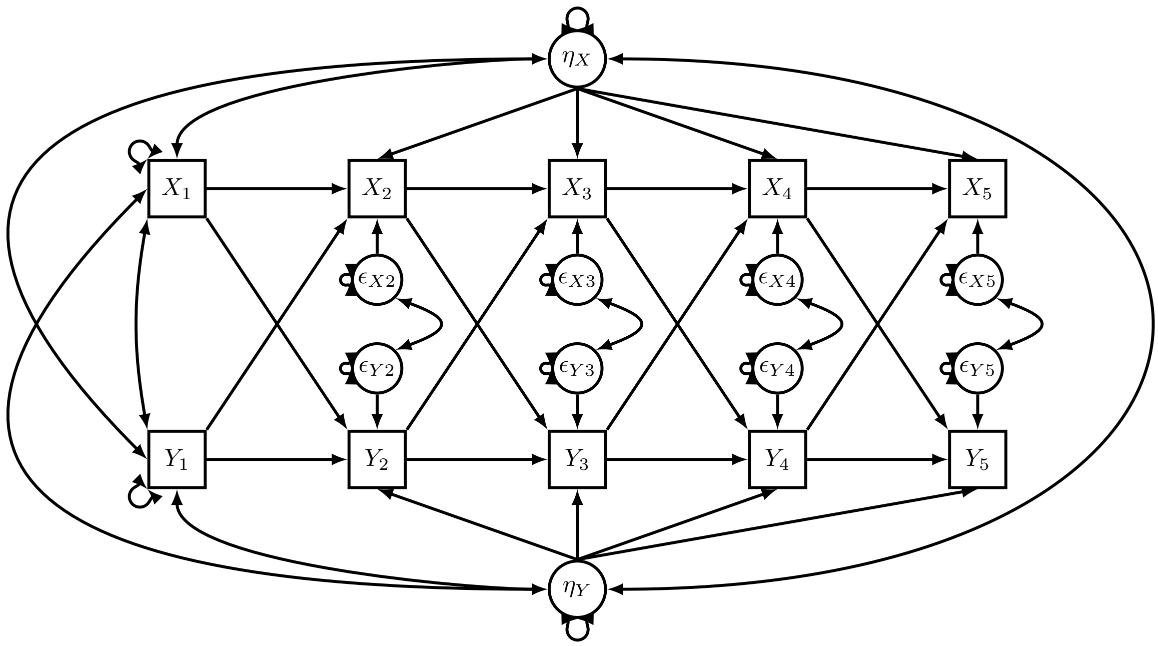
This report is structured as follows. We will start with a conceptual comparison between the RI-CLPM and the DPM. Then, a hypothetical data generating mechanism that includes multiple time-invariant confounders will be introduced. This will be used for a simulation study that assesses the performance of the RI-CLPM and the DPM under the data generating mechanism. Results and implications will be discussed.

A Comparison of the RI-CLPM and the DPM

The RI-CLPM (Hamaker et al., 2015) is often used with the goal to separate within person dynamics from stable between person differences. It decomposes the observed scores of an individual in a stable person-specific deviation from the grand mean, i.e. the random intercept representing the 'between' part of the model, and a temporal deviation from this stable component, i.e. the 'within' part of the model. As can be seen in Figure 1a, the RI-CLPM implies only direct effects of the stable trait on the observed variables, as well as no effect of the stable trait of one variable on the observed scores of the other. A relationship with the observed scores of the other



(a) *Random Intercept Cross-Lagged Panel Model (RI-CLPM)*



(b) *Dynamic Panel Model (DPM)*

Figure 1

Structural Equation Model Representations of Two Popular Models in Panel Research. Boxes Indicate Observed Variables and Circles Indicate Latent Variables.

variables is only modeled using a covariance between the stable trait factors.

The DPM, however, is typically used with the goal to model lagged effects, with a latent factor to control for time-invariant confounders (**alison2009**). It does not explicitly separate within person dynamics from stable between person differences, and regresses the observed scores on each other. The latent factors in the DPM are sometimes called ‘accumulating factors’ (Usami et al., 2019), as their effects on the observed variables are both direct, as well as indirect through lagged relationships between the observed variables themselves, a property that becomes clear from Figure 1b. Furthermore, to account for the fact that measurements are usually sampled at a random moment in time in an ongoing process, the observations at the first timepoints are often allowed to covary freely with each other and the accumulating factors (Hamaker, 2005), which is not necessary in the RI-CLPM.

Hamaker (2005) shows that in specific cases, the RI-CLPM and the DPM are statistically equivalent and yield equivalent estimates of the lagged parameters. Specifically, this is the case when lagged parameters are invariant over time and the factor loadings in the DPM at the first timepoint are constrained to reflect this, rather than specified as free covariances. This also implies that when these conditions hold in a causal model, both the RI-CLPM and the DPM should recover the true effects.

The Data Generating Mechanism

To answer the research question, we will perform a simulation study that uses various scenarios to generate data, all of which are special cases of the same data generating mechanism. To represent this general data generating mechanism, we use the causal Directed Acyclic Graph (DAG) in Figure 2. It shows a dynamic process for $t = 1, ..., 5$ between time-varying variables X and Y and includes two time-invariant baseline confounders: C_1 and C_2. This causal DAG does not yet have parametric assumptions. The DAG is most similar to the dynamic panel model, as observed values are determined by observed values at previous timepoints, but it can also be interpreted as only explicitely modeling the within person process.

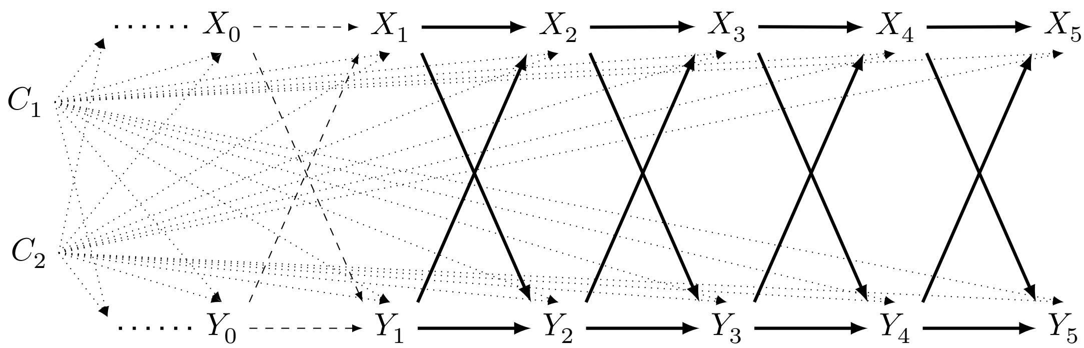


Figure 2

Directed Acyclic Graph for Cross-Lagged Relationships with 2 Confounders. Dashed Lines Indicate the Process Before the Observed Timepoints.

Assuming the underlying process has been going on for long enough that it has stabilized around an equilibrium, when lagged effects as well as the effects of the confounders are time-stable, in both the RI-CLPM and the DPM the latent factor will be a linear combination of the confounders (Usami et al., 2019). Therefore, when these conditions hold, both models should already yield unbiased estimates of the lagged effects, even when the confounders are unobserved. However, when the effects of the confounders are not time-stable, the models may not fully take the confounders into account thus resulting in biased effects. In the next section, a simulation study will be described that simulates from our DAG and evaluates the performance of the RI-CLPM and the DPM when confounders are unobserved.

Methods

In previous sections, the RI-CLPM and the DPM have been compared conceptually and their ability to account for unobserved time-invariant confounding has been touched upon. In this section, a simulation study will be presented that evaluates their performance under three different scenarios. These scenarios are based on the data generating mechanism in Figure 2 and differ with respect to the stability of the effects of the confounders.

Data Generation

Below, the different scenarios of the data generating mechanism that will be simulated are described, each having different effects of the confounders. These all follow the following general data generating mechanism:

For each person i at timepoint t,

$$x_{it} = \phi_{xx}x_{i,t-1} + \phi_{xy}y_{i,t-1} + \gamma_{1t}C_{1i} + \gamma_{2t}C_{2i} + \epsilon_{xit},$$

$$y_{it} = \phi_{yy}y_{i,t-1} + \phi_{yx}x_{i,t-1} + \delta_{c1t}C_i + \delta_{2t}C_{2i} + \epsilon_{yit}.$$

Furthermore, at the first sampled timepoint,

$$x_i = \gamma_1 C_{1i} + \gamma_2 C_{2i} + \epsilon_{xi},$$

$$y_i = \delta_1 C_{1i} + \delta_2 C_{2i} + \epsilon_{yi}.$$

In addition,

$$\epsilon_{xt} \sim \mathcal{N}(0, \psi_x),$$

$$\epsilon_{yt} \sim \mathcal{N}(0, \psi_y),$$

$$C_1 \sim \mathcal{N}(0, \psi_{C_1}),$$

$$C_2 \sim \mathcal{N}(0, \psi_{C_2}).$$

All lagged effect, i.e. cross-lagged effects and autoregressions, and residual variances will be time-invariant, whereas effects of the confounders may be time-varying. For all scenarios lagged effects are set to $\phi_{xx} = 0.2$, $\phi_{yy} = 0.3$, $\phi_{xy} = 0.15$, $\phi_{yx} = 0.1$ (based on Mulder, 2023) and the residual variances are set to $\psi_x = 1$, $\psi_y = 1$, $\psi_{C_1} = 1$, and $\psi_{C_2} = 1$.

Three scenarios will be simulated. For all scenarios, at the start, $\gamma_1 = 0.3$, $\gamma_2 = 0.8$, $\delta_1 = 0.5$, and $\delta_2 = 0.2$. For scenario 1, the effects of the confounders remain stable. For scenario 2, at $t = 3$, the effect of C_1 on x and y decreases and remains stable afterwards, $\gamma_{1t} = 0.1$ and $\delta_{1t} = 0.2$. For scenario 3, at $t = 3$ all effects confounder effects change and afterwards remain stable $\gamma_{1t} = 0.6$ and $\delta_{1t} = 0.2$, $\gamma_{2t} = 0.3$ and $\delta_{2t} = 0.5$.

1000 datasets are simulated according to each scenario with N=500. To allow for convergence to an equilibrium, 50 timepoints are simulated, of which 45 are used as burn-in. The remaining 5 timepoints are used for analysis.

Analysis

To assess the performance of the described models when the confounders are unobserved, all simulated datasets described above will be analyzed using the RI-CLPM and the DPM, as well as versions of these models with free factor loadings, as these may, in part, capture some time-varying heterogeineity due to the unobserved confounders. Models that do not converge or do not result in positive definite covariance matrices will be excluded. All simulations and analyses will be done using R (R Core Team, 2022) and models will be fit using the lavaan package (Rosseel, 2012).

Results

The discussion of results will focus on the $x_4 \rightarrow y_5$ effect, which is equal to $\phi_{xy} = 0.1$. For each of the models, and for each scenario, the bias, Monte-Carlo error, root mean squared error (RMSE), and coverage were computed, based on recommendations by Morris et al. (2019). The bias is computed as the difference between the mean of the empirical sampling distribution, i.e. the distribution of our estimates, and the true effect. The Monte-Carlo error can be interpreted as the standard error of this estimate. The RMSE is the root of the mean squared deviation from the true value and thus expresses the expected deviation for a single sample. The coverage is the proportion of times that the 95% confidence interval contains the true value. The coverage rate is typically seen as acceptable when it exceeds 95%. For a more elaborate explanation and computation of these measures, we refer the reader to Morris et al. (2019).

Upon further inspection of the models, we found that for scenario 3, a very large number of models resulted in a non positive-definite matrix of the latent variables. Specifically, the random intercepts had a correlation higher than 1 in all of the RI-CLPM models, and the accumulating factors had a correlation higher than 1 in most of the latent DPM models (91.9% and 55.8% for the DPM without and with free factor loadings respectively). Therefore, for subsequent explorations, the models were run again, with correlations between random intercepts or accumulating factors set to 1. It should be kept in mind, however, that this constitutes an exploratory approach.

Table 1 shows the bias, relative bias, RMSE, and coverage for the $x_4 \rightarrow y_5$ effect for each of the models in each situation. Furthermore, Figure 3 visualizes these biases. The error bars indicate the Monte Carlo error.

Table 1 and Figure 3 shows that under scenario 1, as expected, both the RI-CLPM and the DPM yield unbiased estimates. Surprisingly, freeing the factor loadings, however, results in bias. Specifically, they yield a negative bias, indicating that they underestimate the effect. Furthermore, the RI-CLPM has an RMSE of 0.06 and the DPM has an RMSE of 0.04, indicating that under this scenario, the DPM shows less average deviation from the true effect and thus may be a better choice. Moreover, only the DPM and the RI-CLPM with free loadings have an acceptable observed coverage.

Under scenario 2, where the effects one of the confounders is time-varying, both the RI-CLPM and the DPM yield biased effects. Specifically, the RI-CLPM underestimates the effect with 49.81%, a quite large relative bias, and the DPM overestimates it with 11.83%. Furthermore, freeing the factor loadings does not seem to capture the time-varying effects of the confounders. Moreover, the DPM has the lowest RMSE of all models with 0.04. Only the versions of the RI-CLPM and the DPM with free loadings have an acceptable coverage.

Under scenario 3, where both confounders have time-varying effects, all models have quite high bias. In this situation the RI-CLPM and DPM with free factor loadings have lower bias than their counterparts with fixed loadings. Furthermore, the versions with free factor loadings have acceptable coverage rates (although they could be considered too high). However, this comes with a higher RMSE.

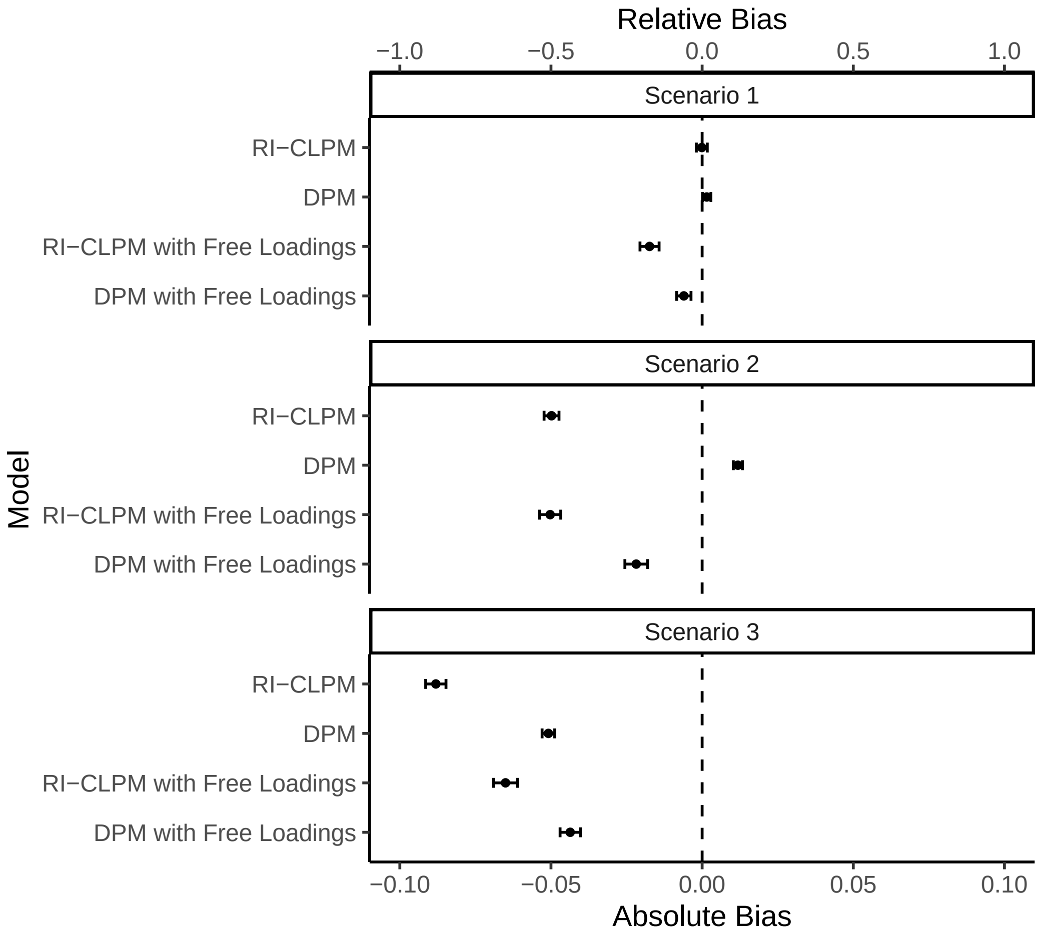


Figure 3

Biases for All Ccenarios. The Dots Indicate the Bias of the Models and the Error Bars Indicate the Monte Carlo Error.

Discussion

Previous research has suggested that when lagged effects and the effects of unobserved time-invariant confounders are stable over time, both the RI-CLPM and the DPM yield unbiased estimates of the true cross-lagged effect, even when the confounders are unobserved (e.g. Usami et al., 2019). This characteristic of both models was replicated by the discussed simulations in scenario 1. However, the extent to which this is a reasonable causal model is debatable. For many phenomena, it may be more realistic to assume effects of confounders to vary over time.

Therefore, we also considered situations where effects of one or two confounders changed over time. In our simulations, the RI-CLPM and the DPM underestimated the

Table 1

Bias, RMSE, and Covarage for Each Scenario on the Effect of x_4 on y_5. The true effect is 0.1

Scenario	Model	Bias	Relative Bias (%)	RMSE[1]	Coverage (%)[2]
1	RI-CLPM	0.00	-0.11	0.06	94.40
	DPM	0.00	1.52	0.04	95.69
	RI-CLPM (free loadings)	-0.02	-17.39	0.10	95.81
	DPM (free loadings)	-0.01	-6.06	0.07	94.73
2	RI-CLPM	-0.05	-49.81	0.08	88.94
	DPM	0.01	11.83	0.04	94.52
	RI-CLPM (free loadings)	-0.05	-50.27	0.11	97.78
	DPM (free loadings)	-0.02	-21.80	0.09	95.42
3	RI-CLPM	-0.09	-88.09	0.11	71.70
	DPM	-0.05	-50.86	0.07	79.10
	RI-CLPM (free loadings)	-0.07	-65.04	0.12	98.00
	DPM (free loadings)	-0.04	-43.64	0.10	96.65

[1] Root-Mean-Square-Error

[2] Inclusion frequency of the true effect for the 95% confidence interval

cross-lagged effects when this was the case. Furthermore, freeing the factor loadings in both models did not fully capture the time-varying nature as negative bias was observed in these models as well. However, because only few scenarios were considered, it cannot yet be concluded that these models generally underestimate cross-lagged effects when confounders are unobserved; other data generating mechanisms may show different patterns. Furthermore, although bias in the estimates was present in these scenarios, conclusions on the direction of the effect generally did not change, as the observed effects were still positive.

The analyses performed in our simulation were used on simulated data with unobserved confounders. When confounders are observed, however, it may be more sensible to model them explicitly. This can be done by including them in the model as covariates (Mulder & Hamaker, 2021), but methods specifically designed for causal inference may be more appropriate (Leite et al., 2019; Schafer & Kang, 2008), especially when many confounders are involved or when effects are nonlinear. In reality, however, only a subset of confounders may realistically be observed, whereas others are unobserved. In this situation, it may be sensible to use a combination of methods, such as using the RI-CLPM or DPM to correct for some unobserved confounders, and using causal inference methods to correct for observed confounders. However, the extent to which this is possible and sensible is yet to be explored.

It is therefore necessary to investigate whether the toolbox of social science researchers can be expanded by examining the use of causal inference methods in cross-lagged panel models, which will be covered by the thesis. The presented results highlight the necessity of such research and highlight the importance of acting with caution when attempting to assess cause and effect.

References

Allison, P. D., Williams, R., & Moral-Benito, E. (2017). Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects. *Socius*, *3*, 2378023117710578. https://doi.org/10.1177/2378023117710578

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

Hamaker, E. L. (2005). Conditions for the Equivalence of the Autoregressive Latent Trajectory Model and a Latent Growth Curve Model With Autoregressive Disturbances. *Sociological Methods & Research*, *33*(3), 404–416. https://doi.org/10.1177/0049124104270220

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

Hoffman, L., & Hall, G. J. (2024). Considering between- and within-person relations in auto-regressive cross-lagged panel models for developmental data. *Journal of School Psychology*, *102*, 101258. https://doi.org/10.1016/j.jsp.2023.101258

Leite, W. L., Stapleton, L. M., & Bettini, E. F. (2019). Propensity Score Analysis of Complex Survey Data with Structural Equation Modeling: A Tutorial with Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(3), 448–469. https://doi.org/10.1080/10705511.2018.1522591

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

Mulder, J. D. (2023). Power Analysis for the Random Intercept Cross-Lagged Panel Model Using the powRICLPM R-Package. *Structural Equation Modeling: A Multidisciplinary Journal*, *30*(4), 645–658. https://doi.org/10.1080/10705511.2022.2122467

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

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.

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing.

Rosseel, Y. (2012). {Lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313. https://doi.org/10.1037/a0014268

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