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

Pepijn Vink (6100252)[1], Ellen Hamaker[1,2], Jeroen Mulder[1,2]

ARTICLE HISTORY
Compiled December 19, 2023

[1] Department of Methodology and Statistics, Utrecht University,
[2] Supervisor,

KEYWORDS
RI-CLPM; DPM; causality; longitudinal; simulation; confounding

1. Introduction

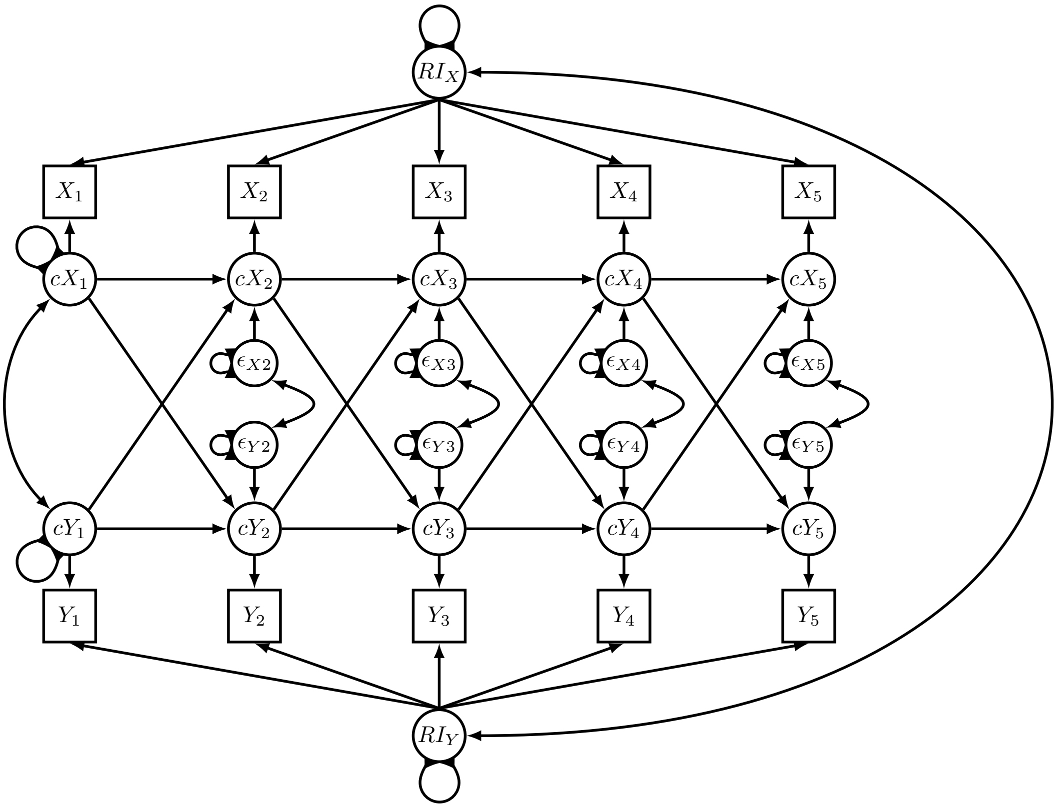
Cross-lagged panel designs are a popular method in psychological research to investigate relationships between two or more variables over time. In the last few years, the random intercept cross-lagged panel model (RI-CLPM, Hamaker et al., 2015; Mulder & Hamaker, 2021) has become a popular method for analyzing cross-lagged relationships in panel data. This is longitudinal, non experimental data with 3-10 measurement moments. 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 as within person dynamics using these within person components. Specifically, the interest is often in the effects of the variables on each

CONTACT: Pepijn Vink (6100252). Email: p.a.vink@uu.nl.

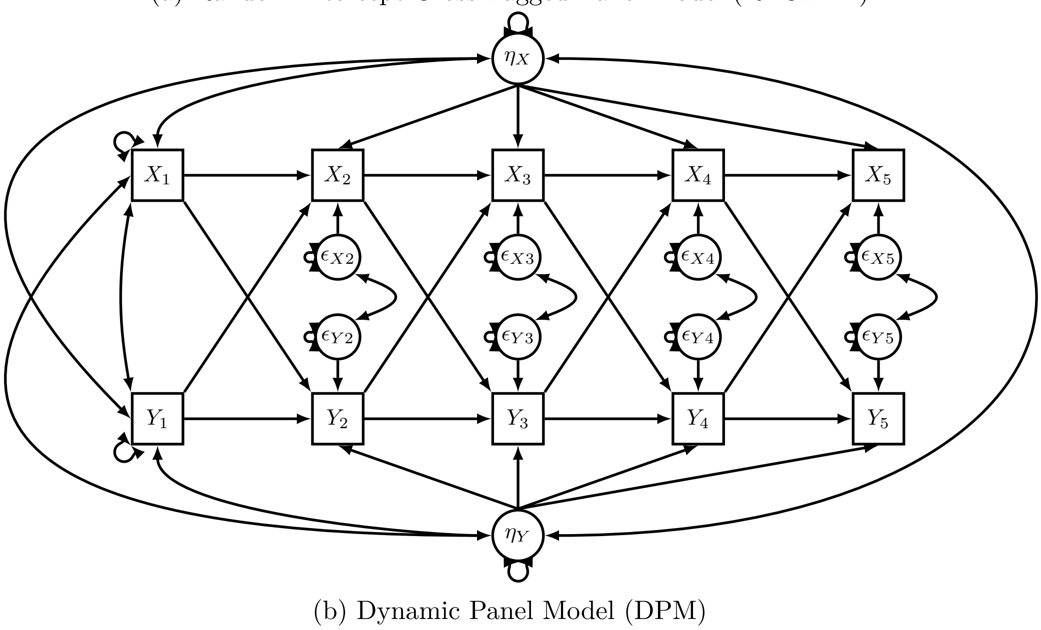
other at later timepoints: the cross-lagged effects. These cross-lagged parameters are often interpreted as causal effects, albeit implicitly. However, one assumption that such causal interpretations rely on is that both time-varying and time-invariant confounders are adequately controlled for.

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 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 some 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 compare time-invariant confounders in panel data with time-stable effects and time-varying effects, and assess their effects on estimates of the RI-CLPM and DPM when they are unobserved. 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.



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



(b) Dynamic Panel Model (DPM)

Figure 1. Two Popular Models in Panel Research. Boxes Indicate Observed Variables and Circles Indicate Latent Variables.

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 and the potential of the RI-CLPM and the DPM to control for unobserved confounding when this mechanism is true will be discussed. After this, a simulation study will be performed to assess the performance of the RI-CLPM and the DPM under the data generating mechanism. Results and implications will be discussed.

2. 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 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. 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.

3. 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.

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

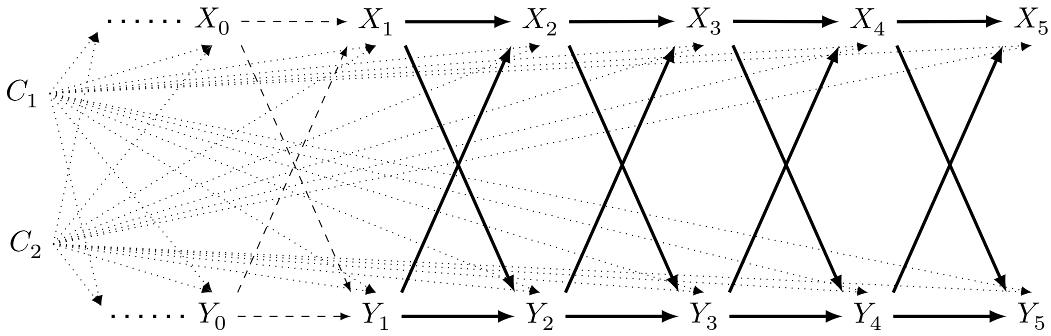


Figure 2. Directed Acyclic Graph for Cross-Lagged Relationships with 2 Confounders. Dashed Lines Indicate the Process Before the Observed Timepoints.

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.

4. 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.

4.1. ***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.

4.2. *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).

5. 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 4.98%, a quite large relative bias, and the DPM overestimates it with 1.18%. 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.

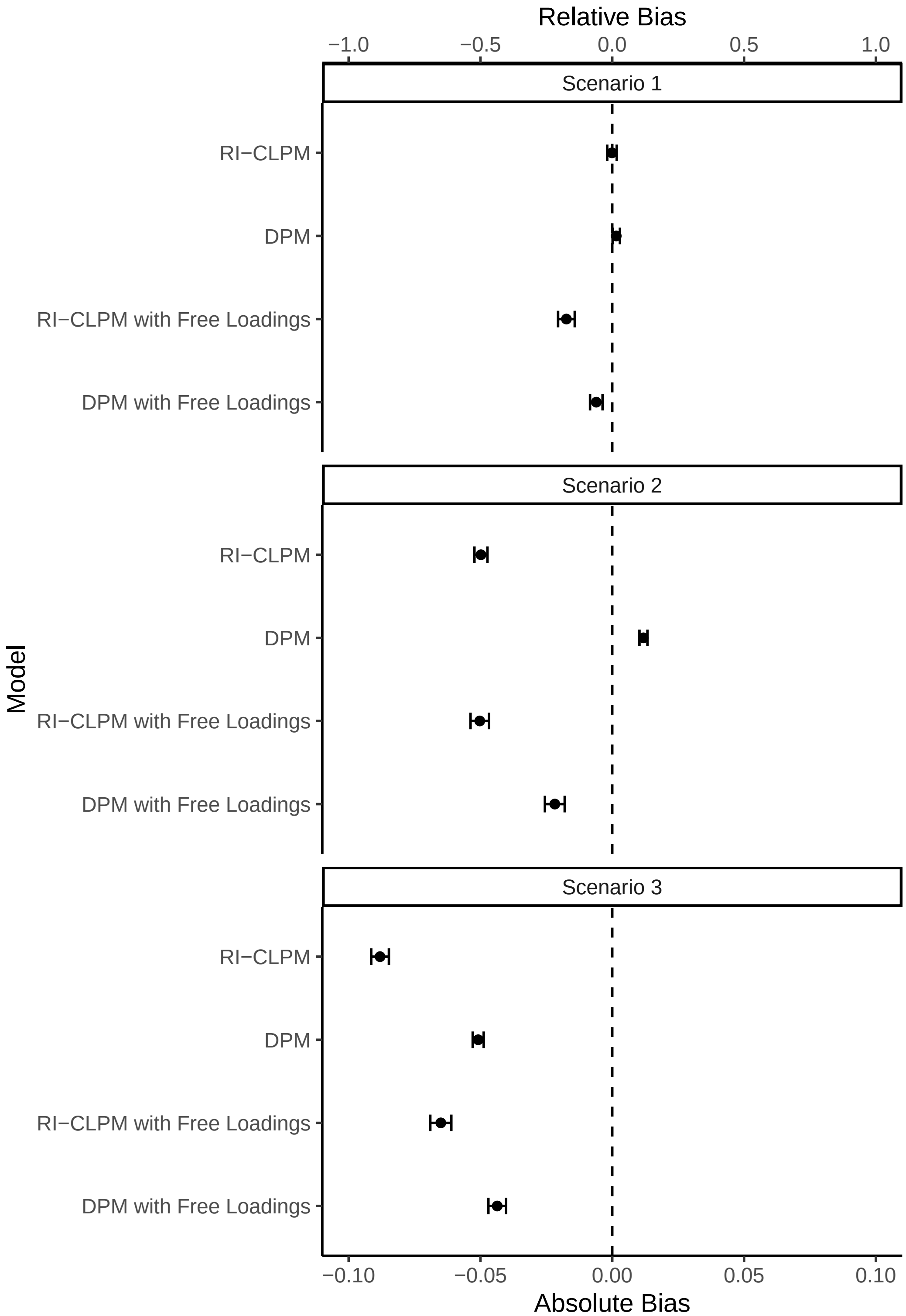


Figure 3. Biases for All Ccenarios. The Dots Indicate the Bias of the Models and the Error Bars Indicate the Monte Carlo Error.

Table 1. Bias, RMSE, and Covarage for Each Scenario.

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

6. Discussion

It has been known that when lagged effects as well as the effect of unobserved time-invariant confounders are stable over time, both the RI-CLPM and the DPM can recover the true effect, even when the confounders are unobserved. This characteristic of both models was replicated by the discussed simulations. 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. Furthermore, when times between measurements are not equal, time-stable lagged effect cannot be assumed either.

Moreover, the biases observed in our simulations were relatively small; although the estimates deviated from the true effects, conclusions regarding presence of an effect may remain unchanged. Furthermore, the advantage of the used analyses is that confounders did not need to be observed.

Still, the scenarios presented in the simulations were relatively simple. When more confounders are involved, methods that need confounders to be observed may be necessary to obtain good estimates of causal effects. When there are few confounders and effects of confounders are linear, regression adjustment may be an option. However, this method may break down when more confounders are involved and the effects are nonlinear, as is often the case in social science research.

It is therefore necessary to expand the toolbox of social science researchers by investigating 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 techniques and highlight the importance of acting with cause when attempting to assess cause and effect.

Pearl and Mackenzie (2019) argue that science has been in a 'causal revolution' and as Haber et al. (2022) show, there is an abundance of causal language present in modern research. It is therefore essential that causal effects are estimated with great care. Evaluating causality using reciprocal effects in longitudinal data can bring researchers in the social science one step closer to obtaining 'true' causal effects, but more will be necessary.

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

Haber, N. A., Wieten, S. E., Rohrer, J. M., Arah, O. A., Tennant, P. W. G., Stuart, E. A., Murray, E. J., Pilleron, S., Lam, S. T., Riederer, E., Howcutt, S. J., Simmons, A. E., Leyrat, C., Schoenegger, P., Booman, A., Dufour, M.-S. K., O'Donoghue, A. L., Baglini, R., Do, S., … Fox, M. P. (2022). Causal and Associational Language in Observational Health Research: A Systematic Evaluation. *American Journal of Epidemiology*, *191*(12), 2084–2097. https://doi.org/10.1093/aje/kwac137

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

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.

Pearl, J., & Mackenzie, D. (2019). *The Book of Why: The New Science of Cause and Effect*. Penguin Books UK.

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

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