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

Pepijn Vink (6100252)[1]

ARTICLE HISTORY
Compiled December 22, 2023

[1] Methodology and Statistics for the Behavioral, Biomedical, and Social Sciences, Utrecht University,

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

1. Introduction

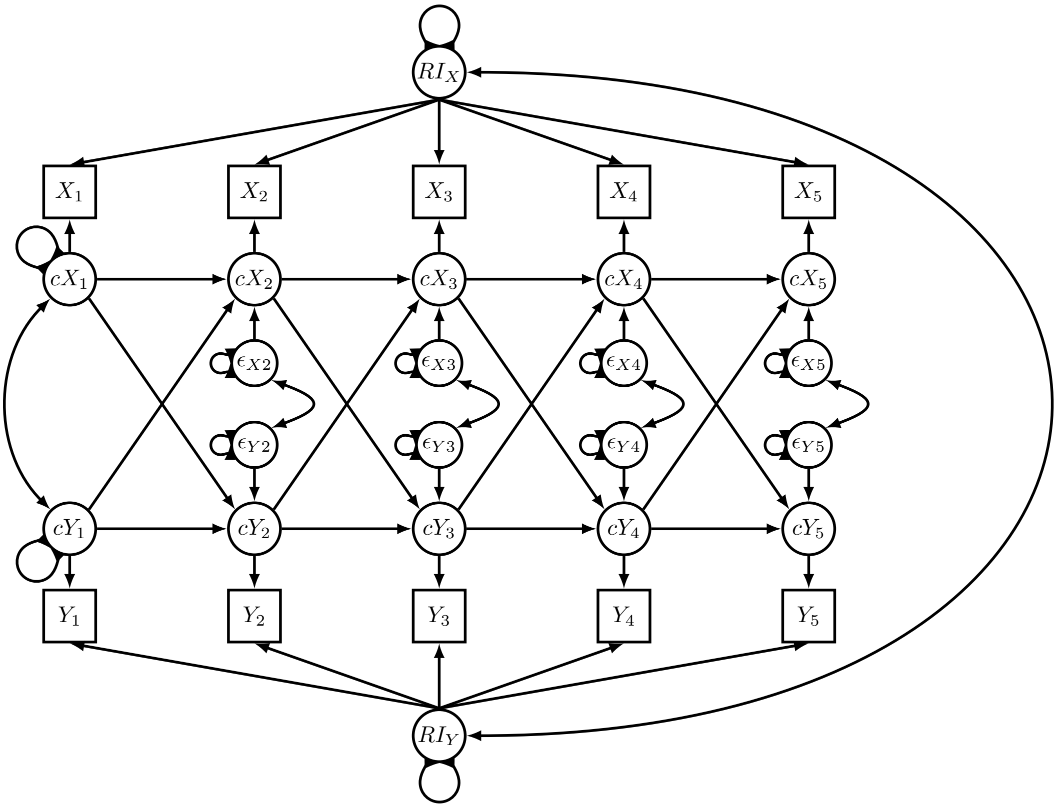
Cross-lagged panel designs, which are based on measuring the same individuals on the same variables for two or more occasions, are commonly used in psychological research to investigate relationships between two or more variables over time. Multiple longitudinal SEM models have been developed for analyzing such data, of which the random intercept cross-lagged panel model (RI-CLPM, Hamaker et al., 2015; Mulder & Hamaker, 2021) is one of the most popular. The RI-CLPM, shown in Figure 1a, is an extension of the cross-lagged panel model (CLPM). The model decomposes 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

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

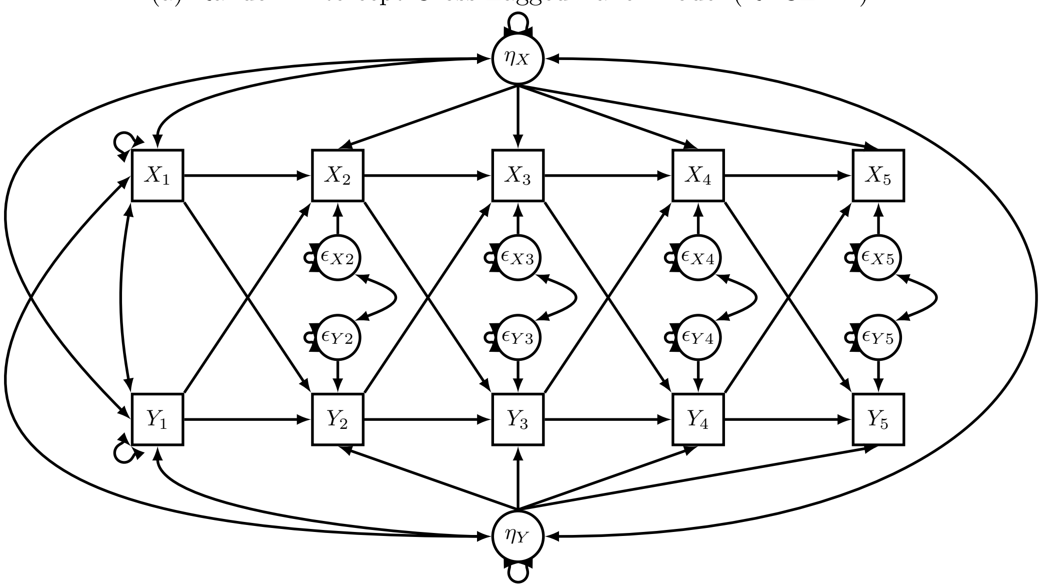
other at later timepoints: the cross-lagged effects. These cross-lagged parameters are often interpreted as causal effects, albeit implicitly. However, a causal interpretation presumes that both time-varying and time-invariant confounders are adequately controlled for, an assumption commonly refered to as absence of unobserved confounding.

Usami et al. (2019) show that when certain assumptions are met, the random intercept in the RI-CLPM controls for unobserved heterogeneity resulting from confounding. 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), 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 process 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,



(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.

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

2. A Comparison of the RI-CLPM and the DPM

The RI-CLPM is commonly used with the goal to investigate causal relationships between variables while accounting for stable between-person differences (Hamaker et al., 2015). It decomposes the observed scores of an individual in a stable person-specific deviation from the grand mean, and a temporal deviation from this stable component through the inclusion of a random intercept factor. These temporal deviations make up the 'within' part of the model where the dynamics are modeled. As can be seen in Figure 1a, this factor only has direct effects on the observed variables, and no effect on the observed scores of the other variable. The random intercept of x is only related to the observed scores of y through a covariance between the random intercepts, and vice versa.

The DPM, however, is typically used with the goal to model lagged effects, with a latent factor to control for unmeasured time-invariant confounders (Allison et al., 2017). It does not explicitly separate within person dynamics from stable between person differences, and lagged effects are included on the observed scores. 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. 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 data generating mechanism, both the RI-CLPM and the DPM should recover the true effects. In the next section, a simulation study is described that simulates from a causal model and evaluates the performance of the RI-CLPM and the DPM when confounders are unobserved.

3. Methods

In this section, a general data generating mechanism is presented, which is used for a simulation study that evaluates the performance of the RI-CLPM and the DPM under three different scenarios that include unobserved confounders. These scenarios differ with respect to the stability of the effects of the confounders.

3.1. ***The Causal Model***

To represent the causal model that we will simulate, 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 that each have an effect on all future observations of x and y. The structure of the DAG is most similar to the dynamic panel model, as observed values are determined by observed values at previous timepoints, but it may also represent confounding on the within-person part of the RI-CLPM. However, it should be noted that the DAG is not equivalent to either model, as each confounder has effects on both x and y, whereas the latent factors in the RI-CLPM and DPM are variable-specific.

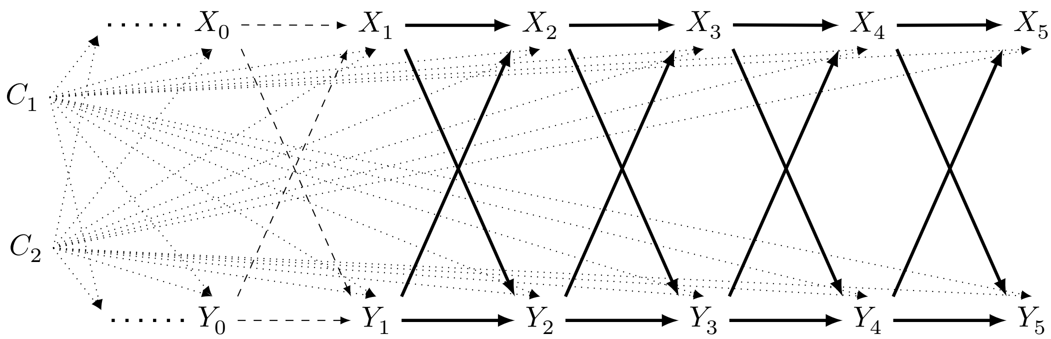


Figure 2. Directed Acyclic Graph for Cross-Lagged Relationships with 2 Confounders. Dashed Lines Indicate the Process Before the Observed Timepoints.

When the underlying dynamic process has been going on for long enough such that scores are independent of arbitrary starting values, and the process has therefore stabilized around an equilibrium, and when lagged effects as well as the effects of the confounders are time-invariant, in both the RI-CLPM and the DPM the latent factors will be a linear combination of the confounders (Usami et al., 2019). Therefore, when these conditions hold, we expect both models to 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 be able to fully account for the effects of confounders, thus resulting in biased effects.

The causal model introduced above was simulated to assess the performance of the RI-CLPM and the DPM under different patterns of confounding. Below, the different scenarios that will be simulated are described, which each have different effects of the confounders. These were all simulated as special cases of the data generating mechanism presented earlier and are represented by the following equations:

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, that is, the cross-lagged effects and autoregressions, and residual variances are time-invariant (as indicated by the absence of a time index t for these parameters), whereas effects of the confounders may be time-varying (as indicated by the presence of a time index t for these parameters). In our simulation, for all scenarios, lagged effects were 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 were set to $\psi_x = 1$, $\psi_y = 1$, $\psi_{C_1} = 1$, and $\psi_{C_2} = 1$.

Three scenarios were 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 effects of C_1 on x and y decrease to $\gamma_{1t} = 0.1$ and $\delta_{1t} = 0.2$, and remain stable afterwards. For scenario 3, at $t = 3$ all effects of confounders change and afterwards remain stable. Specifically, for C_1 its effect on x increases and its effect on y decreases, $\gamma_{1t} = 0.6$ and $\delta_{1t} = 0.2$. For C_2 its effect on x decreases and its effect on y increases, $\gamma_{2t} = 0.3$ and $\delta_{2t} = 0.5$.

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

3.2. *Analysis*

To assess the performance of the described models when the confounders are unobserved, all simulated datasets described above were analyzed using the RI-CLPM and the DPM, as well as versions of these models with free factor loadings for the latent factors, because freeing the factor loadings may, in part, capture the time-varying effects of the unobserved confounders. Models that do not converge or do not result in positive definite covariance matrices were excluded. All simulations and analyses were done using R version 4.3.2 (R Core Team, 2022) and models will be fit using the lavaan package version 0.6-16 (Rosseel, 2012).

4. Results

The discussion of results will focus on the $x_4 \to y_5$ 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, that is the distribution of our estimates, and the true effect. The Monte-Carlo error is computed as the standard error of this estimate. The RMSE is the root of the mean squared deviation from the true value and is thus a measure of the expected absolute deviation for a single sample. Setting $alpha = 0.5$, 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 is close to 95%.

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 or accumulating factors had a correlation higher than

1 in all of the RI-CLPM models and in most of the 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 as it was not based on theory and therefore results may not generalize well to other situations.

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. 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 the effect is underestimated. Furthermore, DPM has a lower RMSE than the RI-CLPM, indicating that under this scenario, the DPM shows less average deviation from the true effect and thus may be a better choice. Moreover, all coverage rates are relatively acceptable.

Under scenario 2, where the effects of one of the confounders are time-varying, both the RI-CLPM and the DPM yield biased estimates. Specifically, based on the relative bias, the RI-CLPM shows a relatively extreme underestimation of the true effect, and the DPM yields a relatively small overestimation. Furthermore, freeing the factor loadings does not seem to capture the time-varying effects of the confounders, as indicated by the extreme relative biases. Moreover, the DPM has the lowest RMSE of all models. The RI-CLPM shows very low coverage, and the RI-CLPM with free loadings may have coverage that is too high.

Under scenario 3, where both confounders have time-varying effects, all models show large bias. In this situation, the RI-CLPM and DPM with free factor loadings have lower bias than their counterparts with fixed loadings. Furthermore, RI-CLPM and the DPM have very low coverage rates, and the versions with free factor loadings have coverage rates that may be too high. The DPM yields the lowest RMSE.

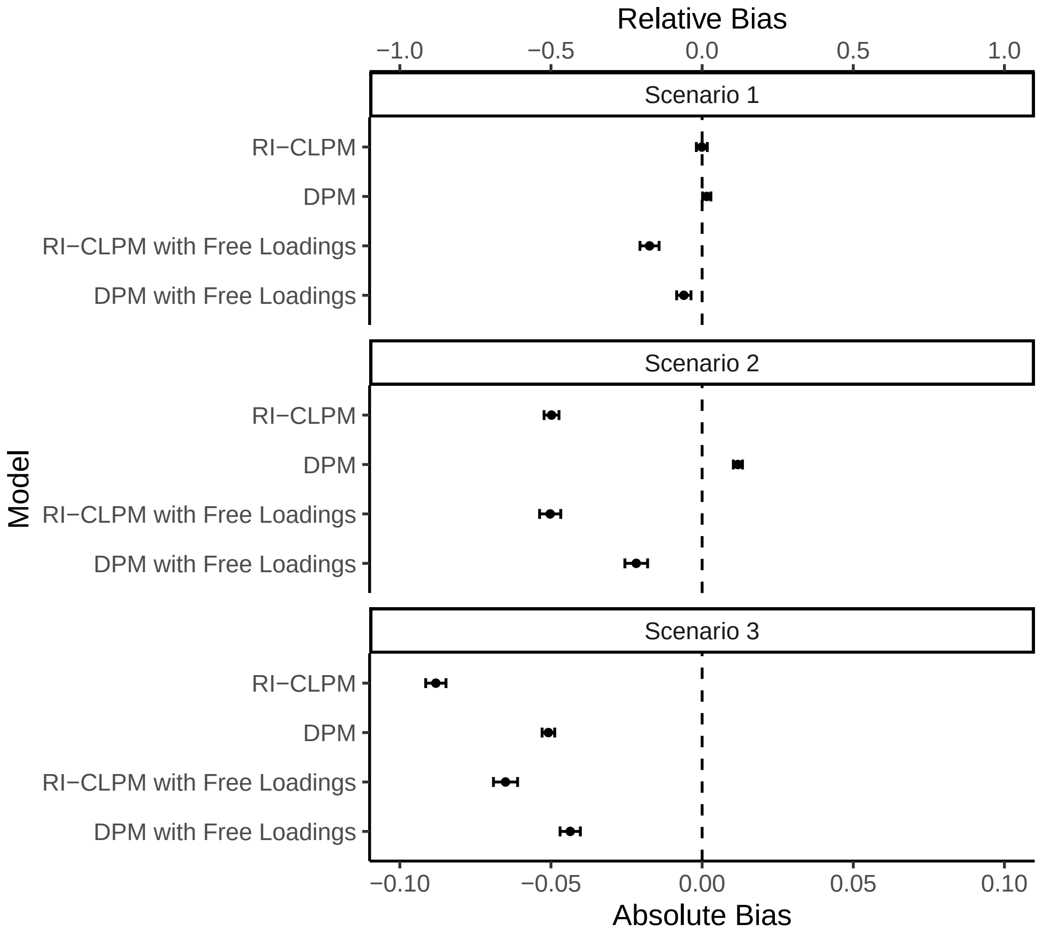


Figure 3. Biases for All Scenarios. The Dots Indicate the Bias of the Models and the Error Bars Indicate the Monte Carlo Error.

5. 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

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

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 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 estimated 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 propensity score based 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 integrating 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 emphasize 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