

The Cross-Lagged Panel Model Almost Always Provides Evidence for Causal Effects

Richard E. Lucas

Michigan State University

Author Note

Richard E. Lucas  <https://orcid.org/0000-0002-7995-3319>

I would like to acknowledge the following graduate students and undergraduate research assistants who helped develop the database of variables used in the empirical analyses: Rebekka Weidmann, Hyewon Yang, Alisar Alabdullah, Samuel A. Barans, Megan E. Denehy, Lisa Stuckman, Siya Vesikar, Jenny Warkentien, Grace Yancho, and Connor Yee. Author roles were classified using the Contributor Role Taxonomy (CRediT; <https://credit.niso.org/>) as follows: Richard E. Lucas: conceptualization, writing, formal analysis, methodology, software, supervision, visualization, editing

Correspondence concerning this article should be addressed to Richard E. Lucas, Michigan State University, Psychology Building, East Lansing, MI 48824, Email: lucasri@msu.edu

Abstract

The cross-lagged panel model is a widely used tool for testing causal effects in longitudinal data. Critiques of this model have focused on the fact that it fails to account for unobserved confounders, which can lead to spurious evidence for causal effects. Prior simulation studies suggest that the risk of spurious effects is high. In this paper, randomly selected pairs of variables that showed cross-sectional correlations between .20 and .70 in a large longitudinal study were analyzed using the cross-lagged-panel-model. Results showed that in 98% of these randomly selected pairs, evidence for significant causal effects emerged. The median effect size of these randomly selected effects was larger than the median effect from the literature. Alternative models that account for stable-trait or state variance led to fewer significant effects. These empirical results confirm recent theoretical and simulation-based critiques of the cross-lagged panel model and suggest that the model is not well-suited for testing causal effects in panel data.

Keywords: cross-lagged panel model, causal inference, longitudinal

The Cross-Lagged Panel Model Almost Always Provides Evidence for Causal Effects

Researchers across the social sciences are frequently interested in investigating causal associations. For variables that are easy to manipulate, either through carefully controlled laboratory studies or more naturalistic intervention studies, causality can be determined through experimental designs. For variables that are more difficult to directly influence, however, either due to practical or ethical concerns, experiments are not an option and alternative methods for establishing causality are required.

In such cases, researchers often turn to longitudinal data to provide evidence for causal effects ([Rohrer & Murayama, 2021](#)). One of the most common approaches to using longitudinal data in this way is to employ a simple cross-lagged panel model (CLPM, [Heise, 1970](#)). The CLPM can be used if there are at least two waves of data and each variable is measured in both waves. Time 2 measures of both variables are predicted from both Time 1 variables. If one variable measured at Time 1 predicts the other variable at Time 2, even after controlling for the Time 1 measure of the outcome, then, with some assumptions, the *cross-lagged path* can be interpreted as an estimate of the causal effect. The logic behind this interpretation is that by controlling for the Time 1 measure of the outcome (and its association with the Time 1 measure of the predictor), the effects of some (unobserved) third-variable confounds can be controlled.

Although the CLPM has been widely used in the literature, relatively serious critiques of the model have been presented. For instance, Hamaker et al. ([2015](#)) noted that if the two constructs under investigation have a stable-trait structure (where at least some variance is stable over time) and these variables are associated at this stable-trait level, then CLPM can lead to spurious evidence for causal effects. They recommended an approach that separates between-person variance from within-person variance when estimating causal effects, though alternative approaches for addressing the impact of time-invariant confounders exist (e.g., [Dishop & DeShon, 2021](#); [Zyphur et al., 2020](#)). Lucas ([2023](#)) used simulations to show that under realistic situations, the bias in estimates of lagged effects that results from stable-trait variance is quite high (relative to the size of effects found in the literature) and will frequently lead to significant

but spurious evidence for causal effects.

Although much of the discussion on the limitations of the CLPM has focused on the biasing effect of stable-trait variance, the model suffers from other problems. For instance, if measures are assessed with error—or even if there is reliable “state” variance that exists at one point and time but does not carry over to future assessments—then spurious lagged effects may emerge (Lucas et al., 2025; Westfall & Yarkoni, 2016). Importantly, Lucas et al. (2025) analyzed the longitudinal structure of hundreds of variables and found that most of these variables exhibited patterns of stability that were consistent with the existence of a substantial state component. Their analyses suggested that ignoring either the state or trait-like variance in measures leads to substantial bias in estimates and spurious evidence for causal effects.

These critiques of the CLPM suggest that significant lagged effects will emerge frequently, even when no true causal effects drive the associations. The concern is that it is too easy to find significant lagged effects and that the existence of such an effect does not provide much (if any) evidence for true causal effects. In other words, the CLPM may not provide a severe test of causal effects because evidence for these effects may frequently emerge even when the variables in question do not have any causal associations (Mayo, 2018). But just how easy is it to find significant lagged effects when using the CLPM?

The goal of this paper is to determine how often cross-lagged effects emerge, even when there is no strong theoretical reason to expect causal effects. Specifically, I examine how often significant cross-lagged effects emerge in randomly selected pairs of variables that exhibit small to moderate cross-sectional correlations. Researchers often investigate causal direction after first establishing that two variables are in fact correlated in cross-sectional studies. For instance, a recent meta-analytic review examined the association between electronic screen use and children’s socioemotional problems (Vasconcellos et al., 2025). The authors first noted that cross-sectional studies reliably show a positive association between these variables. This prompted them to investigate the causal direction of this effect in a meta-analysis of longitudinal studies. The authors used a meta-analytic CLPM and found significant lagged associations both between

screen use and socioemotional problems and between socioemotional problems and screen use, leading them to the causal conclusion that “screen use led to socioemotional problems...and socioemotional problems led to greater screen use” (p. 513). The authors noted that this causal evidence has practical implications, including support for the establishment of screen-time guidelines for children. Given the concerns about spurious effects from the CLPM, it can be useful to determine how often tests of cross-lagged effects will result in significant effects, even when there is no strong theoretical reason to expect causal effects. If such evidence is common, even among randomly selected pairs of variables, then this may suggest that the CLPM is too likely to provide evidence for causal effects.

Of course, when randomly selecting pairs of variables in an atheoretical manner, one can never know whether a significant cross-lagged association reflects a true causal effect or a spurious association. Yet at the same time, researchers like to think that their hypothesis tests are risky and that significant effects would only emerge if the theory that predicted them were true. If significant cross-lagged effects occur frequently, even in randomly selected pairs of variables, then this would suggest that examining lagged associations does not provide a particularly risky or severe test of a hypothesis.

To accomplish the goal of estimating the frequency with which significant cross-lagged effects emerge when the CLPM is used, I first computed cross-sectional correlations among almost all qualifying variables (see methods for details) in a long-running panel study. Next, pairs of variables that correlated between .20 and .70 were selected and subjected to a standard CLPM to see how common significant cross-lagged effects were. The logic behind this selection criterion is that causal associations generally result in correlations between variables, and researchers often use cross-sectional correlations as a guide when selecting variables to probe for causal associations. Results from the CLPM were then compared to alternative models. Specifically, I tested three additional models. The Random-Intercept Cross-Lagged Panel Model (RI-CLPM) is an alternative to the CLPM that, as its name suggests, includes a random intercept that accounts for the stable effects of time-invariant confounding variables ([Hamaker et al., 2015](#)). The

Autoregressive Trait State (ARTS, or factor CLPM) is a model that omits the random intercept but includes a state component at each wave (Lucas et al., 2025; Usami et al., 2019). This state component can account for measurement error and reliable occasion-specific variance, both of which can lead to upwardly biased estimates of lagged effects (Lucas et al., 2025; Westfall & Yarkoni, 2016). Finally, the Stable-Trait Autoregressive-Trait State (STARTS, Kenny & Zautra, 1995) model incorporates includes both a random intercept and a wave-specific state component.

Method

Sample

The current study uses data from the Household, Income, and Labour Dynamics in Australia (HILDA) study (Watson & Wooden, 2012). This ongoing panel study began with a nationally representative sample of households in Australia, and follow-up surveys are conducted yearly. At the time of these analyses, 22 waves of data were available. Tens of thousands of participants have been included in at least one wave of data collection, though sample sizes vary considerably depending on which waves and variables are included. A broad range of topics were assessed in the surveys, including questions about health, relationships, attitudes, income, work, and personality characteristics. Detailed information about the study and its methods can be found in the HILDA User Manual (Summerfield et al., 2024). A list of all variables included in the study can be found here: <https://hildaodd.app.unimelb.edu.au/default.aspx>). Survey organizers provide de-identified data to researchers who sign a data-use agreement, and thus, the analysis of these data do not constitute human-subjects research for which institutional review board approval is needed.

Analytic Approach

The analyses proceeded in a multistep process. First, as part of a previous project that focused on the longitudinal structure of commonly analyzed variables (Lucas et al., 2025), research assistants identified variables in the dataset that had been assessed for at least four years and that were measured with an ordered scale with at least three options. This database of variables was then examined to identify variables that were somewhat redundant (e.g., monthly

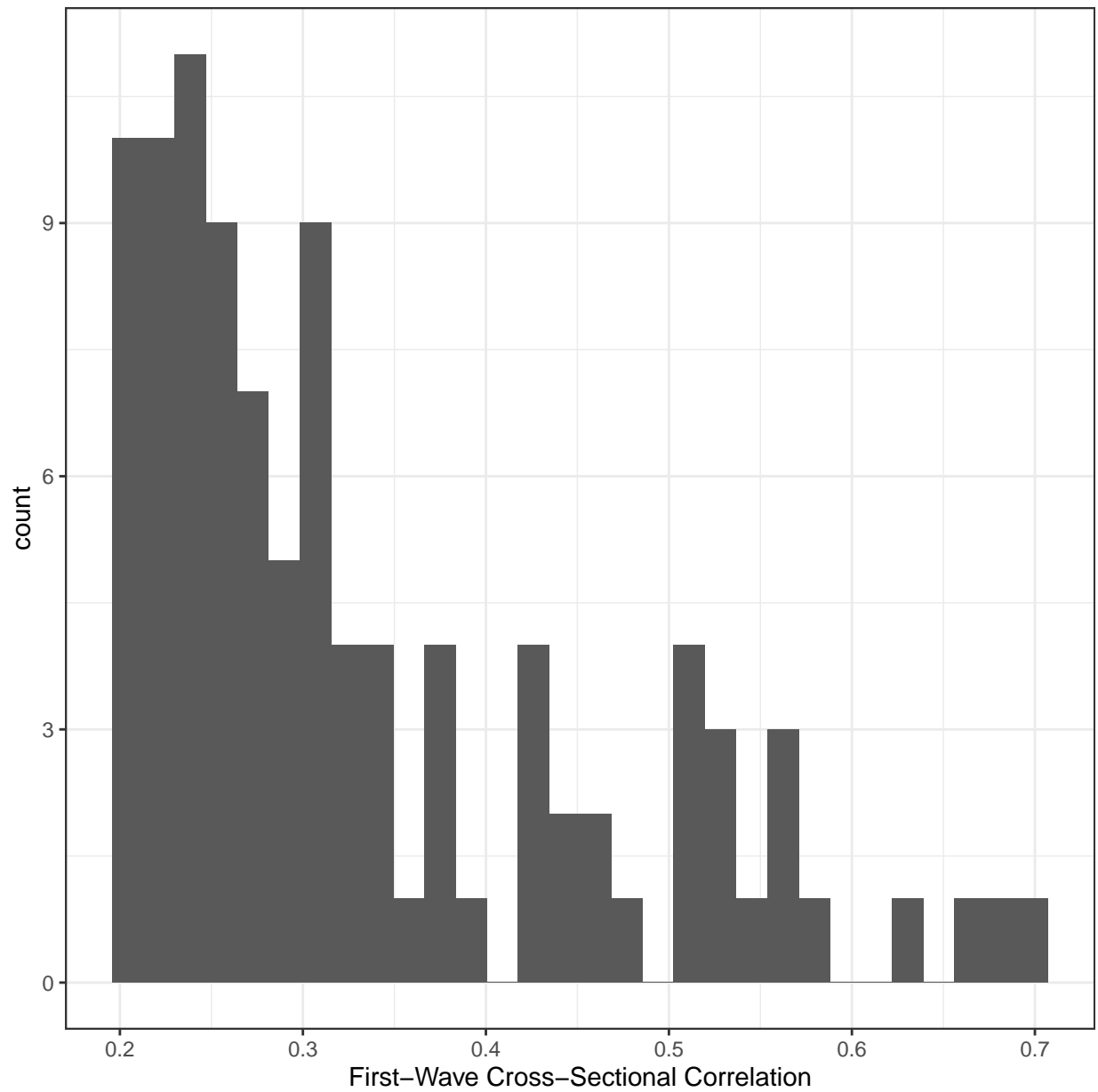
income and yearly income) and those deemed to be redundant were removed. Distributions of variables were examined and those that were highly skewed were transformed using a log function. Finally any variable with very unusual distributions (usually measures that were assessed infrequently or in small subsamples) were removed. This resulted in a list of 433 variables from a broad range of domains studied by social scientists (see supplemental materials for a complete list). Note that not all variables were assessed in all waves.

Next, the first wave in which each pair of variables was measured together was identified and the cross-sectional correlation between these variables at that wave was calculated. Variables that correlated between .2 and .7 were identified, under the assumption that researchers typically examine causal effects in variables that are known to be at least somewhat correlated. Five hundred of these variables were randomly selected for potential CLPM analysis.

CLPM analyses were then conducted for 100 of these selected pairs. Specifically, starting with the first pair from the 500 selected pairs, checks were run to ensure that both variables were assessed for at least five sequential waves (i.e., with yearly intervals between assessments) so that the cross-lagged paths would cover the same interval for all analyses. If a pair of variables did not meet this criterion, then it was skipped and the next pair from the initial list of 500 was selected. Because of the large number of analyses to be run, a maximum of 10 waves and 1,000 randomly selected rows of data were used to reduce computing time (more complex models like the STARTS can take a very long time to run when the number of variables or participants is large). This selection process also ensures that the very high power that comes from the large sample size of the parent study does not lead to high numbers of significant but negligible results. Note that although 1,000 lines of data were sampled, some of these lines will have no data for the waves or variables analyzed, which means that the sample size for each analysis was less than 1,000. Specifically, the average sample size was 660.76 with minimum and maximum samples sizes of 554 and 890, respectively. The distribution of initial cross-sectional correlations is shown in Figure 1. As can be seen, although a range of correlations between .2 and .7 is included, most cross-sectional correlations tended to be between .2 and .3.

Figure 1

Distribution of first-wave cross-sectional correlations for selected variables.



After conducting these CLPM analyses, additional models were tested using the final set of pairs of variables from the CLPM analyses. Specifically, three additional models—the RI-CLPM, the ARTS, and the STARTS—were run for each pair. Because these models are more complex, estimation difficulties were expected, especially for the STARTS. Therefore, I report how many models resulted in estimation problems for each set of analyses.

Figure 2

The Stable-Trait Autoregressive Trait State (STARTS) Model

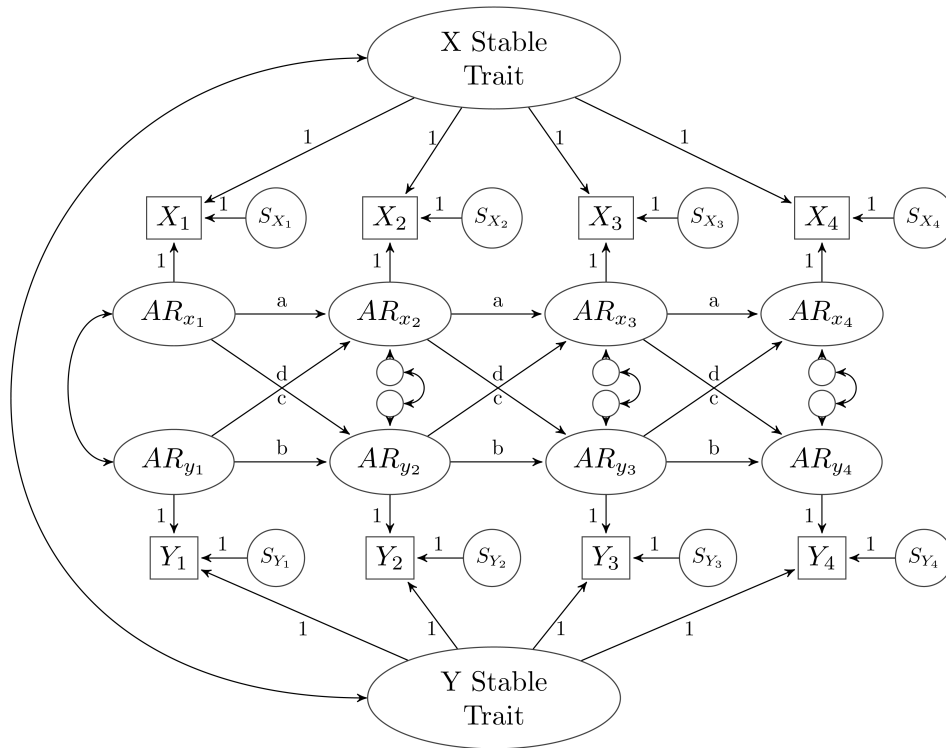


Figure 2 shows a diagram of the STARTS model. Note that this model is the most general model tested and the other three models are nested under the STARTS. Thus, they can be derived from the STARTS by constraining variance components to be 0. Specifically, by constraining the state component (the circles with S labels) to 0, the model becomes the RI-CLPM. Similarly, by constraining the stable-trait components to be 0, the model becomes the ARTS. By constraining

both of these to 0, the model becomes the CLPM. In all models, stability and cross-lagged paths were constrained to be equal across waves. In models that include a state component, full stationarity was imposed to aid estimation and convergence (Kenny & Zautra, 1995). All code for reproducing these analyses is available on the corresponding OSF site:

https://osf.io/5vr8n/?view_only=23deaeb7e30c4fad831ad6539e53cf9e.

Due to confidentiality restrictions, data cannot be shared publicly. However, researchers can sign a data-use agreement with the study organizers to access these data.

Results

Table 1

Results from CLPM, RI-CLPM, ARTS, and STARTS

Model	X Sig.	Y Sig.	Either Sig.	Both Sig.	Estimate	Problems
CLPM	0.96	0.93	0.98	0.91	0.10	0
RI-CLPM	0.40	0.42	0.61	0.22	0.04	0
ARTS	0.19	0.20	0.32	0.08	-0.01	6
STARTS	0.06	0.10	0.16	0.00	-0.01	69

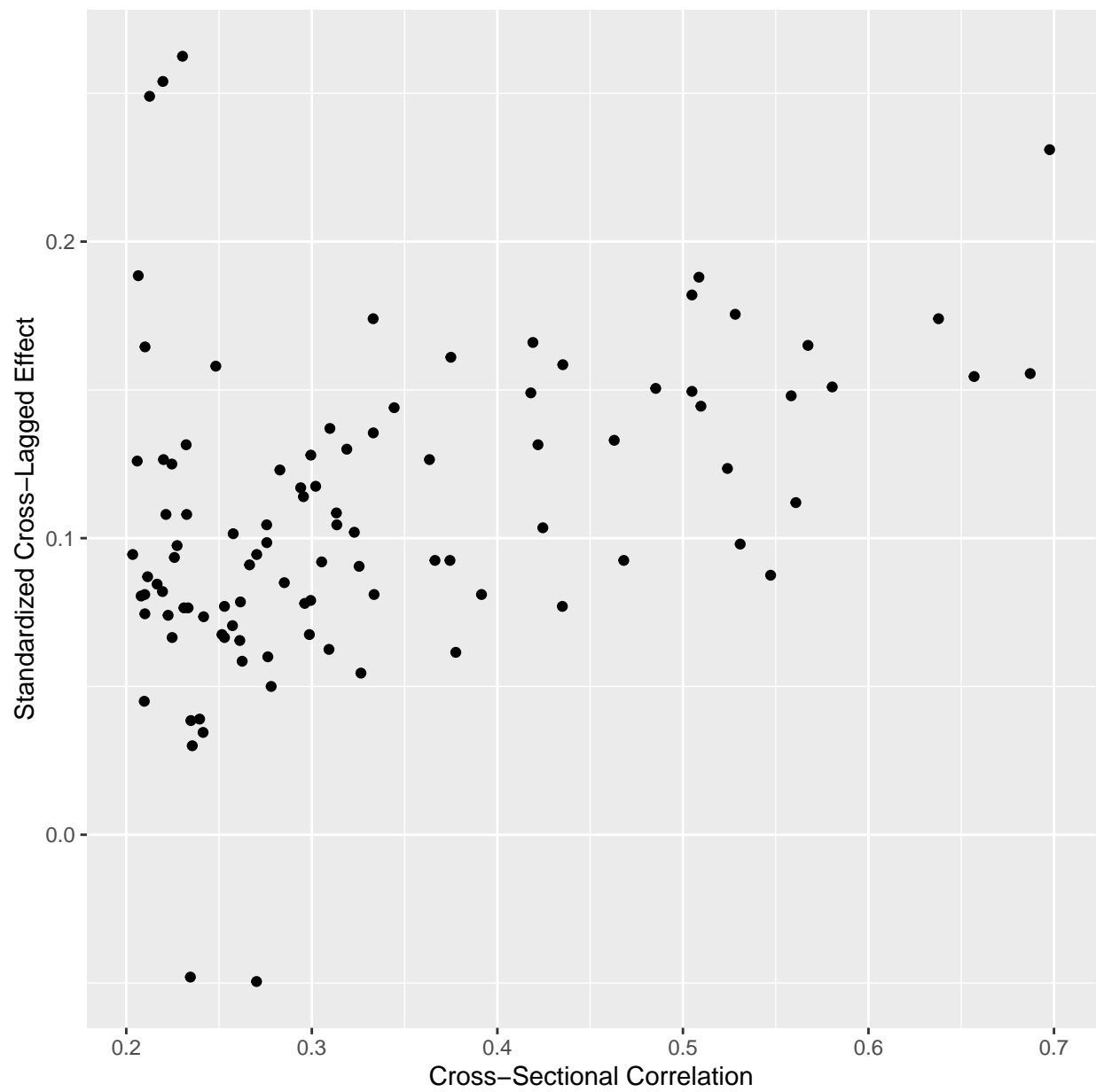
The primary results are presented in Table 1. The first two columns show the percentage of the 100 individual models where the effect of either X on Y or Y on X was significant. Perhaps more importantly, the third column shows the percentage of models in which at least one of these two effects was significant. This may be the most relevant statistic, as statistical significance is often used as a filter for publication. The cross-lagged panel model is typically used when *either* the effect of X on Y or Y on X would be meaningful and interpretable (as there are less restrictive alternative models that could be used when only one causal direction is of interest), and thus, a publishable result would be one where either causal effect was significant. The fourth column shows the percentage in which both effects were significant. As can be seen in the first row, when using the CLPM, at least one significant effect emerged in almost all models run (98%) and in the

vast majority of models, both directional effects were significant (91%). Notably, the median standardized effect size for these estimated effects (including those that were not significant) was 0.098. This can be compared to meta-analytic estimates of typical cross-lagged effects provided by Orth et al. (2022). Their meta-analytic review found that the 25th, 50th, and 75th percentiles of the distribution corresponded to effects of .03, .07, and .12, respectively. The effect size from randomly selected pairs of variables exceeds the 50th percentile of presumably theoretically motivated analyses from the literature. These results confirm concerns about the CLPM, as it suggests that to obtain a statistically significant lagged effect, researchers simply need to run a CLPM with two variables that are at least weakly correlated, and they will be almost guaranteed to find an effect (assuming a moderate sample size).

As would be expected, the size of the estimated cross-lagged paths is related to the size of the cross-sectional correlation. Specifically, the correlation between the two is 0.39. A scatter plot of the values is shown in Figure 3. Thus, although significant cross-lagged paths are extremely likely to emerge even with correlations as small as .20 (assuming a moderate sample size), this likelihood increases as the size of the cross-sectional correlation increases (also see Orth et al., 2022).

These results can be compared to those obtained when alternative models are used. The RI-CLPM has become an increasingly commonly used alternative to the CLPM in the literature. The RI-CLPM improves upon the CLPM by modeling a random intercept for both variables. This random intercept can capture stable individual differences, and thus can control for certain unobserved time-invariant confounders that are likely to bias estimates from the simple CLPM. The second row of Table 1 shows results for this model. As can be seen, fewer significant results emerged when the RI-CLPM was used as compared to the CLPM. Specifically, for any given variable, significant results were about half as likely to emerge when the RI-CLPM was used as compared to the CLPM. However, even with this more conservative model, at least one significant result was found in the majority of models tested (61%). Thus, even when using the RI-CLPM, which addresses important sources of bias in the CLPM, significant lagged effects are likely with

Figure 3



moderate sample sizes and weakly to moderately correlated variables.

The third row of Table 1 shows results for a less commonly used model for examining cross-lagged effects, the ARTS model. This model excludes the random intercept from the RI-CLPM and includes a state component to account for measurement error and reliable occasion-specific variance. Lucas et al. (2025) showed that in univariate contexts, most variables required a state component to account for their longitudinal structure, and stable-trait components (as reflected in the random intercept of the RI-CLPM) was frequently not needed to explain this structure. Moreover, Lucas et al. (2025) also showed that under simulation conditions that were designed to match the typical longitudinal structure of variables measured in panel studies, the ARTS model showed less bias than either the CLPM or RI-CLPM (though this model could also be negatively biased under certain conditions). Consistent with the conclusions from this study, lagged effects were much less likely to be significant when using the ARTS as compared to either the CLPM or RI-CLPM. Individual variables were significant in about 20% of all models, and at least one variable was significant in 32%. It is important to note that the ARTS model did result in estimation problems in 6 of the 100 models tested. These estimation problems likely result from the fact that for most variables in the HILDA, latent autoregressive stability is extremely high, often exceeding .90 (Lucas et al., 2025). When estimates are close to the boundary of what is possible, estimation problems can occur.

Finally, the fourth row of Table 1 shows results for the most complex of the models tested, the STARTS. This model combines the features of the RI-CLPM and the ARTS, including both a stable-trait component and a state component. Lucas et al. (2025) showed that in a univariate context, this model was tied with the ARTS for the best-fitting model when analyzing the longitudinal structure of variables in the HILDA. However, the STARTS is notorious for having convergence and estimation issues. Indeed, the final column of Table 1 shows that estimation problems were very common in this set of 100 analyses. Of those, 69 had problems that prevented the calculation of plausible estimates. Of the models that did successfully converge, however, very few resulted in significant lagged effects. Indeed, the number of significant effects for each

variable barely exceeded chance levels. The overall estimate was also very close to zero (and similar to that from the ARTS).

All of the above models were tested after selecting variables that have cross-sectional correlations of at least .2. The logic behind this selection procedure is that causal associations generally result in correlations, and researchers often use cross-sectional correlations as a guide when selecting variables to probe for causal associations. Yet it is also useful to consider whether significant lagged associations are likely, even among variables that exhibit very weak associations. Thus, as a final test, I repeated the CLPM analyses, this time focusing on pairs of variables that correlated between 0 and .2. Even among this set of very weakly correlated variables, significant cross-lagged effects were likely. Specifically, at least one cross-lagged effect was significant in 63% of all models tested. Thus, even among variables that exhibit very weak cross-sectional correlations, finding significant cross-lagged effects is quite likely, even when variables included in the analyses are selected randomly.

Discussion

Researchers typically strive to develop severe tests of the theories that drive their work (Mayo, 2018). Tests that are severe should provide support for a claim when that claim is true, while also allowing for the possibility of not passing the test. As Mayo notes, “before we have evidence for a claim, it must have passed an analysis that could have found it flawed” (p. xii). The results of the current study suggest that the CLPM does not provide a severe test of causal claims because it almost never fails. Almost all of the CLPMs tested resulted in significant evidence for causal effects, despite the fact that the pairs of variables in these analyses were randomly selected. There is no way to know which of the 98 significant models (if any) represent true causal effects and which represent spurious associations, but the fact that almost all pairs of variables resulted in significant effects should cause concerns about the severity of the test that the CLPM provides.

This result is particularly concerning given prior conceptual, empirical, and simulation-based work that highlights how bias in estimates of causal effects can emerge when the CLPM is used. This prior work shows that the existence of stable-trait variance or

occasion-specific state variance can lead to bias that is very high relative to the size of effects published in the literature (Lucas, 2023; Lucas et al., 2025). Moreover, most variables that social scientists study have longitudinal structures that suggest the existence of state variance, trait variance, or both. Thus, this prior work suggests that it should be quite easy to find significant cross-lagged associations even when no true causal effects exist, and the current analyses support that suggestion. As long as researchers choose variables with at least small cross-sectional associations, significant lagged effects are almost guaranteed to emerge (given a moderate sample size).

Methodologists have suggested alternative models that can address some of the concerns about the CLPM. Most notably, the RI-CLPM can address some forms of time-invariant confounding to which the simpler CLPM is susceptible (Hamaker et al., 2015). Although the RI-CLPM is not inherently more conservative than the CLMP (Lucas, 2023), this model should result in fewer significant lagged effects and smaller effect sizes if confounding at the stable-trait level is common. Consistent with this concern, for any particular effect, the RI-CLPM was less than half as likely to find significant effects than the CLPM. In addition, the typical effect size was less than half the size. This suggests that stable-trait level confounds are common, and that once these confounds are controlled with appropriate modeling strategies, cross-lagged effects become much smaller in size.

Notably, even when using the RI-CLPM to test cross-lagged effects in a set of randomly selected pairs of variables, at least one significant result emerged in more than half of the models tested. Although there is no way to tell which of these significant estimates (again, if any) reflect true causal effects, this number is arguably much higher than one would expect for pairs of variables for which no clear theoretically motivated expectation of causal effects exist. This, too, is consistent with prior empirical and simulation-based analysis of the potential limitations of the RI-CLPM. Specifically, Lucas et al. (2025) noted that the models with stable-trait components (like the RI-CLPM) and models with occasion-specific state components have different implied patterns of stability over increasingly long lags. Lucas et al.'s analysis suggested that patterns of

stability for almost all variables in the HILDA were consistent with structural models that incorporate a state component. The existence of stable-trait variance was also supported for many variables, but less so than for state components. Lucas et al. (2025) also showed that if occasion-specific variance exists, then the RI-CLPM can be positively biased, just like the CLPM. In these cases, models that include a state component—like the ARTS and STARTS—can reduce this bias (though they also have the potential to be negatively biased under some conditions). Consistent with this analysis, lagged effects were much less common when the ARTS and STARTS were used, and the average effect size was very close to zero¹.

Importantly, models with state components—especially the STARTS, which includes both state and stable-trait components—were more likely to result in estimation and convergence problems. These problems typically occur because the models, while technically identified, are often empirically underidentified (Kenny & Zautra, 1995) or because estimates often approach the boundaries of what is possible. For example, Lucas et al. (2025) showed that once state variance was taken into account, the latent autoregressive stability of most variables in year-long panel studies like the HILDA was very high, often exceeding .90. With stabilities this high, residual variances can be very small, and by adding cross-lagged paths, inadmissible solutions may often result. Thus, while models that include state components may provide better descriptions of the underlying longitudinal structure of variables in panel studies, frequent estimation problems may mean that well-justified models may not be testable in certain situations.

So how should researchers proceed when testing lagged effects? The current analyses add to previous calls for the avoidance of the simple lag-1 CLPM by highlighting the lack of severity of this commonly used test. If significant evidence for causal effects is usually found in randomly selected pairs of variables, then this means that finding significant effects in theoretically derived analyses is probably not strong evidence for the underlying theory that motivated this test. Given the concerns raised about the impact of state and trait variance, alternative models should be

¹ Note that this means that many of the significant effects that did emerge in the ARTS and STARTS were negative, even though all variables examined in this analysis were positively correlated.

considered. Specifically, given the widespread evidence for the existence of both stable-trait and state variance in the types of variables that social scientists study, models that include these components should be considered and tested. In addition, sources of local misfit in these models should be examined when these models are estimated, as this misfit can provide evidence for assumptions that may be violated (see [Lucas et al., 2025](#) for more detailed guidance). If results diverge across models—for instance, if the RI-CLPM results in a significant lagged effect and the ARTS does not—then strong theoretical (and ideally, empirical) justification for the assumptions underlying the RI-CLPM need to be provided before accepting this significant result as evidence for the causal effect.

Limitations

The analyses presented here suggest that it is too easy to find significant cross-lagged effects when relying on the simple lag-1 CLPM. However, these results may depend on the specific design features of this study. Most notably, all variables were assessed using self-report methods and the interval between assessments was always one year. Because the bias in estimates from the CLPM depends on the specific pattern of stability coefficients in the data, anything that systematically affects these stability coefficients can affect the performance of the model. It is possible that other types of measures or different interval lengths will result in different conclusions. Given that data like those included in this study are common targets of longitudinal, causally focused analyses, however, these results should be concerning even if they do not generalize to all longitudinal analyses.

Conclusion

Existing conceptual, empirical, and simulation-based research suggests that the simple lag-1 CLPM is very likely to provide spurious evidence for causal effects. The current analyses support and extend these existing critiques of the CLPM by illustrating how frequently evidence for these causal effects emerges even in randomly selected pairs of variables. These analyses show that in weakly to moderately correlated variables, finding evidence for causal effects is almost guaranteed, as long as sample sizes are moderate. In addition, finding support is also quite likely

among pairs of variables that are very weakly correlated. Thus, these results highlight that the CLPM does not provide a severe test of causal associations and significant lagged effects should be interpreted very skeptically unless a comprehensive set of alternative models are tested and ruled out.

References

- Dishop, C. R., & DeShon, R. P. (2021). A tutorial on Bollen and Brand's approach to modeling dynamics while attending to dynamic panel bias. *Psychological Methods*.
<https://doi.org/10.1037/met0000333>
- 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/f67cvh>
- Heise, D. R. (1970). Causal Inference from Panel Data. *Sociological Methodology*, 2, 3–27.
<https://doi.org/10.2307/270780>
- Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, 63(1), 52–59.
<https://doi.org/10.1037/0022-006X.63.1.52>
- Lucas, R. E. (2023). Why the Cross-Lagged Panel Model Is Almost Never the Right Choice. *Advances in Methods and Practices in Psychological Science*, 6(1).
<https://doi.org/10.1177/25152459231158378>
- Lucas, R. E., Weidmann, R., & Yang, H. (2025). *Typical Patterns of Stability in Longitudinal Data: Implications for Model Choice*. Michigan State University.
- Mayo, D. G. (2018). *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge University Press.
- Orth, U., Meier, L., L., Buhler, J. L., Dapp, L. C., Krauss, S., Messerli, D., & Robins, R. W. (2022). Effect Size Guidelines for Cross-Lagged Effects. *Psychological Methods*.
- Rohrer, J. M., & Murayama, K. (2021). These are not the effects you are looking for: Causality and the within-/between-person distinction in longitudinal data analysis. *Advances in Methods and Practices in Psychological Science*, 6(1). <https://doi.org/10.1177/25152459221140842>
- Summerfield, M., Garrard, B., Nesa, M., Kamath, R., Macalalad, N., Watson, N., Wilkins, R., & Wooden, M. (2024). *HILDA User Manual – Release 23*. Melbourne Institute: Applied Economic and Social Research, University of Melbourne.
- 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/gf4fqx>

- Vasconcellos, R. P., Sanders, T., Lonsdale, C., Parker, P., Conigrave, J., Tang, S., Salmela-Aro, K., Vasconcellos, D., Wilhite, K., Tremaine, E., Booker, B., & Noetel, M. (2025). Electronic Screen Use and Children's Socioemotional Problems: A Systematic Review and Meta-Analysis of Longitudinal Studies. *Psychological Bulletin*, 151(5), 513–543.
- Watson, N., & Wooden, M. (2012). The HILDA Survey: A Case Study in the Design and Development of a Successful Household Panel Study. *Longitudinal and Life Course Studies*, 3(3), 369–381.
- Westfall, J., & Yarkoni, T. (2016). Statistically Controlling for Confounding Constructs Is Harder than You Think. *PLOS ONE*, 11(3), e0152719. <https://doi.org/10.1371/journal.pone.0152719>
- 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/gf8rt5>