MULTIVARIATE BEHAVIORAL RESEARCH
2025, VOL. 60, NO. 5, 973–989

Ruling out Latent Time-Varying Confounders in Two-Variable Multi-Wave Studies

David A. Kenny[a] and D. Betsy McCoach[b]

[a]Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut, United States; [b]Department of Educational Psychology, University of Connecticut, Storrs, Connecticut, United States

ABSTRACT
There has been considerable interest in estimating causal cross-lagged effects in two-variable, multi-wave designs. However, there does not currently exist a strategy for ruling out unmeasured time-varying covariates that may act as confounders. In this paper, we propose a new strategy for testing whether an unmeasured time-varying covariate explains all covariation between the two "causal" variables in the data. That model, called the *Latent Time-Varying Covariate* (LTVC) model, can be tested with observations for two variables assessed across three or more measurement waves. If the LTVC model fits well, then a time-varying covariate can explain the covariance structure, which undermines the plausibility of causal cross-lagged effects. Although the LTVC model tends to be underpowered when causal cross-lagged effects are small, if testable stationarity constraints on the LTVC model are imposed, adequate power can be achieved. We illustrate the LTVC approach with three examples from the literature. Additionally, we introduce the LTVC-CLPM model, which is identified given strong stationarity constraints. Also considered are multivariate and multi-factor models, the inclusion of measured time-invariant covariates in model, measurement of the stability of the LTVC, and the lag-lead model. These methods allow researchers to probe the assumption that an unmeasured time-varying confounder is the source of all the *X-Y* covariation. Our methods help researchers to rule out certain forms of confounding in two-variable, multi-wave designs.

KEYWORDS
Longitudinal Data; Structural Equation Modeling; Causal Inference; Latent Confounder

There has been considerable interest in methods to infer causal cross-lagged effects in the multi-wave, two-variable (2VMW) designs (Usami et al., 2019; Zyphur et al., 2020). The two variables, denoted here as X and Y, are measured at three or more times, and persons are measured at the same times. Different structural models have been proposed to measure the causal effect of X_t on Y_{t+1} and the effect of Y_t on X_{t+1}. Usami et al. (2019) discussed seven such models, and Zyphur et al. (2020) discuss three other models. All of these models presume a theory of causality (Huang & Yuan, 2017; Robins & Hernán, 2009) that requires a series of assumptions. A key assumption, sometimes called *exchangeability* (Greenland & Robins, 1986) specifies that there does not exist a confounder, a variable that causes both the putative cause and the outcome. Usami et al. (2019) state

> (T)he crucial point for choosing the cross-lagged model is whether the model provides adequate control of time-varying and time-invariant confounders (p. 652).

They also state

> However, if there are omitted time-varying variables … , these are not properly accounted for in these models, and this is likely to cause biased results (p. 643).

Grosz et al. (2020) have urged non-experimental researchers to talk openly about causal assumptions and their desire to make causal inferences. They recommend that researchers clearly articulate the assumptions underlying their analyses and identify potential confounders and plausible threats to the internal validity of the study. A central difficult issue in causal inference from non-experimental data is ruling out the existence of an unmeasured confounding variable, i.e., a variable that causes both the causal variable and its presumed outcome. The current article outlines an approach for assessing the plausibility that an unmeasured time-varying confounder may *completely* explain the causal cross-lagged paths with 2VMW data.

To date there does not exist a general strategy for handling confounding due to a latent *time-varying* covariate. In this article, we suggest an innovative approach to manage this problem. The guiding idea is to model a latent time-varying covariate that completely explains the *X*-*Y* covariation as a plausible rival hypothesis (Shadish et al., 2002) for the estimation of causal effects from 2VMW data. That is, we propose an analysis that potentially rules out the very plausible rival hypothesis that *all* the *X*-*Y* covariation is explained by a latent time-varying confounder. We call this model *LTVC*.

In LTVC, there is a single latent variable, *C*, that explains all the *X*-*Y* covariation. This latent variable changes over time, and no constraints are placed on how *C* changes over time. The model equations are as follows:

$$X_{it} = f_t C_{it} + E_{it} \quad (1)$$

$$Y_{it} = g_t C_{it} + F_{it} \quad (2)$$

where C_t is a standardized latent variable, whose over-time covariance matrix is free to vary.[1] We do not constrain the loadings on *C* to be time invariant. All residual variables *E* (for *X*) are correlated with each other, and all residual variables *F* (for *Y*) are correlated with each other; however, neither the *E* residuals are correlated with the *F* residuals, and nor are the *E* or *F* residuals are correlated with *C*. A three-wave illustration of this model is presented in Figure 1.

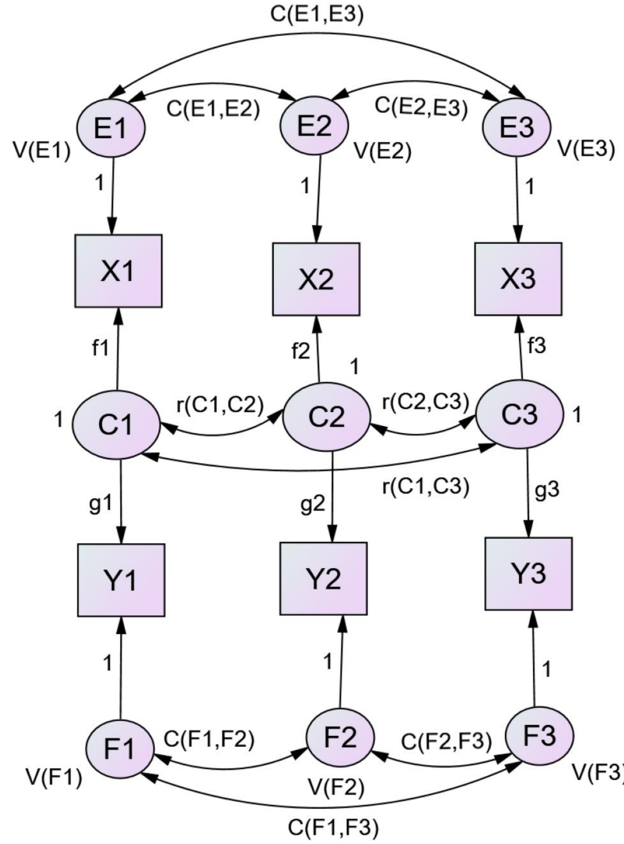

Figure 1. Model of a latent time-varying confounder (LTVC). Note: *C* entirely explains the covariation between the measured variables *X* and *Y*, whose loadings on *C* are *f* and *g* and whose residuals, *E* and *F*, are correlated across time; *V*() symbolizes variance, *C*() covariance, and *r*() correlation.

The LTVC model has the following set of assumptions:

1. The standard linear model assumptions of normality of latent variables and residuals, homogeneity of variance, and independence of residuals.
2. A latent variable, *C*, simultaneously and completely explains the covariation between *X* and *Y* at each time.
3. The residuals for *X* are uncorrelated with the residuals of *Y* and both sets of residuals are uncorrelated with all the *C* variables.
4. There are no paths from X_t to X_{t+1} and from Y_t to Y_{t+1}.

Assumptions 2, 3, and 4 are discussed later in the article. Note too that CLPM or any other 2VMW models are not alternative models to LTVC. That is, having the LTVC be a poor fitting model does not imply any particular 2VMW model would have non-zero causal paths. We presume that if there were a causal effect between *X* and *Y* that the causation would not be instantaneous. If the casual lag were zero, then the LTVC would be a good-fitting model, with the cause having a standardized loading of one.

The LTVC model is under-identified because we are unable to recover unique estimates for all the model parameters. (As discussed in the "*Understanding the structure of change in a latent time-varying covariate*" section later in the article, the squared autocorrelations of the latent variable are identified.) The fundamental reason for this lack of identification is there are only two indicators of the latent factor at each time, and there is no third indicator that can be "borrowed" that has uncorrelated errors of measurement with those two indicators. Nonetheless, with three or more waves of measurement, the model imposes constraints on the variance-covariance structure for the data. The LTVC model with three waves has one over-identifying restriction, which is:

$$\rho_{X_1Y_2}\rho_{X_3Y_1}\rho_{X_2Y_3} = \rho_{X_2Y_1}\rho_{X_1Y_3}\rho_{X_3Y_2}, \quad (3)$$

[1]For all structural equations in this article, we omit the mean or intercept from the equations as all exogenous variables in the model are presumed to have means of zero.

with both terms equaling $f_1g_1f_2g_2f_3g_3\rho_{C_1C_2}\rho_{C_2C_3}\rho_{C_1C_3}$ where f_1 to f_3 and g_1 to g_3 are defined in Figure 1 and ρ_{C1C2}, ρ_{C2C3}, and ρ_{C1C3} are population parameters that replace the sample estimates r_{C1C2}, r_{C2C3}, and r_{C1C3}, respectively. The general equation for the over-identifying restriction with 3 or more waves is

$$\rho_{X_tY_{t+p}}\rho_{X_{t+q}Y_t}\rho_{X_{t+p}Y_{t+q}} = \rho_{X_{t+p}Y_t}\rho_{X_tY_{t+q}}\rho_{X_{t+q}Y_{t+p}} \tag{4}$$

where p and q are integers and $p < q$. In general, for studies with T waves, there are $(T-1)(T-2)(T-3)/6$ constraints like Equation 4, of which $(T-1)(T-2)/2$ are independent. (Throughout this article, we use T to denote the number of waves.) The degrees of freedom of the LTVC model are $(T-1)(T-2)/2$, the number of free overidentifying restrictions.

Even though the LTVC cannot be fully estimated, the model can be tested and shown to be inconsistent with the data. Thus, it is possible to rule out confounding as the sole explanation of cross-variable covariation without specifying exactly how the time-varying covariate changes over time. To some readers, it may seem unusual to estimate a model that is not identified. Normally, the goal of an SEM analysis is to estimate the model's parameters. However, an identified model is not the goal here. Rather, the goal is to rule out the possibility that an unobserved time-varying covariate is sufficient to explain all of the *X*-*Y* covariation. Given this purpose, the relative sizes of the factor loadings on *X* and *Y* (*f* and *g*) are of no theoretical interest. Rather, the key question is whether the data are inconsistent with the model. There are different strategies to determine if the data are inconsistent with the model. The traditional way is to test the null hypothesis that the proposed model is the true model using a chi-square test. Alternatively, we might use a fit statistic to measure how well the proposed model is consistent with the data, e.g., RMSEA or SRMR.

The estimation of under-identified models is not unprecedented in SEM. For instance, a two-factor model with all measures loading on both factors is not identified; nonetheless it is possible to test whether the covariance structure is consistent with the model. Another example is a post-treatment confounder within a mediation analysis discussed by Moerkerke et al. (2015). Although not all the model's parameters can be uniquely estimated, it is still possible to estimate the direct and indirect effects, the key parameters of interest.

Given this lack of identification, estimating the LTVC by an SEM program presents difficulties. The SEM program *Amos* (Arbuckle, 2021) can estimate the model with the correct degrees of freedom if the user chooses the option "Estimate Under-identified Models." *Mplus* (Muthén & Muthén, 1998-2017) typically provides estimates, but reports that the model is not identified and gives no chi-square test or fit statistics. To obtain the chi-square and fit statistics, the model can be re-estimated, fixing one parameter to its estimated value from the initial run. We suggest fixing the estimate of the residual variance of X_1 value from the first run and re-estimating the model. The LTVC model with four or more waves produces fit statistics in *lavaan* (Rosseel, 2012), but yields the wrong degrees of freedom, always one too few. With three waves, *lavaan* reports the model is not identified, but does provide estimates. As with *Mplus*, we suggest taking the estimate of the residual variance of X_1 value from the first run, fixing the parameter to that value, and re-estimating the model, which yields the correct degrees of freedom for the model.

To prevent Heywood cases, we suggest constraining the residual variances for *X* and *Y* to be non-negative. Our reasoning is as follows: Imagine that CLPM is the correct model with large positive paths from X_t to Y_{t+1}. If we were to fit an LTVC model to this data structure, the loadings on *Y* would increase over time, but the loadings on *X* would decrease. If the *X* to *Y* path is large enough in the true model, when estimating LTVC, we would obtain Heywood cases for *Y* on later waves and for *X* on earlier waves. Allowing Heywood cases would enable impossibly large loadings.

Power analyses of the LTVC model

It is important to examine whether the LTVC model (see Equations 1 and 2, as well as Figure 1) can explain the pattern of *X*-*Y* covariances observed in the data. If the LTVC model can explain the pattern of covariances, then we cannot rule out the alternative hypothesis that an omitted latent time varying confounder completely explains that covariation, which would undermine the plausibility of any *X*-*Y* causal effects. Alternatively, if the LTVC model is not a good-fitting model, we can rule out the possibility that omitted latent time-varying confounder can explain all the *X*-*Y* covariation.

The recommendation here is to estimate the LTVC model when conducting a 2VMW causal model to rule out the possibility of a latent confounder as an alternative explanation for the pattern of observed covariances. For this strategy to be viable, there would need to be sufficient power to reject the LTVC. If the power to reject the LTVC were low, then the test would not be diagnostic. Therefore, we undertook a series of power analyses in which the true model was a causal model (e.g., CLPM) model, and the estimated model was an LTVC model. We then evaluated whether we had adequate power to reject the LTVC model under a variety of scenarios in which we varied the stabilities, the causal cross-lagged paths, the cross-sectional correlations, the sample size, and the number of waves. Because the LTVC model implies a non-trivial cross-sectional correlation between X and Y, we presume that the lowest value of that correlation is ±0.2.

We assumed that a given causal model (e.g., CLPM) is the true model, and we then generated a variance-covariance matrix using the parameters for that model. We then conducted a power analysis to determine whether the LTVC model would be rejected under those conditions. We also computed the RMSEA for the LTVC model. We interpreted power values of 0.80 and RMSEA values of 0.08 or above as indicative of a poor-fitting model.

We computed power, not by simulation, but by using a method developed by Satorra and Saris (1985; see Feng & Hancock, 2023). To compute power, we first determined the population variance-covariance matrix implied by a causal model (e.g., CLPM). We then used that covariance matrix to compute the non-centrality parameter (NCP). Using the NCP, df, and the proposed sample size, we determined the probability of a statistically significant chi-square statistic to estimate the power to reject the LTVC model when the generating model was a CLPM model. Using the NCP and the degrees of freedom (df), we also computed RMSEA.

The basic CLPM equations to generate the data are:

$$X_{it} = aX_{i(t-1)} + dY_{i(t-1)} + E_{it} \quad (5)$$

$$Y_{it} = bY_{i(t-1)} + cX_{i(t-1)} + F_{it} \quad (6)$$

Note that E and F are uncorrelated with prior values of X and Y but may be contemporaneously correlated. For simplicity, the total variances of X and Y were fixed to be one across all waves. In addition, the cross-sectional correlations between X and Y were set to the same value across all waves of data collection. To accomplish this, we varied the magnitude of the covariance of the residual terms in Equations 5 and 6.

We computed the power to reject the model with an alpha of 0.05 for sample sizes of 250, 500, 750, and 1000 for the LTVC. We varied several factors, including the number of equally-spaced waves (3, 4, and 5), the cross-lagged paths (c: 0.1, 0.2, 0.3, 0.4, 0.5 and d: 0, 0.1, 0.2, 0.3, 0.4), the stability paths for X and Y (a and b: 0.4, 0.5, 0.6), the cross-sectional correlation between X and Y at time 1 (0.2, 0.3, 0.4, 0.5). We wrote the R program, Power_LTVC, to compute power for all of these cases, which is contained in Supplementary Material: Power_LTVC.

Table 1. Theoretical power and RMSEA estimates for CLPM estimated by LTVC with a lagged path from X to Y (c: 0.1 to 0.5), 3 to 5 waves (T), no path from Y to X ($d = 0$), and both stabilities (a and b), and the X-Y cross-sectional correlation of 0.5, and an alpha of 0.05, with bold values indicating power greater than 0.800 or RMSEA greater than 0.80.

Path	Waves	N				
c	T	250	500	750	1000	RMSEA
.1	3	.055	.061	.066	.071	0.014
.1	4	.056	.061	.067	.073	0.011
.1	5	.056	.061	.067	.074	0.010
.2	3	.084	.120	.156	.192	0.034
.2	4	.171	.314	.457	.585	0.048
.2	5	.425	.765	**.927**	**.981**	0.065
.3	3	.215	.379	.525	.647	0.074
.3	4	**.814**	**.987**	**.999**	**1.000**	**0.123**
.3	5	**.985**	**1.000**	**1.000**	**1.000**	**0.133**
.4	3	.725	**.951**	**.993**	**.999**	**0.162**
.4	4	**.999**	**1.000**	**1.000**	**1.000**	**0.204**
.4	5	**1.000**	**1.000**	**1.000**	**1.000**	**0.206**
.5	3	**.984**	**1.000**	**1.000**	**1.000**	**0.260**
.5	4	**1.000**	**1.000**	**1.000**	**1.000**	**0.288**
.5	5	**1.000**	**1.000**	**1.000**	**1.000**	**0.292**

Table 1 contains the power and the RMSEA values[2] when the population values for a CLPM model in which there the autoregressive (stability) paths for X (path a) and Y (path b) were both .5. The unidirectional cross-lagged path from X to Y, c, equaled 0.1, 0.2, 0.3, 0.4, or 0.5, whereas the cross-lagged path from Y to X, d, was fixed to zero. The cross-sectional correlation between X and Y was fixed to 0.5.

None of the power estimates exceeded 0.80 when the cross-lag path or c was .1. With a unidirectional crosslag path of 0.20, we achieved acceptable power for 5-wave studies with at least 750 cases. With a unidirectional crosslag path of 0.30, acceptable power was achieved for 4 and 5-wave studies, even with sample sizes as low as 250. However, for 3-wave studies and $c = 0.3$, power was low, ranging from 0.22 with 250 cases to 0.65 with 1000 cases, and RMSEA was 0.074. With a unidirectional crosslag path of 0.4, the 3-wave 250 case had power of .73. For all other conditions, power exceeded 0.95 when the unidirectional crosslag path was .40. RMSEA was also above 0.16 when the crosslag path was .40. For an effect of 0.5, power was greater than 0.98 for all conditions, and RMSEA exceeded 0.25.

[2]More detailed results from the power analyses using Power_LTVC are available at osf.io/k7mcj.

Stationarity of parameters

The LTVC is a very general model: The loadings can take on any value and change over time and the covariance structure of the latent variable can take on any form. However, the power to rule out a time-varying covariate as an alternative explanation of the causal relationship between X and Y can be low, especially with small samples and small effect sizes. One way to increase power is to assume stationarity, which is commonly done in the modeling of 2VMW data (Usami et al., 2019). To impose stationarity, the same parameter at different waves equals the same value. Stationarity assumptions reduce the complexity of the model and so increase the precision in the estimation of model parameters and power in rejecting the fit of the model. Conceptually, stationary models also provide advantages for modeling explanatory processes and prediction. Anytime applied researchers use model parameters to make predictions, they implicitly assume stationarity—that the causal effect of the variables does not change across time.

We consider two extensions of LTVC that impose stationarity constraints. To assist in the comprehension, Figure 2 illustrates the set of T-by-T covariances between X and Y. The first model, LTVC-S, presumes that the common factor loadings of X and Y do not change over time and that the factor variance of the common factor, C, also does not change over time. Adding these stationarity assumptions to the LTVC model has two important consequences. First, the model implies equal cross-sectional covariances: $C(X_1,Y_1) = C(X_2,Y_2) = C(X_3,Y_3)$, which are denoted as *csc* in the main diagonal in Figure 2. Additionally, the model implies equal cross-lagged covariances. With three waves, the constraints are $C(X_1,Y_2) = C(X_2,Y_1)$, $C(X_2,Y_3) = C(X_3,Y_2)$, and $C(X_1,Y_3) = C(X_3,Y_1)$, that is $lx_1 = ly_1$ and $lx_2 = ly_2$ in terms of the symbols in Figure 2, but the two lx_1 and the two ly_1 covariances are not forced to be equal to each other.

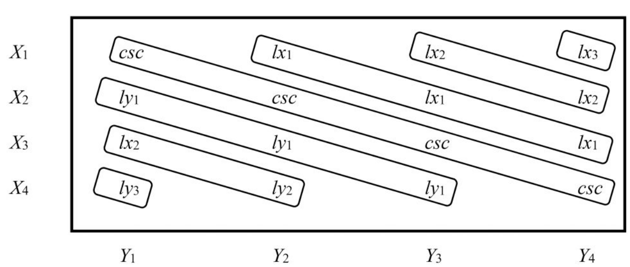

Figure 2. Stationarity constraints on the cross-covariance matrix with four waves.
Note: Cross-sectional covariances (*csc*) and the minor diagonals being the lagged (*l*) covariances with either *x* or *y* as the earlier measured variable and the numeric subscript (1 to 3) being the lag length.

The LTVC-S model, like the LTVC, is not identified because there are only two indicators per factor. However, the covariance matrix implied by the LTVC-S can be uniquely estimated using an SEM program. The equivalent model has no common factors, but instead, there are constraints on X-Y covariances: T equal cross-sectional covariances and $T(T-1)/2$ equal cross-lagged covariances. Although the model, expressed in terms of factor loadings and factor covariances, is not identified, an equivalent model of the LTVC-S implied covariances is identified.

Table 2 presents key information about the LTVC and LTVC-S models. LTVC has no constraints on covariances, whereas LTVC-S has T equal cross-sectional covariances and $T(T-1)/2$ pairs of equal cross-lagged covariances. To test the stationarity assumption in LTVC-S, we create a new model, S-Only that has equal cross-sectional covariances (the main diagonal in Figure 2), but not equal cross-lagged covariances. To test the equality of the cross-lagged covariances of LTVC-S model, we compute the chi-square difference between the LTVC-S model and the S-only model. If the LTVC-S model fits as well as the S-Only model (i.e., chi-square difference between the LTVC-S model and the S-Only model is not statistically significant), then we can conclude a latent time-varying covariate is sufficient to explain the covariation in our data. Instead of using a statistical test, one might alternatively examine the RMSEA, based on the chi-square difference. We adopt the approach of Savalei et al. (2024) of computing the RMSEA using the difference in chi squares, which they denote as RMSEA_D.

We can make a further stationarity assumption that the processes that govern change are time invariant. For instance, the X-Y covariances are the same between waves 1 and 2 as between wave 2 and 3. Such an assumption implies equality of all covariances of the same lag length and presumes the distance between adjacent waves is the same. We call this model LTVC-2S, as it imposes a second set of stationarity constraints. LTVC-2S imposes all the same constraints as LTVC-S and additionally constrains all crosslag covariances of equal lag length to be equal, e.g., $C(X_1,Y_2) = C(X_2,Y_3) = C(X_3,Y_4)$. As illustrated in Figure 2, all the three lx_1 covariances are set equal and equal to the ly_1 covariances. Similarly, the two lx_2 covariances are equal and they equal the two ly_2 covariances. Finally, the lx_3 covariance equal the ly_3 covariance. In a study with T waves, there are $(T-1)(T-2)$

Table 2. Typology of LTVC models.[a]

Model Name	Stationarity	LTVC Over-Identifying Restrictions	Model df	Stationarity *df*	LTVC *df*
LTVC	None	Equation 4	$(T-2)(T-3)/2$	0	$(T-2)(T-3)/2$
S-Only	Equal Cross-Sectional Covariances	None	$T-1$	$T-1$	0
LTVC-S	Equal Cross-Sectional Covariances	Equal Cross-Lagged Covariances	$(T-1)(T+2)/2$	$T-1$	$(T-1)(T-2)/2$
2S-Only	S-Only and Equal Same-Length Lagged Covariances	None	$(T-1)^2$	$(T-1)^2$	0
LTVC-2S	S-Only and Equal Same-Length Lagged Covariances	Equal Cross-Lagged Covariances of the Same Lag Length	$T(T-1)$	$(T-1)^2$	$(T-1)$

[a]All degrees of freedom (*df*) assume that the lag intervals between all adjacent waves are the same.
Note: LTVC, LTVC-S, and LTVC-2S are models that presume a latent time-varying covariate explains the *X-Y* covariation, but they differ in their stationarity assumptions (see column 2). The S-Only and 2S-Only models are used to test those stationarity assumptions.

such constraints. As with the LTVC-S, the LTVC-2S model's covariances are identified.

To test the assumption of time-invariant change, we estimate the 2S-Only model that contains equal same-length lagged covariances, separately for X and Y, as well as equal cross-sectional covariances, but no equal cross-lagged covariances. In Figure 2, the lx_1 covariances and lx_2 covariances are set equal to each other, as well as the ly_1 and ly_2 covariances, but the lx and ly covariances are not be set equal to each other. To test this second stationarity assumption, we compare the fit of the 2S-Only model to the S-Only model. If the 2S-Only model fits as well as the S-Only model, we can impose the second set of stationarity constraints. Then we evaluate the equality of cross-lagged covariances allowing for equality of over-time processes by comparing the fit of LTVC-2S model to the fit of 2S-Only model, which has $T-1$ degrees of freedom.

Power and $RMSEA_D$ for LTVC-S and LTVC-2S models

Earlier we showed that the power and RMSEA to reject an LTVC model was acceptable only when the causal path was moderate to large and sample size was large. Here we examine whether power and $RMSEA_D$ increase when stationarity constraints are imposed. We examine both CLPM and RI-CLPM, when each is tested using LTVC-S and LTVC-2S. To simplify the presentation, we primarily focus on $RMSEA_D$, using 0.08 for the cutoff for a poor-fitting model.

CLPM

In general, we found that power and $RMSEA_D$ were greater for the LTVC-S model than for the LTVC model, and power was greater for the LTVC-2S model than for the LTVC-S model, though improvements when moving from the LTVC-S model to the LTVC-2S model were generally not as dramatic as the improvements when moving from the LTVC model to the LTVC-S model.

Table 3 contains bidirectional models with equal stabilities and gives the $RMSEA_D$ for the model. The path from X to Y, c, varied from 0.1 to 0.5, the path for Y to X, d, from 0 to 0.4, and the cross-sectional correlations were set to .5. Importantly, the a and b stabilities were set equal to 0.5, an assumption that is later relaxed.

When the two causal paths, c and d, were equal, $RMSEA_D$ was always zero. As the difference between the two paths increased by the same amount, $RMSEA_D$ increased as the paths' sizes increased. For instance, consider a 3-wave dataset estimated using LTVC-S. $RMSEA_D$ would be 0.077 when $c = 0.1$ and $d = 0$, but when $c = 0.5$ and $d = 0.4$, $RMSEA_D$ increases to 0.125. Thus, for a fixed value of c—d, the bigger the values of c and d, the greater the value of $RMSEA_D$. Additionally, when the absolute difference between c and d is 0.1, the smallest value of $RMSEA_D$ is 0.060 and when the c — d difference is 0.2, it is .122. With a difference in coefficients of just 0.2, we should be able to rule out the LTVC-S or LTVC-2S models.

Once we allow for unequal autoregressive paths, a and b, the story becomes more complicated. The ability to reject LTVC-S and LTVC-2S depends on the difference between the cross-lagged covariances. When the variables are all standardized and stationary, the crosslag difference equals $c - d + r(b - a)$, where r is the cross-sectional correlation. As we saw in Table 3, the difference, $c - d$, in lagged coefficients affects $RMSEA_D$. But $RMSEA_D$ also depends on the difference between the two autoregressive paths. Assuming $c - d$ is positive, as b gets larger than a, $RMSEA_D$ increases, but as b gets smaller than

Table 3. $RMSEA_D$ values with stabilities and cross-sectional correlation at 0.5, and *X* to *Y* path, *c*, varying from 0.1 to 0.5, *Y* to *X* path, *d*, varying from 0.0 to 0.4 with values larger than 0.080 in bold.

		d									
		0		.1		.2		.3		.4	
C	T	S	2S	S	2S	S	2S	S	2S	S	2S
.1	3	0.077	**0.095**	0.000	0.000	0.078	**0.096**	**0.160**	**0.196**	**0.249**	**0.305**
.1	4	0.067	**0.095**	0.000	0.000	0.069	**0.097**	**0.142**	**0.202**	**0.225**	**0.319**
.1	5	0.060	**0.095**	0.000	0.000	0.062	**0.098**	**0.129**	**0.205**	**0.206**	**0.327**
.2	3	**0.156**	**0.191**	0.078	**0.096**	0.000	0.000	**0.082**	**0.100**	**0.170**	**0.209**
.2	4	**0.136**	**0.192**	0.069	**0.097**	0.000	0.000	0.074	**0.104**	**0.157**	**0.222**
.2	5	**0.122**	**0.193**	0.062	**0.098**	0.000	0.000	0.067	**0.107**	**0.145**	**0.230**
.3	3	**0.239**	**0.293**	**0.160**	**0.196**	**0.082**	**0.100**	0.000	0.000	**0.090**	**0.110**
.3	4	**0.210**	**0.297**	**0.142**	**0.202**	0.074	**0.104**	0.000	0.000	**0.085**	**0.120**
.3	5	**0.189**	**0.299**	**0.129**	**0.205**	0.067	**0.107**	0.000	0.000	0.079	**0.126**
.4	3	**0.330**	**0.404**	**0.249**	**0.305**	**0.170**	**0.209**	**0.090**	**0.110**	0.000	0.000
.4	4	**0.293**	**0.415**	**0.225**	**0.319**	**0.157**	**0.222**	**0.085**	**0.120**	0.000	0.000
.4	5	**0.266**	**0.421**	**0.206**	**0.327**	**0.145**	**0.230**	0.079	**0.126**	0.000	0.000
.5	3	**0.435**	**0.537**	**0.354**	**0.435**	**0.277**	**0.340**	**0.202**	**0.248**	**0.125**	**0.153**
.5	4	**0.393**	**0.564**	**0.327**	**0.466**	**0.262**	**0.373**	**0.197**	**0.280**	**0.127**	**0.179**
.5	5	**0.362**	**0.579**	**0.304**	**0.484**	**0.247**	**0.392**	**0.188**	**0.298**	**0.123**	**0.194**

a, $RMSEA_D$ decreases until it equals zero when $c - d = r(b - a)$, and then it increases as $c - d + r(b - a)$ becomes more negative. Note too that even when both c and d are zero, $RMSEA_D$ is non-zero when a and b are unequal. This result indicates that CLPM is not the alternative model for LTVC. Rather, estimating LTVC models allows us to rule out the plausible rival hypothesis that a latent time-varying covariate completely explains *X*-*Y* covariation.

The $RMSEA_D$ for LTVC-S declines slightly as the number of waves increases from 3 to 5. This is surprising given that power increases as the number of waves increases. Perhaps this occurs because the difference between cross-lagged covariances declines as the lag length increases. Because models with more waves have longer lagged covariances, $RMSEA_D$ declines. In contrast, $RMSEA_D$ for the LTVC-2S model does increase as the number of waves increases.

Generally, holding everything else constant, increasing the cross-sectional correlation, denoted here as r, does appear to increase the power of the LTVC-S and LTVC-2S models slightly, but the effect is quite modest. For instance, when $c = 0.1$, $d = 0$, and a and b equal 0.5, LCTC-S is estimated with 3 waves, and $r = 0.2$, $RMSEA_D$ is 0.068, when $r = 0.5$, it is 0.077. We suspect this is due to increased precision with larger correlations.

RI-CLPM

RI-CLPM is CLPM with the addition of a random intercept that acts as a trait factor (Hamaker et al., 2015). Stable intercepts are added to both *X* and *Y*, and the autoregressive (a and b) and cross-lagged (c and d) paths represent paths among residuals of *X* and *Y*, after removing the stable intercept variance.

To more easily compare power values of CLPM to those from RI-CLPM, we started with the CLPM with standardized variables and added an intercept to that model. For simplicity, the intercept variance was the same for *X* and *Y*. We also allowed the two intercepts to be correlated.

Here we consider the model with a unidirectional cause ($d = 0$) and equal stabilities ($a = b = 0.5$), and a cross-sectional correlation of .5. For RI-CLPM, we add random intercepts explaining between 10 to 50% of the total variance of *X* and *Y* and correlating 0.5 with each other. Focusing on LTVC-S and LTVC-2S, the general pattern of results for RI-CLPM were very similar to those for CLPM, which is not too surprising as the model is a modified version of CLPM. However, on average, the $RMSEA_D$ values were lower for RI-CLPM than for CLPM and this trend increased as the percentage of variance due to the random intercepts increased.

The $RMSEA_D$ was affected by the correlation of the intercepts. Again, using the same values for the parameters and examining LTVC-S and LTVC-2S, the $RMSEA_D$ increased as the correlation of the intercepts increased, although these increases were small. Again, in RI-CLPM, power increased under the assumption of stationarity, and power was generally higher for 2S models than for S models.

Decision tree for ruling out a time-varying confounder

When estimating causal cross-lagged effects (e.g., RI-CLPM), the researcher can evaluate the plausibility of

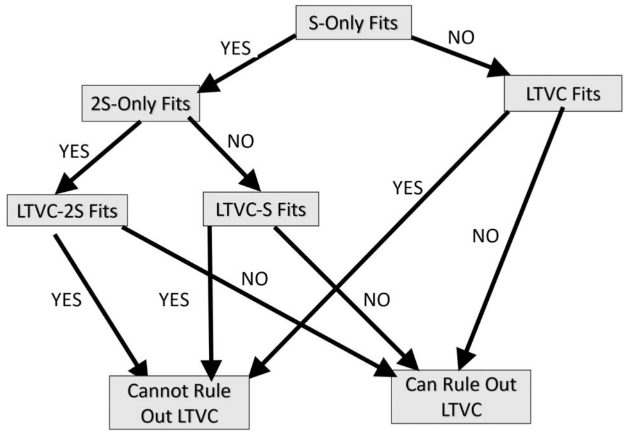

Figure 3. Decision tree for LTVC models.
Note: Begin with "S-Only Fits" and end with "LTVC Model Can Be Ruled Out" or "LTVC Model Cannot Be Ruled Out"; see Table 2 for definitions of the LTVC, S-Only, and 2S-Only models.

whether a latent time-varying covariate successfully explains the observed covariation between the two variables. Figure 3 contains a decision tree that details the steps to evaluate whether the LTVC model fits well. What is meant by the "model fitting well" depends on some explicit empirical cutoff for a poor fitting model, such as a *p* value from the χ^2 test or, especially when sample sizes are large, a good-fitting model might be defined by being below a pre-set RMSEA or SRMR value. Additionally, theory, prior knowledge, and study characteristics would affect the decision as to what is a good model. We strongly urge researchers to preregister their criteria for model rejection. Moreover, we note that because tests of S-Only and 2S-Only evaluate the plausibility of strong stationarity assumptions, it would be advisable to set either to set higher alphas and lower RMSEA values to retain those models.

Note that testing for confounding with the LTVC model is very different from the more familiar tests of model assumptions, e.g., a test of normality. When those familiar tests fail, it signals that the researcher needs to change the analysis, e.g., by transforming the outcome to satisfy the normality assumption. Also, researchers typically hope that the familiar test shows that the estimated model fits well, e.g., no evidence of nonnormality. However, in testing the LTVC model, likely the researchers would hope that the LTVC is not a good-fitting model. Moreover, even if the LTVC model is deemed to be a good-fitting model, the researcher would still estimate the specified causal cross-lagged model. Nonetheless, the LTVC model provides important context: The possibility that a latent time-varying covariate might be the source of all of the causal effects cannot be ruled out.

We begin at the top of Figure 3 by testing the plausibility of stationarity assumption of equal cross-sectional covariances, the S-Only model. If S-Only does not fit well, we must rely on the test of the LTVC model. If we do not reject the LTVC model, the data are consistent with a latent time-varying covariate being able to explain the covariation in the data. Alternatively, if we reject the LTVC model, then LTVC cannot explain the pattern of covariances in the data.

If the S-Only model does fit well, we proceed to see if the 2S-Only model also fits well, relative to the S-Only model. If the S-Only model fits well, but the 2S-Only model does not (i.e., the fit of the 2S-Only model is substantially worse), we compare the fit of LTVC-S to the fit of S-Only model. If we reject the LTVC-S model, we can rule out LTVC. If we fail to reject LTVC-S, then an LTVC could potentially explain our covariance structure. Alternatively, if the 2S-Only model fits as well as the S-Only model, indicating that covariances of the same lag length can be assumed to be equal, then we use the LTVC-2S model to evaluate the plausibility of the LTVC. To do so, we estimate the LTVC-2S model and compare it to the fit of 2S-Only model.[3]

As we discussed earlier, ruling out the LTVC model does *not* mean that CLPM or any other causal model is the correct model. It simply means that the LTVC model is unable to explain the observed pattern of covariation. The rejection of the LTVC model does not indicate the nature of the causal effect, nor does it "prove" that there is a causal relationship between *X* and *Y*.

Illustrations

To demonstrate how these models could be used in practice, we include three examples. For the first example, it is likely that there are no causal lagged effects; the second and third examples are from published papers with lagged effects. Also, the third example includes unequal spacing of the time lags. For each example, we present fit values for all the possible models: LTVC, LTVC-S, LTVC-2S, S-Only, and 2S-Only models. We also present a summary of results from the analyses using the flowchart in Figure 3. All analyses in this article used *Mplus* (Muthén &

[3]Note that the test of stationarity refers to equality of parameters in the original metric. There might be cases in which stationarity holds using the standardized metric. In this case, the standardized loading of *X* and *Y* on the common factor does not change over time. Kenny (2005a) shows how to estimate the standardized LTVC models using SEM. For all three of the examples, the model fit improves with standardized instead of unstandardized stationarity.

Muthén, 1998-2017), and the setups and outputs for these and subsequent analyses are available as Supplemental Material: *Mplus* Setups and Outputs. To identify the LTVC model, we first estimated the model with no constraints. Using the residual variance of the time 1 measure of X from the first run, we fixed its residual variance to that value in a second run. All statistical tests used 0.05 as the value for alpha.

Depression in adolescence

We chose the first example to be one in which there are not likely any lagged effects, as the two variables are two subscales of a construct. Dumenci and Windle (1996) measured depression on 433 adolescent women every 6 months 4 times. Depression is measured using the Center for Epidemiologic Studies Depression Scale or CES-D, which has four subscales, two of which we analyze: Positive Affect (reversed) and Interpersonal Problems.

Table 4 provides the results for all the LTVC models and relevant model comparisons. Using the flowchart in Figure 3, we find the following: Because both S-Only ($\chi^2(3)$ = 4.846, $p = 0.183$, RMSEA = 0.038, SRMR = 0.029) and 2S-Only ($\chi^2(9)$ = 8.302, $p = 0.504$, RMSEA = 0.000, SRMR = 0.028) are good-fitting models, we estimated the LTVC-2S model. When we compared that model to the 2S-Only model, we obtained a good-fitting model, $\chi^2(3)$ = 2.066, $p = 0.559$, $RMSEA_D$ = 0.000. The data were consistent with the view that a latent time-varying covariate could successfully and completely explain the covariation between the two measures of depression. This is what we expected to find as these two variables are presumed to have a common unmeasured confounder: depression.

Depression and self-esteem in college: the BLS study

The second example includes two variables that were hypothesized to be in a causal relationship. Orth et al. (2008), see also Orth et al. (2021), claimed to find an effect from Self-esteem to Depression. The title of their paper is *Low self-esteem prospectively predicts depression in adolescence and young adulthood*. In Study 1 of Orth et al. (2008), the Berkeley Longitudinal Study (BLS), there are 404 United States college students measured four times, once each year.

Table 5 contains LTVC model-fit information for the BLS study. Using the flowchart in Figure 3, we find the following: Because both the S-Only ($\chi^2(3)$ = 6.097, $p = 0.107$, RMSEA = 0.051, SRMR = 0.054) and the 2S-Only ($\chi^2(9)$ = 10.168, $p = 0.337$, RMSEA = 0.018, SRMR = 0.061) are good-fitting models, we estimated the LTVC-2S model. When we compared the LTVC-2S to the 2S-Only model, we obtained an equivocal fit with a statistically significant chi square ($\chi^2(3)$ = 10.628, $p = 0.014$) but an RMSEA below 0.08 ($RMSEA_D$ = 0.079). The decision as to whether the LTVC-2S model fits the data is ambiguous. Ordinarily, unless the sample size was large ($N > 500$), we would rely on the chi square test. Moreover, because all three LTVC models have p values less than 0.05, we would reject the LTVC model. Even if we reject the LTVC model, we know nothing about the correct causal mechanism. Later in this article, we

Table 4. LTVC models for the Dumenci and Windle (1996) study.

Model	χ^2	*df*	*p*	RMSEA	SRMR	C. Model[a]	χ^2 *diff*	*df*	*p*	$RMSEA_D$
LTVC	2.508	3	.474	0.000	.007					
S-Only	4.846	3	.183	0.038	.029					
LTVC-S	8.909	9	.446	0.000	.031	S-Only	4.063	6	.668	0.000
2S-Only	8.302	9	.504	0.000	.028	S-Only	3.456	6	.750	0.000
LTVC-2S	10.368	12	.584	0.000	.028	2S-Only	2.066	3	.559	0.000

[a]Comparison Model.
Note: See Table 2 for model definitions.

Table 5. LTVC models for the BLS study.

Model	χ^2	*df*	*p*	RMSEA	SRMR	C. Model[a]	χ^2 *diff*	*df*	*p*	$RMSEA_D$
LTVC	8.210	3	.042	0.066	.014					
S-Only	6.097	3	.107	0.051	.054					
LTVC-S	17.751	9	.038	0.049	.053	S-Only	11.654	6	.070	0.048
2S-Only	10.168	9	.337	0.018	.061	S-Only	4.071	6	.667	0.000
LTVC-2S	20.796	12	.053	0.043	.058	2S-Only	10.628	3	.014	0.079

[a]Comparison Model.
Note: See Table 2 for model definitions.

Table 6. LTVC models for the Núñez-Regueiro et al. (2022) study.

Model	χ^2	*df*	*p*	RMSEA	SRMR	C. Model[a]	χ^2 *diff*	*df*	*p*	$RMSEA_D$
LTVC	1.334	6	.970	0.000	.004					
S-Only	5.114	4	.276	0.017	.034					
LTVC-S	13.167	14	.513	0.000	.035	S-Only	8.053	10	.624	0.000
2S-Only	16.975	14	.258	0.015	.039	S-Only	11.861	10	.294	0.014
LTVC-2S	17.594	19	.550	0.000	.038	2S-Only	0.619	5	.987	0.000

[a]Comparison Model.
Note: See Table 2 for model definitions.

explore a non-causal model in which Depression is a leading indicator of a latent variable and Self-esteem lags by more than year.

Academic self-concept and achievement

We tested the LTVC models using the data from Núñez-Regueiro et al. (2022) who tested a series of models on a sample of 933 French high school students to examine the reciprocal effects between Academic Self-Concept and Academic Achievement. The students were measured during their first and second years of high school. During each year they were measured three times a year, trimesters, but because students were on internship during the 5^{th} trimester, data were not collected for that trimester. For the 10 time-differences in pairs of measurements, 3 were 1 unit, 3 were 2 units, 2 were 3 units, 1 was 4 units, and 1 was 5 units. Because of this pattern, the formulas for degrees of freedom in Table 2 cannot be used. Núñez-Regueiro et al. included several covariates in their analyses; our analyses did not include them here, but we do discuss them in a later section. We note that Núñez-Regueiro et al. in their article made no claims of any causal effects.

Table 6 contains model-fit information for the Núñez-Regueiro et al. study. Using the flowchart in Figure 3, we find the following: Because both the S-Only ($\chi^2(4) = 5.114$, $p = 0.276$, RMSEA = 0.017, SRMR = 0.034) and the 2S-Only ($\chi^2(14) = 16.975$, $p = 0.258$, RMSEA = 0.015, SRMR = 0.039) are good-fitting models, we estimated the LTVC-2S model. When we compared that to the 2S-Only model, we obtained a good-fitting model, $\chi^2(5) = 0.619$, $p = 0.987$, $RMSEA_D = 0.000$.

Thus, we cannot rule out the possibility that some unmeasured time-varying covariate completely explains the covariation between the two variables. The hypothesis that Academic Self-Concept and Academic Achievement were caused by the same variable, perhaps Academic Ability, remains. However, as we discussed and showed in Table 3, the power to reject the LTVC with reciprocal effects is poor. Núñez-Regueiro et al. (2022) expected and found evidence for reciprocal effects in most of their analyses. In sum, we are left with two competing explanations for the results: reciprocal causal or a latent time-varying confounder.

Model extensions

Here we consider six extensions to our model. First, we develop an observation-level LTVC model that is a blend of LTVC and CLPM. We next consider multivariate models, beyond the bivariate model that we have previously discussed. We also consider the case that the latent variable is a not a single variable, but rather that there are multiple latent variables. Next, we discuss the inclusion of covariates into the LTVC model. Next, we consider modeling the process of change of the latent time-varying confounder, *C*. Lastly, we relax the assumption that the latent time-varying confounder simultaneously causes *X* and *Y*.

The LTVC-CLPM model

In this section, we develop an observation-level LTVC model that is a blend of LTVC and CLPM. Unlike CLPM, there is no cross-sectional correlation between residuals (*E* and *F* in Equations 5 and 6), but rather there is a latent-time varying confounder, which changes in an autoregressive fashion. We refer to this model as the LTVC-CLPM (see Figure 4). The model has a latent variable, *C*, that causes *X* and *Y* at each wave and changes in a first-order autoregressive fashion and the variables *X* and *Y* also have autoregressive paths, *a* and *b*, as well as cross-lagged paths, *c* and *d*. One major advantage this model has over the previous LTVC model with correlated residuals is that this model explicitly includes the possibility of an alternative model, i.e., *c* and *d* being non-zero. This allows the latent time-varying confounding to explain some but not all of the *X*-*Y* covariation. Note that CLPM is a special case of this model in which the autoregressive path for the latent time-varying confounder, *e*, equals zero (Dwyer, 1983, p. 362).

Generally, for observation-level models with latent variables, all the wave-one variables, including C_1, are

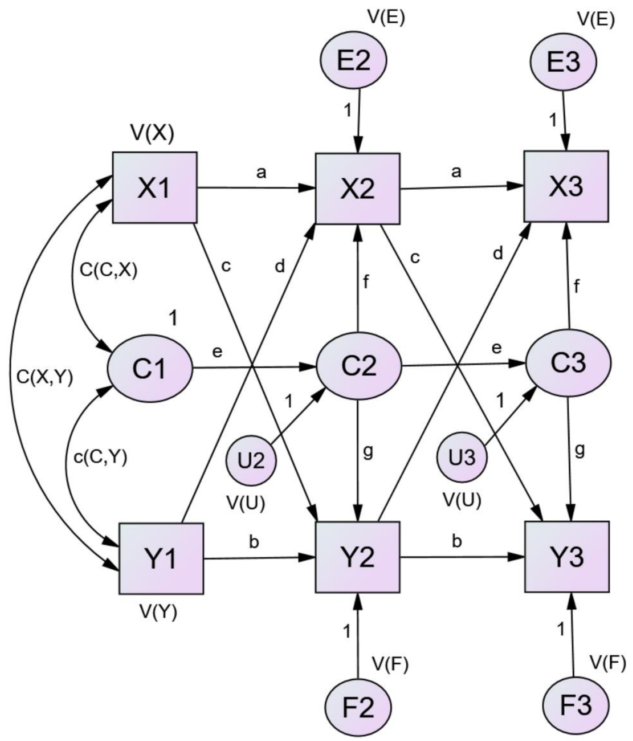

Figure 4. The LTVC model with observation-level autoregessive paths: LTVC-CLPM.
Note: *X* and *Y* are measured variables at 3 Times, *C* a latent time-varying confounder with an *f* path to *X* and a *g* path to *Y*, *a* the autoregessive path for *X*, *b* the autoregessive path for *Y*, *c* the path from *X* to lagged *Y*, *d* the path from *Y* to lagged *X*, and *E* and *F* the residuals for *X* and *Y*, *V*() a variance, and *C*() a covariance.

treated as exogenous, as we have done in Figure 4. To identify the model, we make assumptions of stationarity. The structural equations are:

$$C_{it} = eC_{i,t-1} + U_{it} \quad (7)$$

$$X_{it} = fC_{it} + aX_{i,t-1} + dY_{i,t-1} + E_{it} \quad (8)$$

$$Y_{it} = gC_{it} + bY_{i,t-1} + cX_{i,t-1} + F_{it} \quad (9)$$

As shown in Figure 4, we set equal the factor loadings of *f* and *g*, the autoregressive paths of *a*, *b*, and *e*, the cross-lagged paths of *c* and *d*, and the residual variances of *E* for *X*, *F* for *Y*, and *U* for *C*. As we did for LTVC, we standardize C_1. Lastly, to make the model fully stationarity, we need to constrain the variances and covariances of the exogenous variables, X_1, Y_1, and C_1.

To make those constraints with an SEM program, we chose to reparametrize the model in the same way as Usami et al. (2019). We created a separate phantom variable for each of the 2 *T* measures. Each X_t measure has a phantom variable of L_{Xt} (e.g., $X_1 = s_X L_{X1}$), and each Y_t measure has a phantom variable (e.g., L_{Yt} and $Y_1 = s_Y L_{Y1}$. Two new parameters are added to the model, s_X and s_Y, but we lose two parameters by fixing the variances of L_X and L_Y to one. In this new model, in Figure 4, as well as Equations 7, 8, and 9, we replace *X* with L_X and *Y* with L_Y. This formulation turns the parameters of *f*, *g*, *a*, *b*, *c*, *d*, and *e* into beta coefficients, and s_X and s_Y are the standard deviations of *X* and *Y*, respectively.

The nonlinear constraints that need to be imposed to create stationarity variance covariance matrix are:

$$s_U{}^2 = 1 - e^2 \quad (10)$$

$$r_{cx} = \frac{dge + f(1-be)}{(1-ae)(1-be) - cde^2} \quad (11)$$

$$r_{cy} = \frac{cfe + g(1-ae)}{(1-ae)(1-be) - cde^2} \quad (12)$$

$$r_{xy} = \left[e\left(fbr_{cy} + gar_{cx} + fcr_{cx} + gdr_{cy}\right) + fg + ac + db\right] / (1 - cd - ab) \quad (13)$$

$$s_E{}^2 = 1 - \left[a^2 + d^2 + 2adr_{xy} + ef\left(ar_{cx} + dr_{cy}\right) + fr_{cx}\right] \quad (14)$$

$$s_F{}^2 = 1 - \left[b^2 + c^2 + 2bcr_{xy} + eg\left(br_{cy} + cr_{cx}\right) + gr_{cy}\right] \quad (15)$$

The total number of model parameters is 9, regardless of the number of waves. To validate that the model is identified, we generated several population covariance matrices, then conducted an SEM analysis, and each time the program successfully recovered the population parameters. Note that if *C* has zero stability, $e = 0$, the model is the same as CLPM, but paths *f* and *g* are not identified.

The full set of stationarity constraints imply all the previously discussed 2S constraints, $(T-1)^2$, as well equal lagged covariances of the same lag length for both *X* and *Y*, $(T-1)(T-2)$, and equal variances across time for *X* and for *Y*, $2(T-1)$, resulting in the total number of stationarity constraints equals $(2T-1)(T-1)$.

We illustrate the estimation of the LTVC-CLPM using the earlier discussed Dumenci and Windle (1996) study. To test stationarity, we imposed 21 different constraints on the variance-covariance matrix. The fit for this stationarity model is $\chi^2(21) = 26.986$, $p = 0.171$, RMSEA = 0.026, SRMR = 0.060, a good-fitting model. Fitting the LTVC-CLPM, we obtain $\chi^2(27) = 85.714$, $p < 0.001$, RMSEA = 0.071, SRMR = .073. When we remove the test of stationarity, we obtain $\chi^2(6) = 58.728$, $p < 0.001$, $\text{RMSEA}_\text{D} = 0.142$. Both the chi square test and the RMSEA suggest that LTVC-CLPM is not a good-fitting model.

Even if we believed LTVC-CLPM were a good model, the model estimates are difficult to interpret. The standardized loading of X on C is 0.995 and of Y on C is .554. Examining the beta weights, the autoregressive coefficient for C or e is 0.859, the autoregressive effect for X (absence of Positive Affect) or a is -0.144, and for Y or b is 0.316 (Interpersonal Problems). The negative autoregressive effect is difficult to comprehend. Both lagged effects (c and d) are negative, which is implausible. In terms of the residuals, the largest ones are for the autocorrelations of X and Y.

When we estimated the LTVC-CLPM with the other two example datasets, we also obtained poor fit and anomalous results with huge factor loadings on the LTVC, negative autoregressive paths and cross-lagged paths with the opposite sign than the cross-sectional correlations. The LTVC–CLPM estimates overly large factor loadings on C and high stabilities for C that the autoregressive and cross-lagged paths must compensate for by having the negative paths. Although we do not understand exactly why this is happening, we very speculatively discuss two different possible explanations.

One possibility is that the model is too complicated for there to be stable estimates. Perhaps if we introduced prior knowledge in a Bayesian analysis (e.g., paths a and b are not likely to be negative), we might obtain more reasonable estimates.

Alternatively, perhaps the over-time model is misspecified. One possibility is that besides the LTVC, there are also random intercepts for X and Y, and no autoregressive or cross-lagged paths. Another possibility is there might be a second unmeasured confounder, a state factor, which increases the cross-sectional covariances, but not the lagged ones.

In summary, the LTVC-CLPM has several advantages over the earlier LTVC model presented in Figure 1. First, all of its model parameters are identified. Second, CLPM is a special case of this model, whereas for LTVC, there is no clear alternative causal model. Third, LTVC-CLPM is an observation-level model in that it has lagged paths for X and Y, whereas LTVC is a residual model with no such paths. Several analysts (Andersen, 2022; Lüdtke & Robitzsch, 2022; Murayama & Gfrörer, 2024) have argued that the true model is the observation-level model, whereas the residual-level model is only an approximation used to simplify the estimation of the observation model.

However, the residual-level model of LTVC has several theoretical advantages over the LTVC-CLPM. First, LTVC-CLPM model requires stronger stationarity assumptions than the LTVC model. Second, historically, over-time factor models have presumed the residual model (Duncan, 1972; Kenny, 1973; Newsom, 2024). Thus, in this case we need not necessarily presume that observation model is the true model, and residual model is a simplification. Third, the estimates for the LTVC-CLPM were anomalous, and the models fit poorly, whereas for two of the three examples, we obtained a good-fitting LTVC model. Of course, three studies are too small of a sample to make a generalization, but it does suggest that the residual model LTVC may be a sensible way to model no causal cross-lagged effects. A systematic comparison of the two approaches would be beneficial. Especially informative would be to include studies similar to the Dumenci and Windle (1996) study in which X and Y are likely not causing each other but rather are presumed to measure the same construct.

Multivariate and multi-factor models

Previous discussion was limited to two-variable, multi-wave models; we consider here having three or more variables. With three or more variables and just two waves, a model in which a single factor causes each variable at each time without any constraints on the factor loadings is identified (Duncan, 1972). With four or more variables, this single-factor model can be estimated and tested within each cross-section.

With three or more indicators and just two waves, we can relax the assumption of a single-factor and instead allow for multiple factors. One strategy is to assume quasi-stationarity (Kenny, 1973, 1975; Kenny & Milan, 2012): Each variable is caused by p latent common factors and the variances and covariances of these p latent factors are the same at both times. The common factor's loadings of measure m at time t for factor i is denoted as f_{itm}. The quasi-stationarity assumption is that over time all the loadings for variable m change from time 1 to time 2 by a proportional constant:

$$k_m = f_{12m}/f_{11m} = f_{22m}/f_{21m} = \ \dots \ = f_{p2pm}/f_{p1m} \quad (16)$$

Given this assumption, the ratio of the same cross-sectional covariances between X and Y or C_{X2Y2}/C_{X1Y1} equals $k_X k_Y$ and the ratio of the two cross-lagged covariances between X and Y or C_{X1Y2}/C_{X2Y1} equals k_Y/k_X. Kenny (2025a) shows how to estimate the quasi-stationarity model by SEM and how to integrate it within the LTVC approach.

We estimated both a one-factor and multi-factor quasi-stationarity model with the four waves and all our indicators of Depression from the Dumenci and Windle (1996) study. The model of quasi-stationarity for that data fits quite well ($\chi^2(42) = 36.248$, $p = 0.721$, RMSEA = 0.000, SRMR = 0.016) and better than the single-factor model ($\chi^2(74) = 107.968$, $p = 0.006$, RMSEA = 0.033, SRMR = 0.052) with the improvement in model fit being ($\chi^2(32) = 71.720$, $p < 0.001$).

So far, we have assumed that the latent variable, C, is a single variable. What happens if there are multiple latent variables? The quasi-stationarity assumption can be made but not tested for two-variable data in which the latent variable is the sum of many variables. If the assumption of quasi-stationarity holds, the LTVC over-identifying restrictions in Equation 4 still hold.

Measured time-invariant covariates

Almost always, 2VMW studies have covariates. Some of these covariates are time varying and are measured at each time point, and for those covariates, it would seem sensible to treat them as another variable, like X and Y, making the analysis multivariate. In the remainder of the section, we consider measured time-invariant covariates, i.e., variables that are measured only once because they are likely hardly to change over the course of the study.

We propose two models that include time-invariant covariates. A measured time-invariant covariate either directly causes the X and Y measures; that is, they have direct effects on X and Y, and the covariates are uncorrelated with C. Note this model explains all of the covariation between the covariates and the two variables X and Y. Alternatively, they cause only the latent time-varying covariate and have an indirect effect on X and Y through C. This indirect model of covariates results in the LTVC being an identified model, as the measured covariate serves as the third "indicator" of the latent time-varying covariate. The model can be considered to be a version of the Multiple Indicator Multiple Causes (MIMIC) model in which the covariates are the multiple causes of C, and X and Y are the multiple indicators. The indirect model has the number of covariates times the number of waves fewer covariate paths than the direct model, making the indirect model is a special case the direct model.

As an illustration, we return to the Núñez-Regueiro et al. (2022) study and consider Age, Gender, and Father's Socioeconomic Status as covariates. Including only those cases with complete data on the covariates, the sample size is reduced to 762. We proceed using the flowchart in Figure 3. We need to expand the S-Only assumption to include the requirement that each covariate has the same effect on X and Y at each time, which for each covariate adds $2(T-1)$ or 24 additional constraints to the model. The fit of this new S-Only model is $\chi^2(28) = 61.641$, $p < 0.001$, RMSEA = 0.040, SRMR = .045. Although the RMSEA is acceptable, we believe that we should cautiously reject the assumption of equal effects at each wave and proceed to test the LTVC model with no stationarity assumptions. First assuming direct effects of the covariates, the fit for the model with direct effects to X and Y is $\chi^2(6) = 1.069$, $p = 0.983$, RMSEA = 0.000, SRMR = .004. Comparing this result to those in Table 6, we see that controlling for the covariates does little to alter the conclusion that the LTVC model is a good-fitting model. The fit for the indirect effect model is $\chi^2(20) = 33.945$, $p = 0.026$, RMSEA = 0.030, SRMR = .026. To test the assumptions of the indirect effect model over the direct effect model, we obtain $\chi^2(14) = 32.876$, $p = 0.003$, $\text{RMSEA}_D = 0.042$, suggesting the indirect effect model is poor fitting. (There are only 14 degrees of freedom more for the indirect model over the direct model when there are 15 fewer covariate paths estimated in the indirect model than the direct model because the indirect model is identified, and so one degree of freedom is lost by not fixing a parameter.) Besides being a poorer fitting model than the direct model, the solution for the indirect model is not very interpretable, as the standardized loadings for Y are very near one.

Because the indirect effect model is identified, it has the potential to make LTVC models much more useful. In fact, the LTVC model with indirect effects of the covariate is identified with as few as two waves of data. It might even be able to expand the indirect model to allow for autoregressive and causal cross-lagged effects in X and Y. However, although this model is identified in theory, in practice it may be difficult to estimate. Akin to instrumental variable estimation, we would likely need very strong covariates that explain substantial amounts of variance in both X and Y for this approach to be useful.

Understanding the structure of change in a latent time-barying covariate

Although the LTVC model is under-identified, the squared over-time correlations of the confounding variable are identified. Understanding the structure of change on the latent-time varying confounder can be highly beneficial. Dwyer (1983) has shown the CLPM implicitly presumes that the common factor that explains the time-one covariation has zero over-time correlations (i.e., a state or occasion factor):

> (Latent) variables may not decay rapidly enough to be treated as occasion factors. Such factors inflate *cross-lagged* covariances and can introduce a bias in the estimation of lagged effects that is most damaging (p. 362).

Alternatively, should the latent variable be totally stable, then perhaps estimates from RI-CPLM may have less bias. Thus, an analysis of the stability of C would be beneficial in understanding the amount of bias due to an unmeasured confounding variable.

As an illustration, we use the Núñez-Regueiro et al. (2022) study. We found that LTVC-2S model is a good-fitting model (see Table 6). We compare these new models to the LTVC-2S model because they place constraints on the variance-covariance matrix of the latent time-varying covariate. We first evaluate whether the latent variable is perfectly stable, a trait. Estimating the LTVC-2S model with latent variable correlations of one, $\chi^2(24) = 42.696$, $p = 0.011$, RMSEA = 0.029, SRMR = .041. When we compare the LTVC-2S model to a model with perfectly stable correlations, $\chi^2(5) = 25.102$, $p < .001$. A trait model cannot explain the covariance structure of C.

Another possibility is to presume that there is a state or occasion factor at each time which is completely transitory. For such a model, all over-time correlations are zero. When we estimate the LTVC-2S model with all C correlations fixed to zero, we find $\chi^2(24) = 161.143$, $p < 0.001$, RMSEA= 0.078, SRMR = .204. When we compare the LTVC-2S model to a model with zero correlations, we find $\chi^2(5) = 143.549$, $p < .001$. A state model is also not good-fitting.

The next possibility is that X-Y covariances have an autoregressive structure. The autoregressive trait model for C is $C_t = e_t C_{t\text{-}1} + U_t$, where e_t is the autoregressive coefficient and $C_{t\text{-}1}$ and U_t are uncorrelated. Because we found evidence of 2S stationarity, we can assume that the autoregressive coefficient does not change over time.[4] When we subtract the chi square for LTVC-2S, we obtain $\chi^2(4) = 1.706$, $p = 0.790$, which is a good-fitting model. The lag-one autoregressive coefficient is estimated to be .942. Although this indicates high lag-1 stability, the implied correlation from wave one to wave six is the autoregressive coefficient raised to the fifth power, which equals 0.743.

We have so far discussed how a single type of factor, either trait, state, or autoregressive, can explain the pattern of LTVC covariances. We now examine two such factors at the same time. For instance, the common factor has two parts: one part is autoregressive, and the other part is totally unstable over time, a state factor (Humphreys, 1960). To identify this model, 2S stationarity assumptions must be made. When we estimated this model for the Núñez-Regueiro et al. (2022) study, we did not find evidence for the utility of adding a state factor to the autoregressive model, obtaining $\chi^2(1) = 0.001$, $p = 1.00$. Thus, the autoregressive model is the preferred model to explain the pattern of change of the latent time-varying confounder for this dataset.

Other possible combinations could be estimated. For instance, akin to RI-CLPM, we could have a stable trait factor, i.e., random intercept, as well as an autoregressive factor. Although difficult to fit, the STARTS model (Kenny & Zautra, 2001), which includes a perfectly stable trait factor, an autoregressive factor, and state factor, could be fitted. Another possibility is a linear growth model with a slope and intercept. If there were enough waves, we could consider a cyclical model of change. See Liu and West (2016) for an introduction to modeling cycles and Muthén et al. (2024) for a discussion of statistical estimation.

Lag-lead LTVC model

A key assumption of the LTVC model is that the time-varying confounder causes X and Y simultaneously. It is possible to estimate a version of the LTVC model in which the effect of C on one variable is simultaneous and a lagged effect on the other variable. We can view the Lag-lead Model as the alternative model to LTVC in that the lag length for LTVC is zero.

Kenny (1973) discussed a Lag-lead model in which the lag is fixed to be one unit of time:

$$X_{it} = a_t C_{it} + E_{it} \tag{17}$$

$$Y_{it} = b_t C_{i(t+1)} + F_{it} \tag{18}$$

In this model, X is said to be a *leading* indicator of C, and Y a *lagging* indicator. Like the LTVC model in Figure 1, there are no constraints on the correlational structure of C. This approach has two serious limitations: lag length must be specified, and which variables are the leading and lagging indicators must be known.

[4]Both Rickard (1972) and Kenny (1973) considered a three-wave model with no stationarity constraints in which the latent time-varying confounder has an autoregressive structure. That model has two degrees of freedom. One degree of freedom evaluates the assumption of a first-order autoregressive structure and the other evaluates the LTVC model.

Kenny (2025b) presented a new model that empirically estimates the leading indicator and the lag length using non-linear model constraints. The model presumes that the earlier discussed 2S assumptions hold and that the model of change of the latent time-varying covariate can be specified. For the BLS study, we first estimated a model of change that was autoregressive, and the model fit was poor ($\chi^2(13) = 28.212$, $p = 0.008$, RMSEA= 0.054, SRMR = 0.067). However, a model with a lagged autoregressive factor and a simultaneous state factor fits well ($\chi^2(12) = 12.382$, $p = 0.416$, RMSEA= 0.009, SRMR = 0.063). The lag length for this model was estimated to be −1.234 (95% CI: −1.873 to −0.595), which indicates that Depression leads Self-esteem by about 15 months. Note that lag length is defined in this model in terms of Self-esteem (X) leading Depression (Y), and so a negative value implies Self-esteem is a lagging indicator. When we set the lag length to zero, the fit worsened considerably when compared to the model with lag length free: $\chi^2(1) = 8.495$, $p = .004$.

These results suggest that there is a General Negativity Factor, which first appears in Depression. Perhaps it is triggered by some life event. That factor is seen first in Depression, but then over a year later in Self-esteem.

Conclusion

We have urged researchers interested in showing lagged causal effects with 2VMW data to demonstrate that a model that the covariation between the two variables is not entirely due to a confounding variable. We have developed a testable, but under-identified model of a confounding variable moving through time, as well as a testable and identified model of confounding variable with stationarity assumptions. We investigated the power to reject these time-varying covariate models, and we found that power increases with larger asymmetric causal effects, number of waves, and stationarity of parameters.

It is crucial to realize that sometimes an LTVC model might appear to be a good-fitting model, when in fact it is not the correct model. First, as shown in Table 3, it might well be that that the true model is reciprocal with both X causing Y and Y causing X. Second, if stationarity assumptions cannot be made or effects are small, there may be insufficient power to reject the LTVC model. Moreover, even if we deem that that the LTVC is a poor model, we cannot conclude that the estimated causal model (e.g., CLPC) is correct or that the causal estimates are unbiased. Those estimated effects may still be seriously biased by unmeasured confounders.

An empirically based strategy could be used to validate that the LTVC model represents an appropriate way to model a latent time-varying confounder with 2VMW data. An extensive re-analysis of many datasets in which two variables are likely indicators of a common variable, like the Dumenci and Windle (1996) CES-D reanalysis in this article, would be informative. Moreover, if further improvements in LTVC-CLPM could be made, that model and the LTVC using correlated residuals could be compared.

Instead of estimating an LTVC model, as we have done in this article, an alternative would be to conduct a sensitivity analysis. That is, the researcher would investigate what happens to the estimates of the causal effects, when a latent-time varying confounder is added to the model. A sensitivity analysis does not require that the model be identified. Consider, for example, a researcher who estimates a longitudinal causal model using RI-CLPM and finds evidence of a lagged X to Y effect. One possible sensitivity analysis would be to replace the random intercept with a latent time-varying confounder. The researcher might fix the factor loadings of X and Y, to identify the model, and then see how much the X to Y effect changes as the factor loadings on the LTVC increase. Elaborating the process of how to conduct sensitivity analyses for LTVC models is deserving of further investigation.

In 1963, Donald T. Campbell (Campbell & Stanley, 1963, p. 69) speculated that equal cross-lagged correlations might be indicative of the absence of causal cross-lagged effects. His students (Kenny, 1973; Rickard, 1972) provided a more formal rationale for that approach using latent variable modeling. However, for some 50 years, researchers abandoned the topic of latent confounders for bivariate, multi-wave data. Why did this happen?

One reason is the mistaken belief that confounding can be controlled in a CLPM analysis regression. As first shown by Dwyer (1983, p. 362), CLPM implicitly presumes that the common factor that explains the residual cross-sectional covariation has zero stability over time (i.e., is a state or occasion factor). CLPM makes a very strong and likely implausible assumption about the common factor.

A second reason that researchers have ignored the common factor model is that researchers who gather such data are generally interested in the causal effects. Rogosa (1980) correctly noted that researchers with longitudinal data have little interest in testing confounding but rather are much more interested in estimating and testing causal effects. Although we understand that interest and the desire to find causal effects, it is, nonetheless, beneficial to explicitly rule out confounding as a complete explanation of *X-Y* covariation.

A third reason is that in the early 1980s when this topic was being debated, we did not have the SEM tools that we now have. The focus then was primarily on two-wave studies. Since then, the focus moved from two-wave to multi-wave studies, and we now have powerful SEM programs to estimate such models.

Drawing causal inferences from 2VMW is a challenging enterprise requiring substantive knowledge of the process and measurement issues involved and consideration of alternative non-causal mechanisms. The LTVC models in this article begin to provide us with tools in this endeavor.

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was not supported.

Role of the Funders/Sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Andersen, H. K. (2022). Equivalent approaches to dealing with unobserved heterogeneity in cross-lagged panel models? Investigating the benefits and drawbacks of the latent curve model with structured residuals and the random intercept cross-lagged panel model. *Psychological Methods*, *27*(5), 730–751. https://doi.org/10.1037/met0000285

Arbuckle, J. L. (2021). *Amos (Version 28.0) [Computer Program]*. IBM SPSS.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Rand McNally & Company.

Dumenci, L., & Windle, M. (1996). A latent trait-state model of adolescent depression using the Center for Epidemiologic Studies-Depression Scale. *Multivariate Behavioral Research*, *31*(3), 313–330. https://doi.org/10.1207/s15327906mbr3103_3

Duncan, O. D. (1972). Unmeasured variables in linear models for panel analysis. In H. Costner (Ed.), *Sociological methodology* (pp. 26–82 Jossey-Bass). https://doi.org/10.2307/270729

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences*. Oxford University Press. https://doi.org/10.4236/ojps.2015.51003

Feng, Y., & Hancock, G. R. (2023). Power analysis within a structural equation modeling framework. In R. H. Hoyle (Ed.) *Handbook of structural equation modeling* (2nd ed., pp 163–183). Guilford Press. https://doi.org/10.4135/9781452226576

Greenland, S., & Robins, J. M. (1986). Identifiability, exchangeability, and epidemiological confounding. *International Journal of Epidemiology*, *15*(3), 413–419. https://doi.org/10.1093/ije/15.3.413

Grosz, M. P., Rohrer, J. M., & Thoemmes, F. (2020). The taboo against explicit causal inference in non-experimental psychology. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(5), 1243–1255. https://doi.org/10.1177/1745691620921521

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

Huang, J., & Yuan, Y. (2017). Bayesian dynamic mediation analysis. *Psychological Methods*, *22*(4), 667–686. https://doi.org/10.1037/met0000073

Humphreys, L. G. (1960). Investigations of the simplex. *Psychometrika*, *25*(4), 313–323. https://doi.org/10.1007/BF02289750

Kenny, D. A. (1973). Cross-lagged and synchronous common factors in panel data. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 153–166). Seminar Press. https://doi.org/10.1007/978-1-4612-0897-3

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887–903. https://doi.org/10.1037/0033-2909.82.6.887

Kenny, D. A. (2025a). Testing for standardized and quasi-stationarity of covariance structures in longitudinal data. Retrieved from OSF https://osf.io/7djek

Kenny, D. A. (2025b). The lag-lead model in two-variable longitudinal models [Unpublished paper]. Retrieved from OSF https://osf.io/cfe9s

Kenny, D. A., & Milan, S. (2012). Identification: A non-technical discussion of a technical issue. In R. Hoyle

(Ed.), *Handbook of structural equation modeling* (pp. 145–163). Guilford. https://doi.org/10.1037/12323-00

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In A. Sayer & L. M. Collins (Eds.), *New methods for the analysis of change* (pp. 243–263). American Psychological Association. https://doi.org/10.1037/10409-000

Liu, Y., & West, S. G. (2016). Weekly cycles in daily report data: An overlooked issue. *Journal of Personality*, *84*(5), 560–579. https://doi.org/10.1111/jopy.12182

Lüdtke, O., & Robitzsch, A. (2022). A comparison of different approaches for estimating cross-lagged effects from a causal inference perspective. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(6), 888–907. https://doi.org/10.1080/10705511.2022.2065278

Moerkerke, B., Loeys, T., & Vansteelandt, S. (2015). Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of post-treatment confounding. *Psychological Methods*, *20*(2), 204–220. https://doi.org/10.1037/a0036368

Murayama, K., & Gfrörer, T. (2024). Thinking clearly about time-invariant confounders in cross-lagged panel models: A guide for choosing a statistical model from a causal inference perspective. *Psychological Methods*, Advance online publication. https://doi.org/10.1037/met0000647

Muthén, B., Asparouhov, T., Keijsers, L. (2024). *Dynamic structural equation modeling with cycles* [Unpublished paper]. www.statmodel.com/download/DSEMCycles.pdf

Muthén, L. K., & Muthén, B. O. (1998-2017). *Mplus user's guide* (8th ed.). Muthén & Muthén. https://www.statmodel.com/

Newsom, J. T. (2024). *Longitudinal structural equation modeling: A comprehensive introduction* (2nd ed.). Routledge. https://doi.org/10.4324/9781003263036

Núñez-Regueiro, F., Juhel, J., Bressoux, P., & Nurra, C. (2022). Identifying reciprocities in school motivation research: A review of issues and solutions associated with cross-lagged effects models. *Journal of Educational Psychology*, *114*(5), 945–965. https://doi.org/10.1037/edu0000700

Orth, U., Clark, D. A., Donnellan, M. B., & Robins, R. W. (2021). Testing prospective effects in longitudinal research: Comparing seven competing cross-lagged models. *Journal of Personality and Social Psychology*, *120*(4), 1013–1034. https://doi.org/10.1037/pspp0000358

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology*, *95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Rickard, S. (1972). The assumptions of casual analyses for incomplete causal set of two multilevel variables. *Multivariate Behavioral Research*, *7*(3), 317–359. https://doi.org/10.1207/s15327906mbr0703_5

Robins, J. M., & Hernán, M. A. (2009). Estimation of the causal effect of time-varying exposures. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 553–599). CRC Press. https://doi.org/10.1201/9781420011579

Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*(2), 245–258. https://doi.org/10.1037/0033-2909.88.2.245

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://www.jstatsoft.org/article/view/v048i02 https://doi.org/10.18637/jss.v048.i02

Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*(1), 83–90. https://doi.org/10.1007/BF02294150

Savalei, V., Brace, J. C., & Fouladi, R. T. (2024). We need to change how we compute RMSEA for nested model comparisons in structural equation modeling. *Psychological Methods*, *29*(3), 480–493. https://doi.org/10.0.4.13/met0000537 https://doi.org/10.1037/met0000537

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

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

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/10.1177/1094428119847278