**Published in 2015, in *Psychological Methods, 20*(1), 102-116.**
**doi: 10.1037/a0038889.**

Running head: CROSS-LAGGED PANEL MODEL

# A critique of the cross-lagged panel model

E. L. Hamaker[1], R. M. Kuiper[1] and R. P. P. P. Grasman[2]

1. Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University
2. Psychological Methodology, University of Amsterdam

Author Note:

This study was supported by the Netherlands Organization for Scientific Research (NWO; VIDI Grant 452-10-007).

Correspondence concerning this paper should be addressed to E. L. Hamaker, Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University, P.O. Box 80140, 3508 TC, Utrecht, The Netherlands. Email: e.l.hamaker@uu.nl.

Abstract

The cross-lagged panel model is believed by many to overcome the problems associated with the use of cross-lagged correlations as a way to study causal influences in longitudinal panel data. The current paper however shows that if stability of constructs is to some extent of a trait-like, time-invariant nature, the autoregressive relationships of the cross-lagged panel model fail to adequately account for this. As a result, the lagged parameters that are obtained with the cross-lagged panel model do not represent the actual within-person relationships over time, and this may lead to erroneous conclusions regarding the presence, predominance, and sign of causal influences.

We present an alternative model that separates the within-person process from stable between-person differences, and discuss how this model is related to existing structural equation models that include cross-lagged relationships. Furthermore, we derive the analytical relationship between the cross-lagged parameters from this alternative model and those from the cross-lagged panel model. Through simulations we demonstrate the spurious results that may arise when using the cross-lagged panel model to analyze data that include stable, trait-like individual differences. This is followed by the presentation of a modeling strategy to avoid this pitfall, which we illustrate using an empirical data set. The implications for existing and future cross-lagged panel research are discussed.

A critique of the cross-lagged panel model

In 1980, Rogosa's seminal paper *A critique of the cross-lagged correlation* was published, which successfully conveyed the message that comparing cross-lagged correlations from longitudinal panel data is an inappropriate basis for making causal inferences.[1] One of the key insights stemming from Rogosa's paper is that, if two constructs are characterized by different degrees of stability, the comparison of cross-lagged correlations may lead to spurious conclusions regarding the causal mechanism. Since then, most researchers interested in causality in panel data have abandoned cross-lagged correlations and endorsed what we will referred to in this paper as the *cross-lagged panel model* (CLPM) instead. In the CLPM stability of the constructs is controlled for through the inclusion of autoregressive relationships, and it is therefore often believed that the cross-lagged regression parameters obtained with this model are the most appropriate measures for studying causality in longitudinal correlational data (e.g., Deary, Allerhand, & Der, 2009; Soenens, Luyckx, Vansteekiste, Duriez, & Goossens, 2008; Wood, Maltby, Gillett, Linley, & Joseph, 2008). Specifically, it is common practice to standardize the cross-lagged regression coefficients and compare their relative strength to determine which variable has a stronger causal influence on the other (Bentler & Speckart, 1981).

The current paper forms a sequel to the warning given by Rogosa (1980), in that it will be argued that not only should we *account for stability*, but we also need to account for the *right kind of stability*. It will be shown that if stability of the constructs is to some extent of a trait-like, time-invariant nature, the inclusion of autoregressive parameters will fail to adequately control for this. As a result the estimates of the cross-lagged regression coefficients will be biased, which may lead to erroneous conclusions regarding the underlying causal pattern. This message is not novel in itself: In fact, it has been recognized repeatedly that the "omitted variable problem" may affect the estimation of the cross-lagged coefficients (e.g., Dwyer, 1983; Finkel, 1995; Heise, 1970), and diverse modeling strategies have been proposed to account for unobserved variables that influence the observed variables. However, given the popularity of the CLPM, it seems that either this warning has been lost on a large group of substantive researchers, or many researchers are simply not convinced that this could form a serious problem.

In the current paper, we therefore present a closely related alternative structural equation modeling (SEM) approach that is inspired by considering cross-lagged panel data from a multilevel perspective, implying we need to distinguish between the within-person and the between person level. We show that this alternative SEM approach can lead to very different conclusions than the traditional CLPM when considering the three major objectives of cross-lagged panel research, that is: a) whether or not variables influence each other; b) which of the variables is causally dominant; and c) what the sign of influence is. In doing so we hope to raise awareness about the limitations of the traditional CLPM, and to stimulate researchers to consider alternative SEM approaches.

This paper is organized as follows. In the first section, two models for investigating cross-lagged effects are presented: the traditional CLPM and an extension of this model based on taking a multilevel perspective. We discuss the meaning of each model, the way they predict change, and the minimum number of waves needed for identification. In the second section, we discuss four other SEM approaches that include cross-lagged relationships and discuss how these are related to the model we propose. In doing so, we sketch the broader context of the current account and point the reader in the direction of other alternatives. The third section consists of a more in-depth comparison of the traditional CLPM and the proposed alternative. In the fourth section, a modeling strategy is proposed to ensure that – if present – both forms of stability are accounted for and we illustrate this using an empirical data set. The paper ends with summarizing the most important findings of the present study, discussing the implications for longitudinal research, and providing guidelines for future cross-lagged panel research.

## Two models for studying reciprocal influences

Cross-lagged panel research is concerned with the effect of two or more variables on each other over time. To give an impression of the kinds of questions researchers have tried to tackle using the CLPM, consider the following anthology: Do maternal warmth and praise reduce internalizing and externalizing problems in children with autism (Smith, Greenberg, Mailick Seltzer, & Hong, 2008)? Is the relationship between parenting and adolescent delinquency bidirectional (Gault-Sherman, 2012)? Does gratitude foster social support or vice versa (Wood et al., 2008)? What is the direction of causality between intelligence and academic achievement (Watkins, Lei, & Canivez, 2007)? Is processing speed a foundation for successful cognitive aging (Deary et al., 2009)? What is the role of a pessimistic explanatory style on developing and maintaining social support networks in adolescents (Ciarrochi & Heaven, 2008)? What is the directional nature of the relationship between the quality of the parent-child relationship and a child's

ADHD symptoms (Lifford, Harold, & Thapar, 2008)? And – at a macro social-economic level – what is the direction of causality between intelligence and economic welfare of nations (Rindermann, 2008)?

In this section the traditional CLPM is presented, which is the most typical modeling approach for this kind of research. In addition, an alternative model is presented, which we refer to as the *random intercepts cross-lagged panel model* (RI-CLPM), that accounts for trait-like, time-invariant stability through the inclusion of a random intercept (i.e., a factor with all loadings constrained to 1). This random intercept partials out between-person variance such that the lagged relationships in the RI-CLPM actually pertain to within-person (or within-dyad) dynamics. We discuss how these models predict change, how many measurement waves are needed for identification, and how they are related to each other.

### *The CLPM*

The CLPM can be used if two or more variables have been measured at two or more occasions, and if the interest is in their influences on each other over time. Let $x$ and $y$ denote two distinct variables which were measured multiple times, and which will be analyzed with the CLPM. While this approach typically consists of modeling the covariance structure only, the means are included here as well; note however that no constraints are imposed on them, which is equivalent to analyzing the centered data.

A graphical representation of this model is given in the left panel of Figure 1 (see Appendix 1 for the corresponding SEM specification). The measurement equations can then be expressed as

$$x_{it} = \mu_t + \xi_{it} \tag{1a}$$

$$y_{it} = \pi_t + \eta_{it} \tag{1b}$$

where $\xi_{it}$ and $\eta_{it}$ represent the individual's temporal deviations from the temporal group means $\mu_t$ and $\pi_t$ respectively. These temporal deviations are modeled with the structural equations

$$\xi_{it} = \alpha_t \xi_{i,t-1} + \beta_t \eta_{i,t-1} + u_{it} \tag{1c}$$

$$\eta_{it} = \delta_t \eta_{i,t-1} + \gamma_t \xi_{i,t-1} + v_{it}. \tag{1d}$$

The autoregressive parameter $\alpha_t$ and $\delta_t$ are included to account for the stability of the constructs: The closer these autoregressive parameters are to one, the more stable the rank order of individuals is from one occasion to the next. However, even when the stability coefficients are very high, when enough time passes, the original rank order will be lost. Hence, it is not stability of a trait-like nature, and it is therefore often referred to as *temporal stability* instead (e.g., Heise, 1970).

---

Insert Figure 1 about here

---

The cross-lagged parameters $\beta_t$ and $\gamma_t$ form the key to investigating reciprocal causal effects in this model (Rogosa, 1980): Through standardizing these parameters, a comparison of the relative effects of $x$ and $y$ on each other can be made, which can then be used to determine causal predominance (Bentler & Speckart, 1981). These parameters are often interpreted in terms of *predicting* *change* (e.g., Finkel, 1995; Ribeiro et al., 2011; Rindermann, 2008). To show the reasoning behind this interpretation, we write

$$\begin{aligned} y_{it} - y_{i,t-1} &= \big(\pi_t + \eta_{it}\big) - \big(\pi_{t-1} + \eta_{i,t-1}\big) \\ &= \big(\pi_t - \pi_{t-1}\big) + \big(\delta_t - 1\big)\eta_{i,t-1} + \gamma_t \xi_{i,t-1} + v_{it}, \end{aligned} \quad (2)$$

which shows that the cross-lagged parameter $\gamma_t$ indicates the extent to which the change in $y$ can be predicted from the individual's prior deviation from the group mean on $x$ (i.e., $\xi_{i,t-1} = x_{i,t-1} - \mu_{t-1}$), while controlling for the structural change in $y$ (i.e., $\pi_t - \pi_{t-1}$), and one's prior deviation from the group mean on $y$ (i.e., $\eta_{i,t-1} = y_{i,t-1} - \pi_{t-1}$).

The CLPM is just identified with only two waves of data, which makes it an appealing modeling approach from a practical point of view: In fact, we found that 45% of the datasets published in 2012, which were used to estimate this model, consisted of only two waves of data. In Figure 2 the distribution of all 117 datasets from 2012 is given.[2] This is noteworthy, because it implies that in almost half of the applications, the parameters of the CLPM and their standard errors can be estimated, but it is not possible to evaluate whether the model provides a proper description of the actual underlying mechanism (as the model is just identified and will yield a perfect fit, which is really not meaningful).

---

Insert Figure 2 about here

---

*The RI-CLPM*

As described above, the CLPM only accounts for temporal stability through the inclusion of autoregressive parameters. This implies that in this model it is implicitly assumed that every person varies over time around the same means $\mu_t$ and $\pi_t$, and that there are no trait-like individual differences that endure. At closer consideration, this is a rather problematic assumption, as it is difficult to imagine a psychological construct – whether behavioral, cognitive, emotional or psychophysiological – that is not to some extent characterized by stable individual differences (if not for the entire lifespan, then at least for the duration of the study).

Longitudinal data can actually be thought of as multilevel data, in which occasions are nested within individuals (or other systems, like dyads). When considering this perspective, it becomes clear that we need to separate the *within-person level* from the *between-person level*. This idea motivated the development of the alternative model we present here, which can be thought of as an extension of the CLPM that accounts not only for temporal stability, but also for time-invariant, trait-like stability through the inclusion of a random intercept. This alternative model can be expressed as

$$x_{it} = \mu_t + \kappa_i + \xi^*_{it} \tag{3a}$$

$$y_{it} = \pi_t + \omega_i + \eta^*_{it} \tag{3b}$$

where $\mu_t$ and $\pi_t$ are again the temporal group means. The additional terms $\kappa_i$ and $\omega_i$ are the individual's trait-like deviations from these means: They can be thought of as latent variables or factors whose factor loadings are all constrained to 1, as in case of random intercepts in latent growth curve (LGC) modeling (with the difference that here the group means are allowed to vary freely over time). We have added an asterisk to the temporal deviation terms $\xi^*_{it}$ and $\eta^*_{it}$, to emphasize these terms are different from the individual deviation terms in the traditional CLPM: In the current model they represent the individual's temporal deviations from their expected scores (i.e., $\mu_t + \kappa_i$ and $\pi_t + \omega_i$), rather than from the group means (i.e., $\mu_t$ and $\pi_t$).

Subsequently these deviations are model as

$$\xi^*_{it} = \alpha^*_t \xi^*_{i,t-1} + \beta^*_t \eta^*_{i,t-1} + u^*_{it} \tag{3c}$$

$$\eta^*_{it} = \delta^*_t \eta^*_{i,t-1} + \gamma^*_t \xi^*_{i,t-1} + v^*_{it}, \tag{3d}$$

where the autoregressive and cross-lagged regression parameters differ from the ones in the CLPM, as indicated by the asterisks. That is, the autoregressive parameters $\alpha^*_t$ and $\delta^*_t$ do not represent the stability of the rank order of individuals from one occasion to the next, but rather the amount of within-person carry-over effect (cf., Hamaker, 2012; Kuppens, Allen, & Sheeber, 2010; Suls, Green, & Hillis, 1998): If it is positive, it implies that occasions on which a person scored above his/her expected score are likely to be followed by occasions on which he/she still scores above the expected score again, and vice versa.[3]

The main interest here is however in the cross-lagged parameters $\beta^*_t$ and $\gamma^*_t$, which indicate the extent to which the two variables influence each other. Specifically, $\gamma^*_t$ indicates the degree by which deviations from an individual's expected score on $y$ (i.e., $\eta^*_{it} = y_{it} - \{\pi_t + \omega_i\}$) can be predicted from preceding deviations from one's expected score on $x$ (i.e., $\xi^*_{i,t-1} = x_{i,t-1} - \{\mu_t + \kappa_i\}$), while controlling for the individual's deviation of the preceding expected score on $y$ (i.e., $\eta^*_{i,t-1} = y_{i,t-1} - \{\pi_{t-1} + \omega_i\}$). The cross-lagged relationships pertain to a process that takes place at the within-person level and they are therefore of key interest when the interest is in reciprocal influences over time *within* individuals or dyads. A graphical representation of this model is given in the right panel of Figure 1 (see Appendix 1 for the corresponding SEM specification).

Expressing change in the RI-CLPM, we can write

$$\begin{aligned} y_{it} - y_{i,t-1} &= \left(\pi_t + \omega_i + \eta^*_{it}\right) - \left(\pi_{t-1} + \omega_i + \eta^*_{i,t-1}\right) \\ &= \left(\pi_t - \pi_{t-1}\right) + \left(\delta^*_t - 1\right)\eta^*_{i,t-1} + \gamma^*_t \xi^*_{i,t-1} + \upsilon^*_{it}, \end{aligned} \tag{4}$$

which shows that the cross-lagged parameter indicates the extent to which the change in $y$ can be predicted from the individual's prior deviation from his/her expected score on the other variable (i.e., $\xi^*_{i,t-1} = x_{i,t-1} - \{\mu_t + \kappa_i\}$), while controlling for the structural change in $y$ (i.e., $\pi_t - \pi_{t-1}$) *and* the prior deviation from one's expected score on $y$ (i.e., $\eta^*_{i,t-1} = y_{i,t-1} - \{\pi_{t-1} + \omega_i\}$).

The expressions in Equations 2 and 4 are similar, but unless $\kappa_i$ and $\omega_i$ are zero, the CLPM predicts change from other aspects than the RI-CLPM. In fact, it is easy to see that the traditional CLPM is nested under the current model, as it can be obtained from the latter by fixing the variances and covariance of $\kappa_i$ and $\omega_i$ to zero. To compare the two models statistically, a chi-square bar test should be used, as it requires two parameters to be fixed at the boundaries of the parameter space (see for details: Stoel, Galindo Garre, Dolan, & Wittenboer, 2006).

While the CLPM requires only two waves of data, the RI-CLPM requires at least three waves of data, in which case there is 1 degree of freedom (df).[4] If the intervals are of the same size, and if we assume that the effects the variables have on each other remain stable over time, we could decide to constrain the lagged parameters over time, giving us an additional 4 df (i.e., 5 df in total). Furthermore, we could investigate whether the means can be constrained over time, such that we obtain another 4 df (resulting in 9 df in total). If on the other hand, we are not willing to make these assumptions, and we are not sure whether the effect of the time-invariant stability components $\kappa_i$ and $\omega_i$ are equal over time, we may wish to remove the constraint on the factor loadings. This relaxation may especially be of interest when the observations are made further apart in time, and we expect that we are also measuring some structural changes. However, this would imply that $\kappa_i$ and $\omega_i$ no longer represents random intercepts (as in multilevel modeling), but rather represent latent variables or traits (as common in SEM). Even more so, it would imply we need more waves of data to estimate this model.

*Conclusion*

The CLPM is nested under the RI-CLPM. The latter is an attempt to disentangle the within-person process from stable between-person differences while the former does not differentiate between these two levels that are likely to be present in the data. The question thus rises what happens if the data were generated by the RI-CLPM, but are analyzed using the CLPM: Most likely this will lead to a contamination of the estimated within-person reciprocal effects, but to obtain more insight into this matter, we need to take a closer look at the relationship between the cross-lagged parameters from both models.

However, before doing this, we consider how the RI-CLPM is connected to other longitudinal SEM approaches that include cross-lagged relationships: In doing so we aim to present a broader context for the current exposition and provide some reference points for readers already familiar with (some of) these SEM approaches.

## Relatedness to other existing SEM approaches

There are several other longitudinal SEM approaches that can be used for bivariate data and which include cross-lagged relationships. Here we consider four of these, that is: a) the Stable Trait Autoregressive Trait and State (STARTS) model (Kenny & Zautra, 2001; Kenny & Zautra, 1995); b) the Autoregressive Latent Trajectory (ALT) model (Bollen & Curran, 2006; Curran & Bollen, 2001); c) the Latent Change Score (LCS) model (Hamagami & McArdle, 2001; McArdle & Hamagami, 2001); and d) a modification of the Latent State-Trait (LST) model (Schmitt & Steyer, 1993; Steyer, Schwenkmezger, & Auer, 1990). In this section we discuss the relatedness between the RI-CLPM and these four alternatives, focussing on the substantive and methodological similarities and differences. Note that this section is decidedly not meant as an in depth evaluation of these diverse alternatives: The interested reader is referred to the included citations for further details.

### *STARTS model by Kenny and Zautra*

The STARTS model by Kenny and Zautra (2001), is also known as the Trait State Error (TSE) model (Kenny & Zautra, 1995). It allows the user to decompose observed variance into three components: a) the stable trait, which does not change; b) the autoregressive trait, which changes according to an autoregressive process; and c) the state or error, which is unique to the occasion. Originally, Kenny and Zautra (1995) included constraints over time in their model, such that the relative contributions of these three components remains stable over time, but these constraints may be relaxed if enough measurement waves are available (cf. Lucas & Donnellan, 2007).

Most applications of this model are based on univariate repeated measurements, but Kenny and Zautra (1995) also present a bivariate extension of their model. The RI-CLPM proposed in this paper differs from the bivariate STARTS model in that it does not include measurement error: The RI-CLPM can thus be thought of as a special case of the STARTS model (without the constraints on the lagged relationships over time), in which the observations are modeled without measurement error.

Clearly, the inclusion of measurement error in itself is recommendable, as we know that measurement error is likely to be present in psychological measurements. However, Kenny and Zautra (2001) indicate that the model is often difficult to estimate, and that it may require 10 or more waves of data. Cole, Martin, and Steiger (2005) performed a simulation study and concluded that the (univariate) STARTS model frequently led to improper solutions that were difficult to interpret (i.e., negative variance estimates, or problems with convergence in the form of singularity of the approximate Hessian matrix). They also discuss some of the reasons for this: For instance, when the autoregressive parameter is very close to zero, it becomes difficult to distinguish between variance that is due to measurement error, and variance that is the stochastic input of the autoregressive process. Thus, while extending the model with measurement error may be preferable from a theoretical point of view, the practical consequences (i.e., having to have many more measurement waves), make it a less attractive alternative for the traditional CLPM.

*ALT model by Curran and Bollen*

The ALT model was developed by Curran and Bollen (2001; see also Bollen & Curran, 2006), to “combine the best of two worlds”: It allows people to be characterized by their own trajectory over time (as in the LGC model), while their observations may also exhibit some carry-over effect from one occasion to the next (as in the autoregressive or simplex model). In the bivariate extension of the ALT model presented by Curran and Bollen (2001), the random effects that describe the individual trajectories may be correlated to each other across the variables (as is the case in a bivariate LGC model), and there may also be cross-lagged influences between the observations (as in the CLPM).

While this hybrid model seems to have a lot of potential, applying and interpreting the ALT model is not as straight forward as one may be inclined to think at first: Because the lagged relationships are included in this model between the observations, there is a recursiveness in the model, which has some adverse effects. First, it implies the process needs to be "started up", for which Curran and Bollen (2001) propose two solutions: Either the first observation is treated as exogenous, or nonlinear constraints are imposed on the loadings for the first occasion. While treating the first occasion as exogenous is relatively easy, Jongerling and Hamaker (2011) show that this may lead to rather unexpected growth curves: For instance, in an ALT model with a random constant only (i.e., no linear trend parameter), one may actually be modeling an increasing or decreasing trend over time. Such undesirable effects are not encountered when using the nonlinear constraints to start up the process, but these require the assumption that the lagged effects are constant over time,[5] and are more difficult to impose, especially in the bivariate case.

Second, the recursiveness in the ALT model implies that the random constant and the random change parameter no longer have the original role of individuals' intercepts and slopes (cf. Hamaker, 2005). For instance, the random constant not only affects an observation directly, but also indirectly through (all) previous occasions. Hamaker (2005) has shown that under the assumption that the lagged effects are invariant over time, the ALT model can be rewritten as a LGC model with autoregressive residuals, with the advantage that the random parameters in this reparametrization serve as the random intercept and slope that describe the underlying individuals' deterministic trends. This result has also been extended to multivariate processes, meaning that the bivariate ALT models can be rewritten as a bivariate LGC model with residuals that are characterized by autoregressive and cross-lagged regressive relationships (cf. Hamaker, 2005).

Considering this latter parametrization, the RI-CLPM is related to a bivariate ALT model with only random intercepts and no random slopes. However, in the RI-CLPM we do not constrain the mean structure, meaning that there may be changes–possibly, but not necessarily linear–over time, which are identical for all individuals. If the group means can be constrained to be equal over time, the RI-CLPM is nested under the ALT model with only a random intercept and no slope (using the parametrization proposed by Hamaker, 2005, to avoid the recursiveness in the model).

*LCS model by McArdle and Hamagami*

The LCS model, also known as the Latent Difference Score (LDS) model, was proposed by McArdle and Hamagami (2001; Hamagami & McArdle, 2001), and forms a rather general modeling framework that includes many longitudinal SEM approaches as special cases. What is characteristic of the LCS model is that latent changes (i.e., the differences scores corrected for measurement error), from one occasion to the next are modeled as a function of a *constant change parameter* and a *proportional change parameter* that depends on the preceding score: For this reason the model is also referred to as the Dual Change Score model (McArdle, 2009).

In the bivariate extension of this model, change is not only a function of a constant change parameter and the proportional change parameter, but also of the preceding score on the *other* variable. The cross-lagged paths, going from one variable to the change in the other, are referred to as *coupling parameters*, rather than cross-lagged regression parameters. The interpretation is the same however, in that significant coupling parameters imply that one variable has the tendency to impact changes in the other variable (McArdle & Grimm, 2010). But instead of comparing standardized coefficients in order to determine which variable is causally dominant, the coupling parameters are used to set up a vector field which depicts the expected changes from one occasion to the next on both variables as a function of the current state (see Boker & McArdle, 1995; McArdle, 2009; McArdle & Grimm, 2010). This plot is then used to make statements like: "The resulting flow shows a dynamic process, where scores on Non-Verbal abilities have a tendency to impact score changes on the Verbal scores, but there is no notable reverse effect." (p. 348, McArdle, 2005).

The LCS model has been extended with what has been referred to as "dynamic error", to distinguish it from measurement error (see for instance McArdle, 2001): While measurement error only affects the observation at the current occasion, dynamic error feeds forward through the lagged relationships, affecting the trajectory of the system and making it a stochastic rather than deterministic process. The RI-CLPM can be thought of as closely related to the LCS model with dynamic error, but without measurement error or a constant change parameter. However, the LCS model is characterized by a similar recursiveness as is present in the ALT model, and therefore the random intercept term, which directly affects the first latent score, also influences future occasions indirectly. Because the process is not "started up" as is done in ALT modeling, the recursiveness is not dealt with in such a way that we can ensure the process is stable in the absence of a constant change parameter. As a result, the RI-CLPM is not a special case of the LCS model, although they may be closely related in certain situations.

*The LST model by Steyer and colleagues*

The LST model was originally developed to distinguish between measurement error and the true score (i.e., a latent variable), and to further decompose the true score into a trait-like and a state-like part (Schmitt & Steyer, 1993; Steyer et al., 1990). In practice this typically implies that it is assumed that there is an underlying construct, which is measured by multiple indicators. This underlying construct at a particular occasion is referred to as the state, which is then decomposed into a trait-like part and an occasion-specific part: The trait-like part is included as a second-order factor, relating the states, which are represented by the first-order factors, to each other. The occasion-specific part is the residual part of the state factor, which was not accounted for by the trait.

The LST model has been extended with autoregressive relationships either between the state factors (introducing a similar recursiveness as exists in the ALT model and the LCS model), or between the occasion-specific components (to avoid the detrimental recursiveness in the model): The latter has been coined the Trait State Occasion (TSO) model (Cole et al., 2005). Recently, the TSO has been modified by Luhmann, Schimmack, and Eid (2011) to handle single indicator data. In this modified model, the measurement error term is omitted, the trait factor is modeled as a separate factor with free factor loadings over time (rather than a second-order factor), and second-order autoregressive relationships are included. Note that if the measurement error term had been kept (and the second-order autoregressive relationships were omitted), the model would be identical to the STARTS model.

Luhmann et al. (2011) also propose a bivariate version of the model, which includes cross-lagged regression paths between the occasion-specific components (and no second-order autoregressive relationships). The RI-CLPM can be seen as a special case of this bivariate single indicator LST model, in which the factor loadings for the traits are constrained to 1 over time. In applying this model to empirical data, Luhmann et al. focus on decomposing the variance into separate parts, as is also the goal in applying the STARTS model and the original LST model. Furthermore, they decompose the covariance between the two variables into a part accounted for by the traits, a part accounted for by the autoregressive and cross-lagged regressive relationships, and a part due to the relationship between the residuals of the occasion-specific factors.

*Conclusion*

Clearly, the models discussed above show some overlap with each other and with the RI-CLPM presented in the current paper. When considering these diverse modeling strategies, two observations seem of key importance. First, if researchers are specifically interested in *decomposing the variance* into trait-like and state-like components and the means are not of interest, the STARTS model and the models based on the LST model are most relevant. In contrast, if the interest is in *individual developmental trajectories*, the ALT model and the LCS model are more appropriate, as they are based on modeling both the mean structure and the covariance structure and allow for individuals to have their own growth curves. Second, the STARTS model, the ALT model and the LST model are most typically applied to *univariate data* (even though the original LST model uses multiple indicators); while bivariate (or multivariate) extensions are possible, they do not form the core focus and the cross-lagged regression parameters are not the key interest. In contrast, the LCS model is most typically used to investigate how two variables *influence each other* (based on the expected change described with the vector field), although it can also be applied to univariate data.

The above observations are relevant, because they help pitting the RI-CLPM against these alternatives. The main inspiration for proposing the RI-CLPM is that we want to obtain estimates of cross-lagged regression parameters that truly reflect the underlying reciprocal process that takes place at the within-person level. The model thus requires bivariate (or multivariate) data, the mean structure is not (necessarily) of interest, and the focus is on *how* (i.e., positive or negative cross-lagged coefficients), and *how much* (i.e., compare standardized absolute values of cross-lagged coefficients) the variables influence each other. Hence, because the focus is on the covariance structure rather than the mean and covariance structures, we could say that the RI-CLPM is more closely related to the STARTS model and the LST and TSO models. However, the goal is not to decompose the variance and covariance into trait-like and state-like parts, but to determine how the variables influence each other through the cross-lagged relationships at the within-person, state-like level, while controlling for trait-like differences at the between-person level. With this goal in mind, the RI-CLPM can be thought of as more closely related to the bivariate ALT model or the LCS model, although there is no inherent interest in individual developmental trajectories.

In sum, it can be stated that all models discussed in this section could serve as alternatives to the CLPM: Each model forms an attempt to separate between-person trait-like differences from the within-person reciprocal process. While some of these models include desirable properties such as measurement error and/or differences in developmental trajectories, the advantage of the RI-CLPM is that it is most closely related to the CLPM and requires only three waves of data. Since two or three waves of data are currently the norm in cross-lagged panel research, the RI-CLPM is more likely to be considered by researchers as a feasible alternative than models that require (many) more waves. In the following sections we focus on the CLPM and the RI-CLPM, but we return to the issue of other alternatives in the discussion.

## Comparing the cross-lagged parameters

Cross-lagged panel research is characterized by three major objectives: first, the aim is to determine whether the variables have a significant effect on each other; second, the question is which variable is causally dominant; and third, researchers want to know whether a variable has a positive or negative influence on the other variable. If researchers use the CLPM when the data were actually generated by the RI-CLPM, the question is whether this alters their conclusions with respect to these three objectives. In this section we focus on these issues through considering the cross-lagged regression parameters from both models analytically and in simulations.

*Analytical comparison*

In Appendix 2 we show that the standardized cross-lagged regression parameter in the CLPM from variable $x$ to variable $y$ can be expressed as a function of the parameters of the RI-CLPM, that is

$$\begin{aligned}\gamma_t \frac{\mathrm{SD}(x_{i,t-1})}{\mathrm{SD}(y_{it})} = &\left[1 - \left\{\mathrm{cov}(\omega_i, \kappa_i) + \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\right\}^2\right]^{-1} \\ &\times \Bigg[\mathrm{cov}(\omega_i, \kappa_i) + \delta^*_t \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1}) + \gamma^*_t \mathrm{var}(\xi^*_{i,t-1}) \\ &\quad - \left\{\mathrm{cov}(\omega_i, \kappa_i) + \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\right\} \\ &\quad \times \left\{\mathrm{var}(\omega_i) + \delta^*_t \mathrm{var}(\eta^*_{i,t-1}) + \gamma^*_t \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\right\}\Bigg], \end{aligned} \tag{5}$$

which shows that it is a complex function of: a) the cross-lagged regression coefficient from variable $x$ to variable $y$, that is $\gamma^*_t$; b) the within-person autoregressive parameter of variable $y$, that is $\delta^*_t$; c) the covariance between the within-person deviations at the previous time point, that is $\mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})$; d) the variance of the within-person deviation at the preceding occasion, that is $\mathrm{var}(\eta^*_{i,t-1})$; e) the variance of the trait-like component, that is $\mathrm{var}(\omega_i)$; and f) the covariance between the trait-like components, that is $\mathrm{cov}(\omega_i, \kappa_i)$.

Considering the first objective of cross-lagged panel research, that is, is there a significant effect of one variable on the other, the relationship in Equation 5 is not very informative, although it may be expected that the two models will not necessarily lead to same conclusion regarding the presence of a cross-lagged relationship.

With respect to the second objective, the question is whether the difference in absolute values of the standard cross-lagged coefficients is of the same sign across the two models. That is, the question is whether

$$|\gamma \frac{\mathrm{SD}(x_{i,t-1})}{\mathrm{SD}(y_{it})}| - |\beta \frac{\mathrm{SD}(y_{i,t-1})}{\mathrm{SD}(x_{it})}| \quad \text{and} \quad |\gamma^* \frac{\mathrm{SD}(\xi^*_{i,t-1})}{\mathrm{SD}(\eta^*_{it})}| - |\beta^* \frac{\mathrm{SD}(\eta^*_{i,t-1})}{\mathrm{SD}(\xi^*_{it})}|.$$

are either both positive, leading to the conclusion that $x$ is causally dominant, or both negative, leading to the conclusion that $y$ is causally dominant. If these differences are not of the same sign, this implies that using one model leads to the conclusion that $x$ is causally dominant, while the other model leads to the conclusion that $y$ is causally dominant. Clearly, that is not a desirable situation. For instance, when investigating the reciprocal influences of mothers' harshness and children's behavioral problems, the RI-CLPM may indicated that the mothers are causally dominant and form the driving force in this potentially negative spiral, while the CLPM may point to the children as being the instigator of maladaptive patterns. However, due to the rather complex relationships between the models' differences of absolute standardized cross-lagged parameters, it is difficult to evaluate when these models will lead to conflicting conclusions, although in general we may expect that larger trait-like differences are likely to have a stronger effect than in case of small between-person differences.

The third objective concerns the sign of the cross-lagged parameters. Thus the question is: If $\gamma^* > 0$, will $\gamma > 0$, and when $\gamma^* < 0$, will $\gamma < 0$? Naturally, the same question applies to $\beta^*$ and $\beta$. Although this is not immediately apparent from the expression in Equation 5, the many unrelated terms from the two levels strongly suggest that $\gamma^*$ and $\gamma$ not necessarily have the same sign. This is again quite disturbing, as it suggests that using the CLPM may lead to the conclusion that mothers' harshness has a damping effect on children's behavioral problems, while the RI-CLPM may indicate that mothers' harshness actually exacerbates the children's behavioral problems.

*Simulations*

In order to further investigate the effect of using the CLPM instead of the RI-CLPM with respect to the three objectives of cross-lagged panel research identified above, we performed a series of simulations based on four models. We emphasize that the models used here were handpicked, to illustrate several specific situations that can arise, and we do not claim that these are necessarily reflecting realistic scenarios. Specifically, we used Mplus (Muthén & Muthén, 1998-2012), to simulate two-wave bivariate data according to a RI-CLPM, which were subsequently used to estimate the traditional CLPM. For each model, 1000 replications were generated, of $N = 200$ each. Saving the parameter estimates in a separate file, which we then imported into R (R Core Team, 2012), we computed the standardized cross-lagged parameters (as Mplus does not allow for the computation of standardized parameters in case of Monte Carlo simulations).

In the first model, we had autoregressive parameters of .5 and no cross-lagged regression coefficients. The within-person variances of both variables was set to 1, and the covariance between the two variables was .4. Since we made sure the process was stationary (meaning the variances and covariances are stable over time; cf. Hamilton, 1994), this implies that the residual variances at the second wave were .75 and the residual covariance was .3. The between-person variances were set to 3 for each variable, and the covariance at this level was set to -2. Hence, this represents a process which is characterized by a negative correlation at the between-person level, while there is a positive correlation at the within-person level,[6] which can be seen as an instance of Simpson's paradox (cf., Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). In the upper-left panel of Figure 3, the standardized cross-lagged parameter estimates of this model are plotted. It clearly shows that the point estimates are far from the generating values (indicated by the diamond). The average $\beta$ estimate was -.118 (SD=.036, average SE=.036), and the average $\gamma$ estimate was -.120, (SD=.037, average SE=.036). Considering whether the 95% confidence intervals of these parameter estimates contained zero, we obtained coverage rates of .105 for the $\beta$ parameter, .103 for the $\gamma$ parameter, which implies that in about 90% of the cases, the CLPM would lead to the conclusion that there is at least one significant negative cross-lagged parameter, although no cross-lagged relationships were present in the model that generated the data.

Insert Figure 3 about here

The second model is based on autoregressive parameters of .5 and cross-lagged regression parameters of .3. The within-person variances were set to 1, and the within-person covariance to .5, which through the additional constraint of stationarity implies that the innovations variances were .51 and the covariance between the innovations was .03. The between-person variances were set to 2, and the between-person covariance was set to -1. In the upper-right panel of Figure 3 the standardized cross-lagged parameter estimates are plotted. Based on 1000 replications, the average $\beta$ estimate was -.003 (SD=.032, average SE=.034), and the average $\gamma$ estimate was -.003 (SD=.034, average SE=.034). Coverage rates for the 95% confidence interval containing zero were .951 and .945 respectively, indicating that in about 95% of the cases it would be concluded that these parameters are non-significantly different from zero, although there were substantial cross-lagged relationships in the model that generated the data.

The third model is based on autoregressive parameters of .5, and a cross-lagged parameter $\beta$ of -.3 (from variable $y$ to variable $x$), and a cross-lagged parameter $\gamma$ of .1 (from $x$ to $y$). The within-person variances were both set to 1, and the covariance to -.5, such that the innovation variances were .51 for the $x$-variable and .79 for the $y$-variable, while their covariance was -.29 (to ensure stationarity). The between-person variances were set to 2 and their covariance was set to 1. The standardized cross-lagged parameter estimates are given in the lower-left panel of Figure 3. It shows that while the original combination of parameter values is in the area that is characterized by a standardized $|\gamma|$ that is *smaller* than the standardized $|\beta|$, indicating that variable $y$ is causally dominant, most point estimates fall in the area in which the standardized $|\gamma|$ is actually *larger* than the standardized $|\beta|$, leading to the opposite conclusion that variable $x$ is causally dominant. The average estimate for $\beta$ is .002 (SD=.039, average SE=.040), and for $\gamma$ it is .151 (SD=.034, average SE=.033). For $\beta$ (which equalled -.3 in the generating model), the coverage rate of the 95% confidence interval containing zero was .958, which implies that in about 95% of the cases the conclusion would be that there is a nonsignficant relationship from $y$ to $x$. The coverage rate $\gamma$ (where true $\gamma$ is .1) was .010, which implies that in 90% of the cases a significant relationship from variable $x$ to $y$ would be detected. This further shows that the CLPM may result in the wrong variable being identified as being causally dominant.

Finally, in the fourth model the autoregressive parameters were set to .5, the $\beta$ parameter (from variable $y$ to$x$) to .3 and a $\gamma$ parameter of .1. The within-person variances are both 1, the covariance is .5, which in combination with the restriction of stationarity leads to the innovation variances of variables $x$ and $y$ being .72 and .60 respectively, and a covariance between the innovations of -.056. The between-person variances ware set to 3 and their covariance to -2. The standardized point estimates of the cross-lagged parameters are presented in the lower-right panel of Figure 3, showing that, while the generating cross-lagged parameters implied that variable $y$ was causally dominant, the parameter estimates almost always lead to the conclusion that variable $x$ is causally dominant. The average point estimate for $\beta$ was -.023 (SD=.037, average SE=.036), and for $\gamma$ it was -.093 (SD=.033, average SE=.033). The 95% confidence intervals included zero with a rate of .897 for $\beta$, and .192 for $\gamma$, meaning that in almost 90% of the cases we would fail to detect the relationship from variable $y$ to $x$ (which in reality was .3), while in more than 80% of the cases we would detect a significant negative relationship from variable $x$ to $y$ (which in reality was .1). This illustrates another disturbing fact: The CLPM may result in a significant estimate of a cross-lagged parameter that actually has a different sign than the corresponding cross-lagged parameter in the generating model.

*Conclusion*

While the algebraic relationship in Equation 5 shows that the cross-lagged parameters from the two models are not necessarily identical, it is not easy to see how they will differ, especially in the light of the three objectives of cross-lagged panel research. The simulations we presented here show however that the CLPM can lead to spurious results regarding all three objectives in this line of research, that is, it can be misleading with respect to: a) the *presence* of causal relationships (models 1 and 2); b) the *causal priority* of two variables (models 3 and 4); and c) the *sign* of the causal relationship (model 4).

The simulations here were designed to illustrate these specific situations, without the intention to represent typical psychological processes. The fact is that we do not know what would be typical values for the parameter of the RI-CLPM, because this is not a model that is currently used in practice. In the simulations here the between-person variance was relatively large (i.e., two or three times as large as the within-person variance), and in general it can be stated that the results from the CLPM deviated more from the generating RI-CLPM when the between-person variances increased. Furthermore, the correlation at the between-person level also influences the results, especially if it is of the opposite sign of the correlation that exists at the within-person level (i.e., in the presence of Simpson's paradox). Finally, sample size affects the variability in estimates and their standard errors (i.e., both are inversely related to sample size), but the bias resulting from estimating a model that does not distinguish between within-person dynamics and between-person trait-like differences does not vanish when sample size increases.

## Modeling strategy

To avoid the pitfall exposed above, we propose a modeling strategy that allows us to investigate whether there are trait-like, time-invariant individual differences present in the constructs that are studied, which should be accounted for through the inclusion of a random intercept. This strategy is based on the fact that the CLPM is nested under the RI-CLPM, such that if three or more waves of data are available, both models can be fitted to the data and can be compared using a chi-bar-square test for the difference in chi-squares (Stoel et al., 2006). We illustrate this strategy using data that are reported in Soenens et al. (2008), concerning the effect of diverse aspects of parenting style on depressive symptoms of adolescents and vice versa. The data were obtained from 396 students and consist of three waves, with intervals of one year, starting in the fall of the first year in college.

We begin with considering the relationship between *Parental Psychological Control* (based on items like "My parents are less friendly to me if I don't see things like they do"), and *Adolescents' Depressive Symptomatology.* First, we fit a model in which the means of each variable are constrained over time (i.e., $\mu_t = \mu$ and $\pi_t = \pi$), while the covariance structure is unconstrained: Models in which the group means do not change over time facilitate interpretation, although it time-invariant means are no prerequisite for the models considered here. The fit of this model is not satisfactory (chi-square is 13.75, 3 df, $p = .008$; RMSEA = .078; CFI = .990; SRMR = .024), and inspection of the means shows that especially the mean of *Adolescents' Depressive Symptomatology* at the first wave is higher than at the other two waves: This measurement is from the first semester that the participants are in college, and the elevated average may thus reflect the difficulties associated with getting adjusted to these new circumstances. Freeing this mean leads to appropriate model fit ($\chi^2(3) = 3.33$, $p = .344$; RMSEA= .017; CFI= 1.000; SRMR= .011). Although constraining this first mean does not affect our results for the lagged parameters in a substantive way, the results reported below are based on the model in which this first mean for Adolescents' Depressive Symptomatology is not constrained to be equal to the means at subsequent waves.

Second, we model the covariance structure using the RI-CLPM, while keeping the constraints on the means (except for the first mean of *Adolescents' Depressive Symptomatology*), and time-invariant lagged parameters. This model fits well (chi-square is 9.85, 8 df, $p = .276$; RMSEA = .024; CFI = .998; SRMR = .025). Finally, we fit the CLPM, with the same constraints on the means and lagged parameters as used in the previous model. This model does not fit well (chi-square is 66.18, 11 df, $p < .001$; RMSEA = .113; CFI = .943; SRMR = .042), although some SEM users may claim on the basis of the CFI and the SRMR that the model fits approximately. Note that since the null-model here consists of fixing two parameters on the boundary of the parameter space (i.e., two variances fixed to zero), the standard chi-square difference test will be too conservative (see Stoel et al., 2006). The chi-square difference is $66.18 - 9.85 = 56.33$, with 3 df, which is significant at an $\alpha$ of .05 (that is, $p < .01$).

To show that the substantive interpretation of the underlying process depends on the model one uses, we consider the standardized cross-lagged regression parameter estimates from both models presented in Figure 4. It shows that both models lead to significant positive cross-lagged parameters. However, while the RI-CLPM indicates that the effect of *Parental Psychological Control* on *Adolescents' Depressive Symptomatology* is only slightly larger than the reverse effect (i.e., .240 versus .212 and .265 versus .205 between wave 1 and wave 2), the CLPM leads to the conclusion that the effect of parents on adolescents is much larger than that of adolescents on their parents (i.e., .239 versus .139 and .248 versus .134 between wave 1 and wave 2). Hence, using the CLPM would lead to the conclusion that parents are causally dominant, while the RI-CLPM leads to the conclusion that the reciprocal process is much more symmetric.

---

Insert Figure 4 about here

---

We apply the same procedure for the variables *Parental Responsiveness* (based on items like “My parents make me feel better after I discussed my worries with them”), and *Adolescents’ Depressive Symptomatology*. Here, both the first mean of the adolescents’ variable, and the last mean of the parents’ variable were estimated freely, in order to obtain a fitting model (chi-square is .933, 2 df, $p = .627$; RMSEA = .000; CFI = 1.000; SRMR = .006): The last mean of *Parental Responsiveness* was significantly lower than that at the other two measurement waves, which may reflect the increasing independence of the adolescents in the third year of college. The RI-CLPM fitted well (chi-square is 11.86, 7 df, $p = .105$; RMSEA = .042; CFI = .996; SRMR = .031), while the CLPM did not lead to a well fitting model (chi-square is 76.01, 10 df, $p < .001$; RMSEA = .129; CFI = .939; SRMR = .048), although again, some SEM users may claim this model is not too far off. The chi-square difference is $76.01 - 11.86 = 64.15$, with 3 df, which is significant at an $\alpha$ of .05 (that is, $p < .01$).

Comparing the standardized lagged parameter estimates from both models given in Figure 4, the RI-CLPM leads to the conclusion that there are no reciprocal influences between *Parental Responsiveness* and *Adolescents’ Depressive Symptomatology*, whereas the CLPM leads to the conclusion that there is a significant negative effect from *Parental Responsiveness* to subsequent *Adolescents’ Depressive Symptomatology* (and while there is no significant effect from adolescents to parents, it would be concluded that parents are causally dominant here).

In conclusion, the modeling strategy illustrated above shows that it is possible to investigate whether the constructs are characterized by time-invariant, trait-like individual differences, and that using the traditional CLPM can lead to erroneous conclusions regarding the pattern of mutual influences. Hence, researchers should make sure to use an alternative that decomposes the variance into between-person differences and the within-person process. If the constructs are not characterized by time-invariant, trait-like individual differences, running the RI-CLPM will not affect the results substantially, although in that case one can also use the simpler CLPM instead.

## Discussion

Rogosa summarized his critique on the cross-lagged correlation methodology–which he referred to as CLC–saying: "CLC may indicate the absence of direct causal influence when important causal influences, balanced or unbalanced, are present. Also, CLC may indicate a causal predominance when no causal effects are present. Moreover, CLC may indicate a causal predominance opposite to that of the actual structure of the data; that is, CLC may indicate that $X$ causes $Y$ when the reverse is true." (p. 246, Rogosa, 1980). In the current paper, similar problems have been exposed in the context of the CLPM. That is, the CLPM may indicate there are reciprocal effects when these do not exist (model 1), and may fail to detect them when they do exist (model 2). Furthermore, the CLPM may identify one variable as being causally dominant, when in fact the other variable is (models 3 and 4). Finally, the CLPM may indicate a negative influence from one variable on another, while in reality the effect is positive (model 4).

The source of these problems is the failure to adequately separate the within-person and the between-person level in the presence of time-invariant, trait-like individual differences. As a result, the estimates of lagged parameters are confounded by the relationship that exists at the between-person level (see Hamaker, 2012 for other situations in which this confounding may occur). As it is reasonable to assume that most psychological constructs that are studied with cross-lagged panel designs are to some extent characterized by time-invariant stability reflecting a trait-like property (at least for the duration of the study), it follows that many lagged parameters reported in the literature will not reflect the actual within-person (causal) mechanism.

This is especially problematic if one wishes to use the results from cross-lagged panel research as a basis for future interventions. For instance, the results obtained with the traditional CLPM for adolescent depression and parental responsiveness in this paper, would lead the researcher to conclude that increasing parental responsiveness should result in a reduction in depressive symptoms on part of the adolescent; however, the RI-CLPM shows that this result is an artefact, and that there is actually no lagged effect from parental responsiveness to adolescents' depression. Note that this does not imply that the two variables are unrelated: In fact, the trait-like individual differences are negatively correlated (estimated correlation is -.443, SE= .067, $p < .001$), indicating that parents who tend to be more responsive on average, tend to have adolescents who suffer less from depressive symptomatology on average. However, we cannot derive a causal mechanism from these results, which explains this relationship and that can be used as the foundation for an intervention. This shows that "getting it right" with respect to the cross-lagged relationships is not just an academic concern.

We found that 45% of the studies that make use of the CLPM are based on only two waves of data. In these cases, the CLPM is saturated, and hence no statements regarding model fit can be made: That is, the model will always fit perfectly, and the interest in estimating this model is simply in obtaining estimates of the cross-lagged regression parameters which are corrected for the temporal stability of the constructs. This implies that to date, it is impossible to tell what portion of the results reported in the literature based on the CLPM provide truthful reflections of the actual reciprocal mechanisms, and what portion is flawed and if so, how serious these errors are.

Researchers interested in studying lagged relationships are therefore well advised to employ the following approach. First, a minimum of three measurement waves are required: Only then can the within-person process be controlled for stable between-person differences through the inclusion of a random intercept. Second, start with a model in which only the means are constrained over time, while the covariance structure is estimated freely: This allows one to determine whether there are structural changes over time. If this model proves tenable, subsequent models can be specified for the covariance structure, while leaving the means constrained over time. If the first model proves untenable however, the researcher should identify the source of misfit, and consider freeing certain means (as we did in the empirical applications included in this paper), or refrain to an alternative modeling approach, such as LGC or ALT modeling (Hamaker, 2005). If there is no need to refrain to an alternative model based on the mean structure, the researcher is advised to continue with the extended models which were proposed in this paper, and compare the CLPM to the RI-CLPM in order to determine whether the constructs are characterized by trait-like between-person differences, or that it can be assumed that all individuals vary around the same mean or trend.

In addition, if researchers expect their measurements to contain measurement error, they are advised to either obtain a large number of repeated measurements (say $> 10$) such that they can estimate a bivariate STARTS model, or to obtain multiple indicators (e.g., test halves) such that they can fit a bivariate TSO model. In both cases they will be able to distinguish between the within-person process and stable trait-like between-person differences, while controlling for measurement error.

Finally, although we restricted our focus here on an alternative model in which the effect of the between-person differences is constant over time (such that it can be modeled as a random intercept and there is a clear connection with multilevel modeling), we recognize that this will not always be a tenable assumption. That is, when observations are taken further apart in time, the effect of stable individual differences may vary over time, such that the constraints on the factor loadings should be relaxed. This implies that more waves of data are necessary, and that other alternatives such as discussed in this paper should also be considered. Our intention is not to try to convince researchers that the RI-CLPM is necessarily the best alternative for the CLPM – there will be many instances where one of the other alternatives is more suited – but rather, to raise awareness regarding the limitation of the CLPM for uncovering within-person reciprocal processes.

References

Bentler, P. M., & Speckart, G. (1981). Attitudes “cause” behaviors: A structrual equation analysis. *Journal of Personality and Social Psychology*, *40*, 226-238.

Boker, S. M., & McArdle, J. J. (1995). Statistical vector field analysis applied to mixed crosssectional and longitudinal data. *Experimental Aging Research*, *21*, 77-93.

Bollen, K. A., & Curran, P. (2006). *Latent curve models: A structural equation approach.* Hoboken, NJ: Wiley and Sons.

Burt, K. B., Obradović, J., Long, J. D., & Masten, A. S. (2008). The interplay of social competence and psychopathology over 20 years: Testing transactional and cascade models. *Child Development*, *79*, 359-374.

Ciarrochi, J., & Heaven, P. C. L. (2008). Learned social hopelessness: The role of explanatory style in predicting social support during adolescence. *Journal of Child Psychology and Psychiatry*, *49*, 1279-1286.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods*, *10*, 3-20.

Cole, D. A., Nolen-Hoeksma, S., Girgus, J., & Paul, G. (2006). Stress exposure and stress generation in child and adolescent depression: A latent trait-state-error approach to longitudinal analyses. *Journal of Abnormal Psychology*, *115*, 40-51.

Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 105–136). Washington DC:

American Psychological Association.

Deary, I. J., Allerhand, M., & Der, G. (2009). Smarter in middle age, faster in old age: A cross-lagged panel analysis of reaction time and cognitive ability over 13 years in the West of Scotland Twenty-07 Study. *Psychology and Aging*, *24*, 40-47.

Dwyer, J. H. (1983). *Statistical models for the social and behavioral sciences.* Oxford: Oxford University Press.

Erickson, D. J., Wolfe, J., King, D. W., King, L. A., & Sharkansky, E. J. (2001). Posttraumatic stress disorder and depression symptomatology in a sample of gulf war veterans: A prospective analysis. *Journal of Consulting and Clinical Psychology*, *69*, 41-49.

Finkel, S. E. (1995). *Causal analysis with panel data.* Thousand Oaks, CA: Sage.

Gault-Sherman, M. (2012). It's a two-way street: The bidirectional relationship between parenting and delinquency. *Journal of Youth and Adolescence*, *41*, 121-145.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, *37*, 424-438.

Green, B. L., Furrer, C. J., & McAllister, C. L. (2011). Does attachment style influence social support or the other way around? A longitudinal study of early Head Start mothers. *Attachment & Human Development*, *13*, 27-47.

Hamagami, F., & McArdle, J. J. (2001). Advanced studies of individual differences: Linear dynamic models for longitudinal data analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (p. 203-246). Mahwah, NY: Lawrence Erlbaum

Associations.

Hamaker, E. L. (2005). Conditions for the equivalence of the autoregressive latent trajectory model and a latent growth curve model with autoregressive disturbances. *Sociological Methods and Reasearch*, *33*, 404-418.

Hamaker, E. L. (2012). Why researchers should think "within-person" a paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (p. 43-61). New York, NY: Guilford Publications.

Hamilton, J. D. (1994). *Time series analysis.* Princeton, NJ: Princeton University Press.

Heise, D. R. (1970). Causal inference from panel data. *Sociological Methodology*, *2*, 3-27.

Jongerling, J., & Hamaker, E. L. (2011). On the trajectories of the predetermined ALT model: What are we really modeling? *Structural Equation Modeling*, *184*, 370-382.

Kenny, D. A., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of Consulting and Clinical Psychology*, *63*, 52–59.

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (p. 243-263). Washington DC: American Psychological Association.

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpsons paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, 513.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984-991.

Lifford, K. J., Harold, G. T., & Thapar, A. (2008). Parent-child relationships and ADHD symptoms: A longitudinal analysis. *Journal of Abnormal Child Psychology*, *36*, 285-296.

Lindwall, M., Larsman, P., & Hagger, M. S. (2011). The reciprocal relationship between physical activity and depression in older european adults: A prospective cross-lagged panel design using SHARE data. *Health Psychology*, *30*, 453-462.

Lucas, R. E., & Donnellan, M. B. (2007). How stable is happiness? using the STARTS model to estimate the stability of life satisfaction. *Journal of Research in Personality*, *41*, 1091-1098.

Luhmann, M., Schimmack, U., & Eid, M. (2011). Stability and variability in the relationship between subjective well-being and income. *Journal of Research in Personality*, *45*, 186-197.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sórbom (Eds.), *Structural equation modeling: Present and future* (p. 342-380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J. (2005). Five steps in latent curve modeling with longitudinal life-span data. *Advances in Life Course Research,*, *10*, 315-357.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577-605.

McArdle, J. J., & Grimm, K. J. (2010). Five steps in latent curve and latent change score modeling with longitudinal data. In K. van Montfort, J. Oud, &

A. Satorra (Eds.), *Longitudinal research with latent variables* (p. 245-274). Heidelberg, Germany: Springer-Verlag.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (p. 139-175). Wahington, DC: American Psychological Association.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide. Seventh edition.* Los Angeles, CA: Muthén & Muthén.

R Core Team. (2012). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Available from `http://www.R-project.org/` (ISBN 3-900051-07-0)

Ribeiro, L. A., Zachrisson, H. D., Schjolberg, S., Aase, H., Rohrer-Baumgartner, N., & Magnus, P. (2011). Attention problems and language development in preterm low-birth-weight children: Cross-lagged relations from 18 to 36 months. *BMC Pediatrics*, *11*, 59-70.

Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence*, *36*, 127-142.

Rogosa, D. R. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, *88*, 245–258.

Schmitt, M. J., & Steyer, R. (1993). The latent state-trait model (not only) for social desirability. *Personality and Individual Differences*, *14*, 519-529.

Smith, L. E., Greenberg, J. S., Mailick Seltzer, M., & Hong, J. (2008). Symptoms and behavior problems of adolescents and adults with autism: Effects of mother-child relationship quality, warmth, and praise. *American Journal of*

*Mental Retardation*, *113*, 387-402.

Soenens, B., Luyckx, K., Vansteekiste, M., Duriez, B., & Goossens, L. (2008). Clarifying the link between parental psychological control and adolescents' depressive symptoms: Reciprocal versus unidirectional models. *Merrill-Palmer Quarterly*, *54*, 411-444.

Steyer, R., Schwenkmezger, P., & Auer, A. (1990). The emotional and cognitive components of trait anxiety: A latent state-trait model. *Personality and Individual Differences*, *11*, 125-134.

Stoel, R., Galindo Garre, F., Dolan, C., & Wittenboer, G. van den. (2006). On the likelihood ratio test in structural equation modelling when parameters are subject to boundary constraints. *Psychological Methods*, *11*, 439-455.

Suls, J., Green, P., & Hillis, S. (1998). Emotional reactivity to everyday problems, affective inertia, and neuroticism. *Personality and Social Psychology Bulletin*, *24*, 127-136.

Watkins, M. W., Lei, P.-W., & Canivez, G. L. (2007). Psychometric intelligence and achievement: A cross-lagged panel analysis. *Intelligence*, *35*, 59-68.

Wood, A. M., Maltby, J., Gillett, R., Linley, P. A., & Joseph, S. (2008). The role of gratitude in the development of social support, stress, and depression: Two longitudinal studies. *Journal of Research in Personality*, *42*, 854-871.

# Appendix 1

Specifying a CLPM for three occasions can be done with the *measurement equation*

$$\begin{bmatrix} x_{i1} \\ y_{i1} \\ x_{i2} \\ y_{i2} \\ x_{i3} \\ y_{i3} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \pi_1 \\ \mu_2 \\ \pi_2 \\ \mu_3 \\ \pi_3 \end{bmatrix} + \begin{bmatrix} \xi_{1i} \\ \eta_{1i} \\ \xi_{2i} \\ \eta_{2i} \\ \xi_{3i} \\ \eta_{31} \end{bmatrix}, \tag{6a}$$

and *structural equation*

$$\begin{bmatrix} \xi_{1i} \\ \eta_{1i} \\ \xi_{2i} \\ \eta_{2i} \\ \xi_{3i} \\ \eta_{31} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_2 & \beta_2 & 0 & 0 & 0 & 0 \\ \gamma_2 & \delta_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_3 & \beta_3 & 0 & 0 \\ 0 & 0 & \gamma_3 & \delta_3 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_{1i} \\ \eta_{1i} \\ \xi_{2i} \\ \eta_{2i} \\ \xi_{3i} \\ \eta_{31} \end{bmatrix} + \begin{bmatrix} \xi_{1i} \\ \eta_{1i} \\ u_{2i} \\ v_{2i} \\ u_{3i} \\ v_{31} \end{bmatrix}, \tag{6b}$$

where the covariance matrix of the latter residual vector is

$$\mathbf{\Psi} = \begin{bmatrix} \sigma^2_{x_1} & & & & & \\ \sigma_{x_1 y_1} & \sigma^2_{y_1} & & & & \\ 0 & 0 & \sigma^2_{u_2} & & & \\ 0 & 0 & \sigma_{u_2 v_2} & \sigma^2_{v_2} & & \\ 0 & 0 & 0 & 0 & \sigma^2_{u_3} & \\ 0 & 0 & 0 & 0 & \sigma_{u_3 v_3} & \sigma^2_{v_3} \end{bmatrix}. \tag{6c}$$

Note that the variances and covariance between $\xi_{i1}$ and $\eta_{i1}$ are identical to those of $x_{i1}$ and $y_{i1}$ in this model.

Specifying the RI-CLPM for three waves of data in SEM software is based on the *measurement equation*

$$\begin{bmatrix} x_{i1} \\ y_{i1} \\ x_{i2} \\ y_{i2} \\ x_{i3} \\ y_{i3} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \pi_1 \\ \mu_2 \\ \pi_2 \\ \mu_3 \\ \pi_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \xi^*_{1i} \\ \eta^*_{1i} \\ \xi^*_{2i} \\ \eta^*_{2i} \\ \xi^*_{3i} \\ \eta^*_{31} \\ \kappa_i \\ \omega_i \end{bmatrix}, \tag{7a}$$

and *structural equation*

$$\begin{bmatrix} \xi^*_{1i} \\ \eta^*_{1i} \\ \xi^*_{2i} \\ \eta^*_{2i} \\ \xi^*_{3i} \\ \eta^*_{3i} \\ \kappa_i \\ \omega_i \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha^*_2 & \beta^*_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma^*_2 & \delta^*_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha^*_3 & \beta^*_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & \gamma^*_3 & \delta^*_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi^*_{1i} \\ \eta^*_{1i} \\ \xi^*_{2i} \\ \eta^*_{2i} \\ \xi^*_{3i} \\ \eta^*_{3i} \\ \kappa_i \\ \omega_i \end{bmatrix} + \begin{bmatrix} \eta^*_{1i} \\ \xi^*_{1i} \\ u^*_{2i} \\ v^*_{2i} \\ u^*_{3i} \\ v^*_{3i} \\ \kappa_i \\ \omega_i \end{bmatrix}, \tag{7b}$$

where the covariance matrix of the latter residual vector is

$$\boldsymbol{\Psi} = \begin{bmatrix} \sigma^2_{\xi^*_1} & & & & & & & \\ \sigma_{\xi^*_1\eta^*_1} & \sigma^2_{\eta^*_1} & & & & & & \\ 0 & 0 & \sigma^2_{u^*_2} & & & & & \\ 0 & 0 & \sigma_{u^*_2v^*_2} & \sigma^2_{v^*_2} & & & & \\ 0 & 0 & 0 & 0 & \sigma^2_{u^*_3} & & & \\ 0 & 0 & 0 & 0 & \sigma_{u^*_3v^*_3} & \sigma^2_{v^*_3} & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma^2_{\kappa} & \\ 0 & 0 & 0 & 0 & 0 & 0 & \sigma_{\kappa,\omega} & \sigma^2_{\omega} \end{bmatrix}. \tag{7c}$$

Note that in contrast to the previous model, here the variances and covariance of $\xi^*_{1i}$ and $\eta^*_{1i}$ are not identical to those of $x_{i1}$ and $y_{i1}$ (unless $\kappa_i = \omega_i = 0$ for all $i$).

# Appendix 2

The standardized cross-lagged parameters in the traditional CLPM can be expressed as partial correlations (e.g., Heise, 1970). Focussing on the cross-lagged parameter $\gamma_t$ from $\xi_{i,t-1}$ to $\eta_{it}$, and making use of the fact that $\xi_{i,t}$ and $\eta_{it}$ are the group mean centered variables$x_{it}$ and $y_{it}$, we can write

$$\gamma_t \frac{\sigma(x_{i,t-1})}{\sigma(y_{it})} = \frac{\rho(x_{i,t-1}y_{it}) - \rho(y_{i,t-1}x_{i,t-1})\rho(y_{i,t-1}y_{it})}{1-\rho(y_{i,t-1}x_{i,t-1})^2}. \tag{8}$$

In order to see how the cross-lagged parameter $\gamma$ from the traditional CLPM is related to the cross-lagged parameters $\gamma^*$ of the RI-CLPM, we need to express the correlations used on the righthand side of Equation 8 in terms of the parameters of the latter model. If we assume that all the observed variables are standardized, the correlation between $y_{i,t-1}$ and $y_{it}$ can be expressed as

$$\begin{aligned}
\rho(y_{i,t-1}y_{it}) &= \mathrm{E}\Big[\big\{\omega_i + \eta^*_{i,t-1}\big\}\big\{\omega_i + \eta^*_{it}\big\}\Big] \\
&= \mathrm{E}\big[\omega_i^2\big] + \mathrm{E}\big[\eta^*_{i,t-1}\eta^*_{it}\big] \\
&= \mathrm{var}(\omega_i) + \mathrm{E}\Big[\eta^*_{i,t-1}\big\{\delta^*_t\eta^*_{i,t-1} + \gamma^*_t\xi^*_{i,t-1} + v^*_{it}\big\}\Big] \\
&= \mathrm{var}(\omega_i) + \mathrm{E}\big[\delta^*_t\eta^{*2}_{i,t-1}\big] + \mathrm{E}\big[\gamma^*_t\eta^*_{i,t-1}\xi^*_{i,t-1}\big] \\
&= \mathrm{var}(\omega_i) + \delta^*_t\mathrm{var}(\eta^*_{i,t-1}) + \gamma^*_t\mathrm{cov}(\eta^*_{i,t-1},\xi^*_{i,t-1}), \qquad (9)
\end{aligned}$$

while the correlation between $y_{i,t-1}$ and $x_{i,t-1}$ can be expressed as

$$\begin{aligned}
\rho_{y1x1} &= \mathrm{E}\Big[\big\{\omega_i + \eta^*_{i,t-1}\big\}\big\{\kappa_i + \xi^*_{i,t-1}\big\}\Big] \\
&= \mathrm{E}\big[\omega_i\kappa_i\big] + \mathrm{E}\big[\eta^*_{i,t-1}\xi^*_{i,t-1}\big] \\
&= \mathrm{cov}(\omega_i,\kappa_i) + \mathrm{cov}(\eta^*_{i,t-1},\xi^*_{i,t-1}), \qquad (10)
\end{aligned}$$

and the correlation between $y_{it}$ and $x_{i,t-1}$ can be expressed as

$$\begin{aligned}
\rho(x_{i,t-1}y_{it}) &= \mathrm{E}\Big[\big\{\kappa_i + \xi^*_{i,t-1}\big\}\big\{\omega_i + \eta^*_{it}\big\}\Big] \\
&= \mathrm{E}\big[\kappa_i\omega_i\big] + \mathrm{E}\big[\xi^*_{i,t-1}\eta^*_{it}\big] \\
&= \mathrm{cov}(\omega_i, \kappa_i) + \mathrm{E}\Big[\xi^*_{i,t-1}\big\{\delta^*_t\eta^*_{i,t-1} + \gamma^*_t\xi^*_{i,t-1} + v^*_{it}\big\}\Big] \\
&= \mathrm{cov}(\omega_i, \kappa_i) + \mathrm{E}\big[\delta^*_t\xi^*_{i,t-1}\eta^*_{i,t-1}\big] + \mathrm{E}\big[\gamma^*_t\xi^{*2}_{i,t-1}\big] \\
&= \mathrm{cov}(\omega_i, \kappa_i) + \delta^*_t\mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1}) + \gamma^*_t\mathrm{var}(\xi^*_{i,t-1}) \qquad (11)
\end{aligned}$$

Using these expressions for the correlations in Equation 8, we can now write

$$\begin{aligned}
\gamma_t\frac{SD(x_{i,t-1})}{SD(y_{it})} = &\frac{\mathrm{cov}(\omega_i, \kappa_i) + \delta^*_t\mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1}) + \gamma^*_t\mathrm{var}(\xi^*_{i,t-1})}{1 - \big\{\mathrm{cov}(\omega_i, \kappa_i) + \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\big\}^2} \\
&- \frac{\big\{\mathrm{cov}(\omega_i, \kappa_i) + \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\big\}\big\{\mathrm{var}(\omega_i) + \delta^*_t\mathrm{var}(\eta^*_{i,t-1}) + \gamma^*_t\mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\big\}}{1 - \big\{\mathrm{cov}(\omega_i, \kappa_i) + \mathrm{cov}(\eta^*_{i,t-1}, \xi^*_{i,t-1})\big\}^2}.
\end{aligned}$$

Similarly, the relationship between $\beta_t$ and $\beta^*_t$ can be derived.

Footnotes

[1]While the omitted variable problem implies that we cannot make strong causal statements based on correlational data, it does not prohibit the use of the concept of *Granger causality* (Granger, 1969). However, many researchers using cross-lagged regression refrain from using the term *causal*, and use terms like *reciprocal relationship* (Erickson, Wolfe, King, King, & Sharkansky, 2001; Lindwall, Larsman, & Hagger, 2011), *role* (Ribeiro et al., 2011), *cross-domain effects* (Burt, Obradović, Long, & Masten, 2008), *exposure* (Cole, Nolen-Hoeksma, Girgus, & Paul, 2006), *impact* (Gault-Sherman, 2012), or *influence* (Green, Furrer, & McAllister, 2011), instead. It may be argued however, that these alternative terms also imply a causal mechanism, and even more so, that an interest in causality is actually the driving force behind these studies. Therefore, we decided to use the terms *causal* and *causality* in the current paper, although we acknowledging that strong causal statements can only be based on experimental designs, and we should confine ourselves to the concept of Granger causality.

[2]We used PsychINFO and searched for peer reviewed papers that appeared in 2012 and which made reference to the term “cross-lagged” in either the title, the abstract or the key words. We found 115 peer reviewed publications of which two were on time series analysis, one on multilevel modeling, and one did not include an application. The 111 remaining publications reported on 117 datasets.

[3]One could also say these autoregressive parameters indicate the stability of the rank-order of individual deviations, but this is less appealing from a substantive viewpoint.

[4]The number of observed statistics in the covariance matrix is equal to

(6*7)/2=21, while the number of parameters for the covariance structure equals 20, that is: 2 variance and 1 covariance for the between-person structure (i.e., the random intercepts), 2 variances and 1 covariance for the first occasion at the within-person level, 4 lagged parameters for the first interval, 4 lagged parameters for the second interval, 2 residual variances and 1 residual covariance at the second occasion at the within-person level, and 2 residual variances and 1 residual covariance at the third occasion at the within-person level.

[5] Actually, one only has to assume the lagged relationships were invariant before the observations started, which is rather abstract when considering the model as a local description instead of an everlasting truth; hence, this is not a very restrictive assumption in practice.

[6] A possible example could be the relationship between number of words typed per minute and the number of typos: At the within-person level there is a positive relationship, as a person tends to make more mistakes when (s)he types faster, while at the between-person level there is a negative relationship as people who have more experience tend to type faster while making fewer mistakes, and vice versa.

Figure Captions

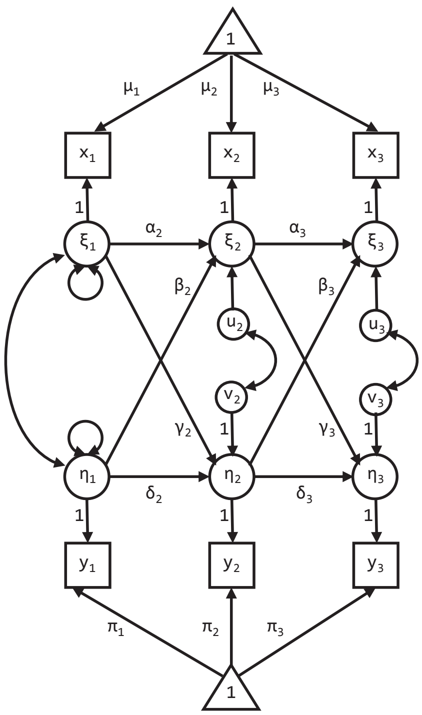
*Figure 1.* Two bivariate models for three waves of data: the standard CLPM, and the alternative RI-CLPM. Squares denote observed variables; circles represent latent variables; triangles represent means.

*Figure 2.* Histogram of number of waves per data set from 111 peer reviewed publications referring to cross-lagged research in 2012.

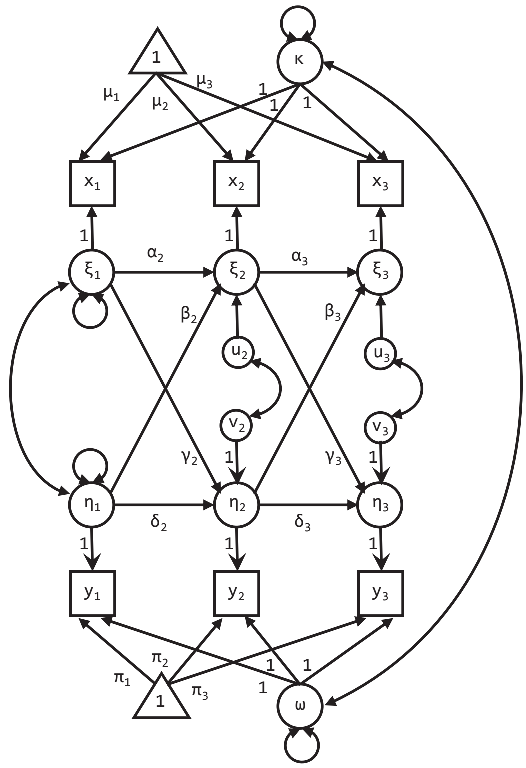
*Figure 3.* Standardized cross-lagged parameter estimates obtained with the traditional CLPM. Generating values from the RI-CLPM are denoted by the diamond. Areas A indicate solutions in which $|\beta| < |\gamma|$ such that variable $x$ is causally dominant; areas B indicate solutions in which $\beta| > |\gamma|$ such that variable $y$ is causally dominant. Only 250 estimates (of the 1000 replications) per model are plotted for reasons of clarity.

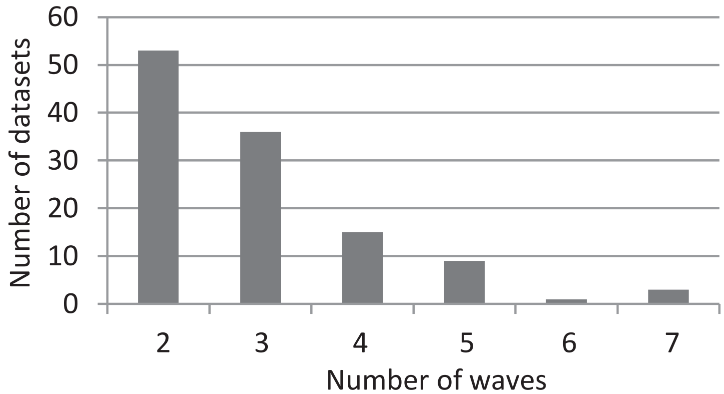
*Figure 4.* Standardized parameter estimates for Soenens data obtained with the RI-CLPM (above the arrows) and the CLPM (below the arrows). Standard errors are given between parentheses. $^*$ indicates significant at $\alpha = .05$; $^{**}$ indicates significant at $\alpha = .01$; $^{***}$ indicates significant at $\alpha = .001$.

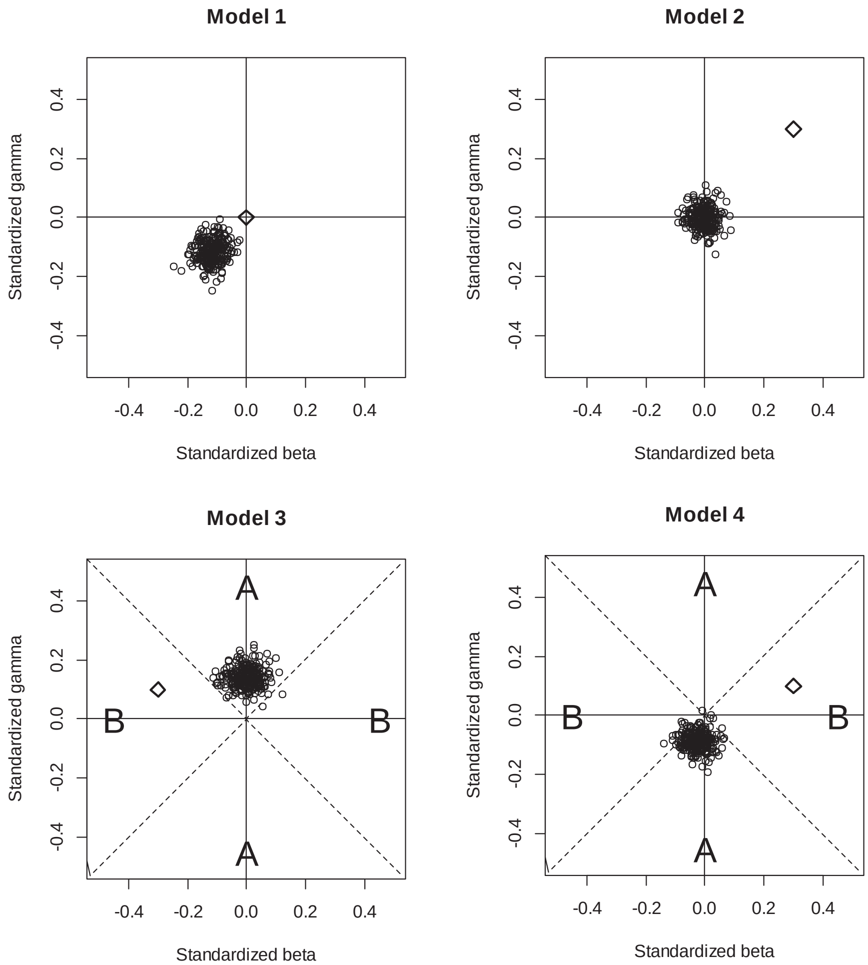
CLPM with means

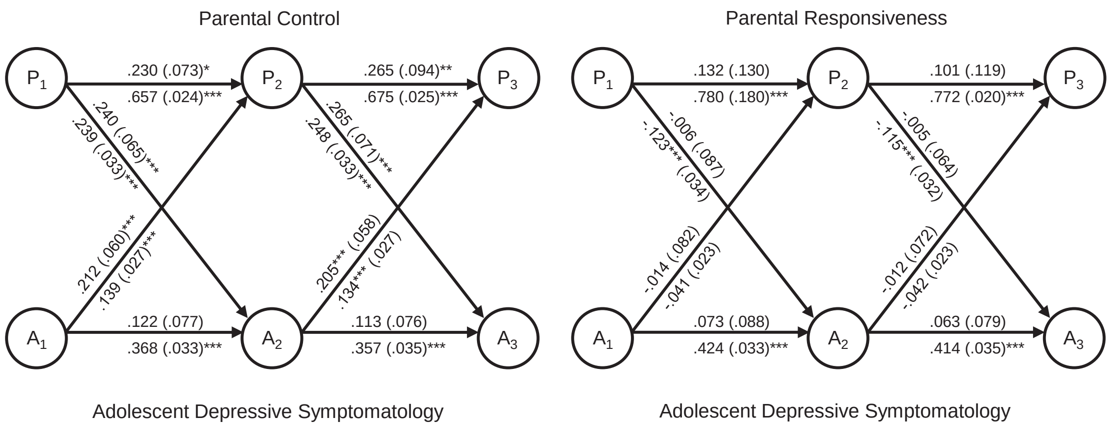



RI-CLPM with means




Number of datasets
60
50
40
30
20
10
0
2
3
4
5
6
7
Number of waves


Model 1
Model 2
Model 3
Model 4
Standardized gamma
Standardized beta
-0.4
-0.2
0.0
0.2
0.4
A
B


Parental Control
$P_1$
$P_2$
$P_3$
.230 (.073)*
.657 (.024)***
.265 (.094)**
.675 (.025)***
.240 (.065)***
.239 (.033)***
.265 (.071)***
.248 (.033)***
.212 (.060)***
.139 (.027)***
.205*** (.058)
.134*** (.027)
$A_1$
$A_2$
$A_3$
.122 (.077)
.368 (.033)***
.113 (.076)
.357 (.035)***
Adolescent Depressive Symptomatology
Parental Responsiveness
$P_1$
$P_2$
$P_3$
.132 (.130)
.780 (.180)***
.101 (.119)
.772 (.020)***
-.006 (.087)
-.123*** (.034)
-.005 (.064)
-.115*** (.032)
-.014 (.082)
-.041 (.023)
-.012 (.072)
-.042 (.023)
$A_1$
$A_2$
$A_3$
.073 (.088)
.424 (.033)***
.063 (.079)
.414 (.035)***
Adolescent Depressive Symptomatology