Effectiveness of the Deterministic and Stochastic Bivariate Latent Change Score Models for Longitudinal Research

Pablo F. Cáncer and Eduardo Estrada[*]

Department of Social Psychology and Methodology, Universidad Autónoma de Madrid,

(Spain)

*Correspondence should be sent to: eduardo.estrada.rs@gmail.com

Orcid: Pablo F. Cáncer (0000-0001-9279-8440) and Eduardo Estrada (0000-0003-0899-4057)

Abstract

The Bivariate Latent Change Score (BLCS) model is a popular framework for the study of dynamics in longitudinal research. Despite its popularity, there is little evidence of the ability of this model to recover latent dynamics when the latent trajectories are affected by stochastic innovations (i.e., dynamic error). The deterministic specification of the BLCS model does not account for the effect of these innovations in the system. In contrast, the stochastic specification of the BLCS model includes parameters that capture the effect of such innovations at the latent level. Through Monte Carlo simulation, we generated two developmental processes and examined the recovery of the parameters in the deterministic and stochastic BLCS models under a broad range of empirically relevant conditions. Based on our findings, we provide specific guidelines and recommendations for the application of BLCS models in developmental research.

Keywords: latent change score model; stochastic dynamical systems; stochastic innovations; structural equation models; longitudinal data analysis

Effectiveness of the Deterministic and Stochastic Bivariate Latent Change

Score Models for Longitudinal Research

Latent Change Score (LCS) models (Hamagami & McArdle, 2001; McArdle, 2001, 2009) are a useful and popular approach to the study of dynamics in longitudinal data. In particular, the bivariate version of these models (BLCS) allows to examine the interrelations between two variables that unfold over time (McArdle, 2001), allowing for a better understanding of the dynamic and multivariate nature of developmental processes.

BLCS models are characterized by their ability to simultaneously describe latent trajectories and time-lagged (i.e., auto-regressive) relations between latent variables over time. The key feature of these models is that they represent the processes of interest as dynamical systems in which the changes, instead of the levels, are the focus and are modeled as latent variables. Specifically, these models include, at every time point t, and for each process under study, one latent variable that captures the changes between t and $t-1$. BLCS models are particularly useful for the study of processes in which the average scores are expected to grow or decline over time (i.e., the mean structure is non-stationary), such as the development of cognitive abilities over the life span (e.g., McArdle et al., 2002). As such, they have been used to study change in numerous empirical constructs, including depressive symptoms and perceptual speed in aging (Bielak et al., 2011), biometric genetic influences in fluid intelligence in twins (Finkel et al., 2013), or the interplay between reading and writing skills in children (Ahmed et al., 2014), among many others.

Despite the popularity of BLCS models, there is little evidence on their ability to recover the dynamics of bivariate systems when the latent trajectories are affected by

stochastic innovations. From a substantive point of view, innovations represent events that impact the change at the latent level, and whose influence lingers on later states of the system (see Schuurman et al., 2015). Consider, for example, the developmental trajectories of reading and arithmetic abilities from childhood to early adulthood. The learning and development of these skills can be affected, either positively or negatively, by multiple events, such as an interpersonal conflict, changing school, or attending support classes, among many others. The impact of these events is constant during development, and they can affect each individual differently, leading to deviations in their developmental trajectories that may be relevant for the phenomenon under study. In longitudinal research, such events are typically modelled as *random shocks* or *stochastic innovations*, representing deviations from the expected latent trajectories with a lingering influence on the system.

The specification of BLCS models that includes innovation parameters is typically referred to as the *stochastic* BLCS model (Ji & Chow, 2019). However, the vast majority of specifications of this model in the Psychology literature are *deterministic* (i.e., they do not account for stochastic innovations). Thus, they assume that the changes in a system are perfectly predicted by the latent states at the previous occasion. Given the frequency with which we are exposed to random shocks such as those described above, the deterministic assumption seems unlikely in the empirical practice. If individuals in a sample are affected by stochastic innovations during their development, the stochastic model is more theoretically plausible than its deterministic counterpart, which will be misspecified. In practice, however, the choice of one model specification over the other is not obvious. Is the deterministic model incapable of recovering bivariate dynamics in the presence of stochastic innovations? If the unaccounted deviations in the latent trajectories were captured as measurement errors,

the misspecification would not be particularly harmful. In contrast, if the parameters capturing latent dynamics were affected by the random shocks, the deterministic model would lead to potentially incorrect conclusions regarding latent dynamics and interrelations between processes. And more importantly, how common are stochastic innovations in the context of developmental research? Due to the scarcity of empirical applications of stochastic BLCS models, there is little evidence on the extent to which developmental trajectories are affected by innovations. In the context of child development, Ferrer and colleagues (Ferrer et al., 2007, 2010) studied the development of reading abilities using a stochastic BLCS model, and found that individual differences in the trajectories due to innovations represented around 5 to 15% of the variance of the initial latent scores. In empirical constructs other than reading abilities, however, the proportion of between-individual variance due to innovations may be above that range. One of the main goals of this manuscript is to evaluate the ability of the deterministic and stochastic BLCS models to capture latent dynamics in the presence of stochastic innovations similar to those found in the empirical literature.

Another relevant issue in the application of BLCS models concerns sampling requirements. In fact, several recent studies have expressed the need for comprehensive simulation studies to evaluate the quality of the estimates under different combinations of sample size and number of repeated measures (e.g., Ji & Chow, 2019; Kievit et al., 2018). Research designs with relatively large samples and few repeated measures are the most frequent in developmental studies, in which BLCS models are typically used (e.g., Ahmed et al., 2014; Bielak et al., 2011; Finkel et al., 2016; Gerstorf et al., 2007; Ghisletta et al., 2006; King et al., 2006; Liao et al., 2018; Quinn et al., 2015; Snitz et al., 2015; Ziegler et al., 2015). In this type of research, collecting large samples can be challenging, and using many repeated measures may not be feasible due to time,

logistic, and financial constraints. It is therefore of great interest to examine the performance of both the deterministic and stochastic BLCS model under different sampling conditions, especially those more accessible to most empirical researchers: few repeated measures and varying sample sizes.

The **aim of the present study** is two-fold. First, we aim to examine the extent to which a deterministic BLCS model (i.e., with no parameters to capture stochastic innovations) yields reliable estimates when the trajectories are affected by stochastic innovations. Second, we aim to conduct a comprehensive evaluation of the ability of the deterministic and stochastic BLCS models to capture the dynamics of a bivariate system under a broad range of sampling conditions. In the following section, we provide a detailed description of the BLCS model, and elaborate on the interpretation of stochastic innovations. Next, we analyze the results of a Monte Carlo study and evaluate the performance of the deterministic and stochastic specifications of the BLCS model under multiple conditions of variance due to innovations, sample size, and number of repeated measures. We conclude the article offering several recommendations on the use of BLCS models and the design of longitudinal studies for developmental research.

The deterministic Bivariate Latent Change Score model

In the standard specification of BLCS models, the observed variables are decomposed into latent true scores and measurement error. Thus, observed scores for variables X and Y at any given time t are given by:

$$
\begin{aligned}
X_{[t]} &= x_{[t]} + e_{x[t]} \\
Y_{[t]} &= y_{[t]} + e_{y[t]}
\end{aligned}
\tag{1}
$$

where e_x and e_y are measurement errors, and x and y are the corresponding latent variables. These errors are stationary, normally distributed with mean zero, variances σ_{ex}^2 and σ_{ey}^2, and covariance $\sigma_{ex,ey}$ at any given time point. In this specification, no further covariances are allowed between the errors and other elements in the system (McArdle & Hamagami, 2001, 2004).

BLCS models represent a bivariate dynamic process in which the state of the system at each time point is dependent on previous states. Therefore, it requires the specification of the state of the system at the first measurement occasion (sometimes called "latent intercept"), and also the specification of the time-lagged effects defining the trajectories' change over time. Initial conditions are defined by: (a) the latent initial scores (x_0 and y_0), representing the state of the latent processes at the first measurement occasion; and (b) the latent additive components (x_a and y_a), representing a constant amount of change added at each measurement occasion. This latent structure follows a multivariate normal distribution with mean vector and covariance matrix:

$$
\begin{bmatrix} x_0 \\ x_a \\ y_0 \\ y_a \end{bmatrix} \sim N \left(\mu = \begin{bmatrix} \mu_{x0} \\ \mu_{xa} \\ \mu_{y0} \\ \mu_{ya} \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_{x0}^2 & & & \\ \sigma_{x0,xa} & \sigma_{xa}^2 & & \\ \sigma_{x0,y0} & \sigma_{xa,y0} & \sigma_{y0}^2 & \\ \sigma_{x0,ya} & \sigma_{xa,ya} & \sigma_{y0,ya} & \sigma_{ya}^2 \end{bmatrix} \right) \tag{2}
$$

Once the initial state of the system is specified, two latent variables are defined to capture changes between true scores at adjacent time points. Thus, scores for each process at time t are a function of the respective true scores at time $t-1$ plus the latent changes:

$$
\begin{aligned}
\Delta x_{[t]} &= x_{[t]} - x_{[t-1]} \\
\Delta y_{[t]} &= y_{[t]} - y_{[t-1]}
\end{aligned}
\tag{3}
$$

In the deterministic BLCS model, each of these latent change scores $\Delta x_{[t]}$ and $\Delta y_{[t]}$ are typically modelled, at any time t, as a function of three components: (a) the additive components (either x_a or y_a), representing an additive linear effect on the system; (b) a self-feedback parameter β, representing the influence of the same variable at the previous occasion, $t - 1$; and (c) a coupling parameter γ, representing the influence of the other variable at $t - 1$. Therefore, the equations for change at time t are expressed as:

$$\Delta x_{[t]} = \alpha_x \times x_a + \beta_x \times x_{[t-1]} + \gamma_y \times y_{[t-1]}$$
$$\Delta y_{[t]} = \alpha_y \times y_a + \beta_y \times y_{[t-1]} + \gamma_x \times x_{[t-1]}$$

(4)

where x_a and y_a influence the system through the coefficients α_x and α_y. The latter two coefficients can be freely estimated to express different amounts of additive change at each time point (McArdle & Nesselroade, 2014), although this is infrequent. Following common practice, they are fixed to 1 in this study (e.g., McArdle, 2009).

BLCS models are used to describe processes of exponential growth (or decay). Consider, for example, the exponential trajectories typically found in the development of intellectual abilities from childhood to early adulthood (e.g., Figure 1). Most cognitive abilities show a rapid growth during the first years of life followed by a progressive deceleration, until they reach a peak between 20 and 30 years of age—the exact age depends on the specific ability and the individual (McArdle et al., 2002). In BLCS models, this maximum level towards which the trajectory tends is modeled as an asymptote. Importantly, the means ($\mu_{x,asym}$ and $\mu_{y,asym}$) and variances ($\sigma^2_{x,asym}$ and $\sigma^2_{y,asym}$) of the trajectories in the asymptotes are not directly estimated, but can be obtained as a function of the self-feedbacks, couplings, and additive components:

$$\begin{bmatrix} \mu_{x,asym} \\ \mu_{y,asym} \end{bmatrix} = \left(\begin{bmatrix} \beta_x & \gamma_x \\ \gamma_y & \beta_y \end{bmatrix}^{-1} \right) \begin{bmatrix} \mu_{xa} \\ \mu_{ya} \end{bmatrix}$$

$$\begin{bmatrix} \sigma^2_{x,asym} & \sigma_{xy,asym} \\ \sigma_{xy,asym} & \sigma^2_{y,asym} \end{bmatrix} = \left(\begin{bmatrix} \beta_x & \gamma_x \\ \gamma_y & \beta_y \end{bmatrix}^{-1} \right) \begin{bmatrix} \sigma^2_{xa} & \sigma_{xa,ya} \\ \sigma_{xa,ya} & \sigma^2_{xa} \end{bmatrix} \left(\begin{bmatrix} \beta_x & \gamma_x \\ \gamma_y & \beta_y \end{bmatrix}^{-1} \right)^{T}$$

(5)

The mathematical relations in Equation 5 imply that the means of the additive components provide information about the position of the asymptotes with respect to the initial states. In the context of cognitive development, larger means for the additive components indicate a larger growth from one time point to the next, which results in a greater distance between the initial latent level and the asymptotes—that is, in higher values for the maximum levels towards which the mean trajectories tend. Likewise, the variances of the additive components contain information about inter-individual variability in the asymptotes. Larger variances for the additive components indicate larger inter-individual differences in the asymptotes or maximum levels of the trajectories. For a detailed account of the relations between LCS model parameters and their interpretation, see Cáncer et al. (2021).

Incorporating innovations: The stochastic BLCS model

The specification of latent changes in the previous section represents a deterministic system (Hamagami & McArdle, 2001; Ji & Chow, 2019), which is by far the most frequent specification in substantive applications of the BLCS model. It assumes that, once the initial conditions are known, individual changes can be perfectly predicted at any given point in time, and all the observed deviations from the change equations are exclusively due to measurement error. However, this is a rather unrealistic assumption in developmental research. Consider, for example, yearly measurements of cognitive abilities and academic performance from childhood to early adulthood (Peng & Kievit,

2020). Longitudinal changes in both variables could be captured by a BLCS model in which, at any given time point, the yearly changes are a function of the three elements described in equation 4. However, time-specific innovations (i.e., random shocks) may affect the system and deviate the yearly change from the deterministic trajectory, either positively or negatively. For example, a child could lose an academic year due to an illness, or a subset of the sample may have a substitute teacher affecting their academic achievement. If the researcher expects time-specific shocks due to known factors affecting all the individuals in the sample, such factors can be measured and included in the model as an exogenous time-specific variable. However, in many scenarios, there may be innovations in the system that cannot be accounted for by a specific variable.

These innovations are very different from measurement errors (see Schuurman et al., 2015). *Measurement errors* are disturbances in the observed scores caused by unobserved events or circumstances that are specific of the measurement occasion, such as a misunderstanding of the instructions, distraction, or fatigue. They are time-specific and inherent to the observed scores, and thus do not affect future states of the latent process. In contrast, *innovations* account for unobserved events with an impact on the latent process, which are carried over to later states of the system through the time-lagged parameters. They can be thought of as individual-specific "shocks" whose impact in the developing system is modeled. This carry-over effect causes past innovations to have a lingering influence on the system, potentially leading to long-run deviations from the deterministic latent trajectories. Figure 1 depicts 20 individual trajectories for two observed processes without innovations/disturbances (left panels), with measurement errors only (middle panels), and with dynamic innovations only (right panels). Note that, in Figure 1, measurement errors act as time-specific white noise, leading to up and down fluctuations around the expected latent trajectories. Innovations, however, have an

accumulative effect over time that modifies the development of the individual trajectories, affecting later time points, and eventually deviating the trajectories from their expected deterministic course.

[FIGURE 1]

Innovations are usually modeled as process random noises, and often termed *dynamic fluctuations*, *dynamic errors, impulses,* or *prediction errors* (Oud & Delsing, 2010; Schuurman et al., 2015; Voelkle et al., 2018; Voelkle & Oud, 2015; Zyphur et al., 2020). It is possible to expand equation 4 to account for them by adding a stochastic residual term to the equations for change:

$$\Delta x_{[t]} = \alpha_x \times x_a + \beta_x \times x_{[t-1]} + \gamma_y \times y_{[t-1]} + d_{x[t]}$$
$$\Delta y_{[t]} = \alpha_y \times y_a + \beta_y \times y_{[t-1]} + \gamma_x \times x_{[t-1]} + d_{y[t]}$$

(6)

Where $d_{x[t]}$ and $d_{y[t]}$ are random variables normally distributed with time-invariant mean 0, variances σ_{dx}^2 and σ_{dy}^2, and covariance $\sigma_{dx,dy}$. In the standard stochastic BLCS specification, the innovations in both latent processes can be correlated at any given time point ($\sigma_{dx,dy}$), and no further covariances are allowed between the errors and other elements in the system. In a bivariate system, these covariances are interpreted as the linear relation between the "shocks" affecting each of the latent processes over time. Typically, the parameters regarding dynamic errors are assumed to be equal across occasion (i.e., time invariant). Figure 2 depicts the full stochastic BLCS model, which is specified by 24 parameters: 4 means, 8 variances, 8 covariances, and 4 time-lagged effects.

Importantly, innovations are dynamic in nature because a) they capture external shocks that are unaccounted for by previous states and b) their effects on the system are

carried over to future states. Note that, in Figure 2, the innovation variance introduced in $\Delta x_{[t-1]}$ goes directly into the latent level $x_{[t-1]}$ through a regression weight of 1. Then, the variance of $x_{[t-1]}$ (which includes all the innovation variance entered in the system up to that moment) is propagated through three regression paths. The first path goes into the latent level $x_{[t]}$ through a regression weight of 1, and the other two paths predict the latent changes in $\Delta x_{[t]}$ and $\Delta y_{[t]}$ (through β_x and γ_x, respectively). Therefore, the impact of the external shocks on each occasion is "passed along" to later states of the system, leading to the lingering deviations depicted in Figure 1 (right-side column).

[FIGURE 2]

Stochastic BLCS models are very rare in the developmental literature. Some exceptions include the study of the interrelations between cognition and reading in dyslexic readers (Ferrer et al., 2010), and the developmental dynamics between reading and cognition during childhood (Ferrer et al., 2007). One possible reason for this scarcity may be that they include a large number of parameters, which may lead to estimation problems. In this regard, Usami et al., (2019) pointed out that the inclusion of both measurement errors and innovations may lead to convergence errors and improper solutions (i.e., solutions with non-available or unreasonable large parameters or standard errors). In a simulation study, Ji & Chow, (2019) observed that measurement and dynamic errors could not be reliably distinguished under certain conditions. However, these authors examined only a restricted set of scenarios, and their focus was on the effect of misspecification of the initial conditions. Moreover, and to the best of our knowledge, no previous studies have systematically examined the consequences of estimating a deterministic BLCS model when the latent trajectories are affected by random shocks or innovations.

The present study

In sum, developmental systems are often impacted by random shocks that affect individuals differently, shaping their trajectories and altering their patterns of change. Deterministic BLCS models, in which the changes are assumed to be fully determined by time-lagged effects and initial conditions, are not able to capture these influences. In contrast, stochastic BLCS models include innovation parameters that account for these influences and model their effect on the system.

In the present work, we conducted an extensive Monte Carlo simulation to evaluate the recovery of population parameters in a bivariate system under a broad set of empirically relevant conditions. As described previously, most BLCS specifications in the empirical literature do not include stochasticity in the latent changes. Therefore, the goals of this manuscript are: a) to examine the consequences of not accounting for stochastic innovations when they are affecting the latent trajectories (i.e., how robust the deterministic model is when fitted to data with varying levels of stochastic innovations), and b) to evaluate the effectiveness of the deterministic and stochastic BLCS models to recover dynamic features of two developmental processes under different conditions affecting the sampling and the populational trajectories.

Method

We generated repeated measures for two processes x and y that unfold over time. The processes were generated according to the stochastic BLCS model described in Equation 6. The generating parameters were chosen to represent trajectories that are typical of the development of cognitive abilities from childhood to early adulthood (e.g., Kail & Ferrer, 2007; Schmitt et al., 2017; van der Maas et al., 2006), and were based on previous

empirical studies (Ferrer et al., 2007, 2010; Shaywitz et al., 1990). They are reported in table 1.

[TABLE 1]

Based on the parameters in Table 1, we simulated three hundred data sets for each of the 80 conditions created by the combination of the following four factors:

- Proportion of variance due to stochasticity in the innovations (*Ps*): four conditions (*Ps* = {.0, .05, .15, .25})

- Sample size: four conditions (*N* = {50, 100, 200, 500})

- Number of repeated measures: five conditions (*T* = {3, 4, 5, 7, 10})

Proportion of variance in the process due to stochasticity in the innovations (Ps)

The specification of the stochastic BLCS model (i.e., with 24 parameters) has been rarely applied in substantive research. Therefore, there is little evidence on the extent to which the evolution of developmental variables is affected by random innovations that accumulate over time. Ferrer and colleagues studied the development of reading abilities during childhood and obtained variances for the innovations that represented around 5 to 15% of the variance of the latent initial scores (Ferrer et al., 2007, 2010).

In order to cover a broad range of scenarios, we used four values for the *Proportion of latent variance due to the stochasticity of the innovations*, *Ps* = (.0, .05, .15, .25). We computed it based on the total amount of latent variance in $t=2$. *Ps* quantifies the proportion of such variance that is due to the innovations' variance. For example, for the latent process *x*,

$$Ps_x = \frac{\sigma^2_{dx}}{\sigma^2_{x[t=2]} + \sigma^2_{dx}} \qquad (7)$$

A value of $Ps = .15$ implies that, from the total latent variance in the process x at $t=2$ (excluding the measurement error variance), 15% is introduced by the variance of the innovations at $t=2$, while the remaining 85% is due to the latent variances of x and y at $t=1$, plus influence of the corresponding latent additive component. In other words, if the variance of the latent process x in $t=2$ took a value of 10, proportions of stochasticity of .0, .05, .15, and .25 would lead to variance due to innovation variances of 0, .5, 1.5, and 2.5, respectively. The covariance between innovations was set according to a correlation of 0.2 across all conditions. Note that a high correlation between "shocks" would imply that the stochastic deviations have a common cause across time points and individuals. This is very unlikely, as these "shocks" represent the impact of multiple individual-specific events over time. Since this correlation is likely to be low and to not vary widely across applications, we do not expect it to have a meaningful impact in the performance of the BLCS model.

Figure 3 depicts a set of trajectories generated with the parameters from Table 1 using different Ps values. The proportion of variance due to stochastic innovations can be thought of as the extent to which the states of the process are influenced by external shocks or disturbances. In other words, it represents the amount of "randomness" or unpredictability in a system. If the proportion of variance due to innovations is high, the values of the process will be more influenced by external shocks and will be more unpredictable. This would result in a lower degree of autocorrelation, as the current state of the process would be less related to its past states. As illustrated in Figure 3, conditions with more stochasticity due to innovations simulate shocks with a greater impact on the system. These shocks are then propagated through the self-feedbacks and couplings to

subsequent states, ultimately resulting in larger and more persistent deviations from the deterministic trajectories. Note that the deterministic BLCS model is a correct specification when $Ps = 0$, whereas the stochastic specification is needed to adequately account for the innovation variance in the rest of Ps conditions. The condition of $Ps = 0$ was included to evaluate: a) the performance of the deterministic model when it is correctly specified, and b) whether the stochastic model captures innovations as null when they are zero in the population.

[FIGURE 3]

Number of measurement occasions

One of our goals was to explore how many measurement occasions are needed to adequately capture the features of longitudinal bivariate processes. To this end, the number of time points chosen were 3, 4, 5, 7, 10. We focused on time points under 10 because they are more frequent in developmental studies (e.g., Ahmed et al., 2014; Estrada et al., 2019; Finkel et al., 2013, 2016; Gerstorf et al., 2007; Ghisletta & Lindenberger, 2003; Liao et al., 2018; Small et al., 2012), and because longer studies are often not feasible in developmental research.

Sample size

BLCS models have been applied to data sets containing hundreds of participants (e.g., Estrada et al., 2019; Ferrer et al., 2007; Finkel et al., 2016; Ghisletta et al., 2006; Lövdén et al., 2005, 2007; Malone et al., 2004; Quinn et al., 2015; Small et al., 2012; Ziegler et al., 2015), or even thousands (e.g., Bielak et al., 2011; Grimm, 2007; Infurna & Gerstorf, 2013; Liao et al., 2018; McArdle & Prindle, 2008; Sargent-Cox et al., 2012; Snitz et al., 2015). Due to the complexity and number of parameters of the BLCS

model, such large samples are likely to be necessary. However, most developmental researchers face economic and other constraints that usually make it difficult to gather very large samples, particularly when repeated measures are required. In order to explore the minimal requirements to adequately estimate the models, we included the following four sample sizes: 50, 100, 200, and 500 individuals per sample.

Estimation and data analysis

Our main goal was to examine the ability of the deterministic and stochastic specifications of the BLCS model to recover developmental trajectories, especially when they are affected by stochastic innovations at the latent level. For each sample in each condition, we estimated: a) one deterministic BLCS model, which does not account for potential innovations in the trajectories and has 21 free parameters (Equation 4), and b) one stochastic BLCS model, which includes two dynamic error variances and one covariance to capture the effects of stochastic innovations in the system, resulting in a total of 24 free parameters (Equation 6). For the specification and estimation of the models, we used OpenMx in R (RAM parameterization estimated with maximum likelihood; cf., Ghisletta & McArdle, 2012; Neale et al., 2016). The estimation was performed through the *OpenMx* functions *mxModel* and *mxRefModels* (Boker et al., 2018). The R code for generating the data sets and estimating the BLCS models is available at: https://github.com/PFernandez-Cancer/stochasticBLCS.

Results

In this section we evaluate, across all simulation conditions: 1) the rates of improper solutions, 2) the bias of the parameter estimates, 3) the variability of the parameter estimates, and 4) the bias of the standard errors. In BLCS applications, the focus is usually on the time-lagged dynamics (i.e., self-feedbacks and cross-lagged

effects). Because of this, and given that innovations' variances have been mostly

unexplored before, here we focus on how self-feedbacks and couplings are recovered

when the trajectories are affected by these innovations. Results for the means and

variances are also described, but due to space constraints, only the self-feedbacks and

couplings are depicted in Figures 5-7. Results regarding the covariances, additional

figures, and extended numerical results are available in the supplemental materials.

Improper solutions

First, we examined the solutions that included invalid parameter estimates, leading to

uninterpretable results. Solutions were considered improper, and removed from

subsequent analysis, when they contained at least one of the following: (a) not available

(NA) parameter or standard error estimates; and (b) unreasonably large parameter or

standard error estimates with absolute value above |10|. Figure 4 shows the percentage

of improper solutions for the stochastic BLCS model.

[FIGURE 4]

The incidence of improper solutions was very low for both specifications of the

BLCS model. For the deterministic model, improper solutions ranged from 0 to 10% in

conditions with three repeated measures, and from 0 to 1.7% in the remaining

conditions. Improper solutions for the deterministic model were not included in Figure 4

due to this low incidence. For the stochastic model, improper solutions ranged from 0 to

28%, and were also more frequent in conditions with three repeated measures. As

expected, increasing the number of repeated measures and the sample size reduced the

rates of improper solutions. In contrast, the proportion of stochasticity in the

innovations had a much smaller impact on the performance of the stochastic model,

with increasing proportions of stochasticity generally resulting in slightly lower

percentages of improper solutions. We discuss strategies for dealing with improper

solutions in the Discussion section.

Bias of the estimates

We examined the accuracy of the estimates as the bias of each parameter in each

condition: $bias = \overline{\theta}_{est} - \theta$, where θ is the true value and $\overline{\theta}_{est}$ is the average estimate

across all replications in a given condition[1].

Self-feedbacks and couplings

Figure 5 depicts the bias for self-feedback and coupling parameters across all

conditions. When the trajectories were not affected by stochastic innovations ($Ps = 0$),

both the deterministic and the stochastic models required a minimum of four repeated

measures and 100 individuals, or five repeated measures and 50 individuals, to produce

unbiased estimates (bias range from −.04 to .04). As expected, increasing the proportion

of stochasticity led to more biased estimates for the deterministic model, probably

because it also increased the degree of misspecification. Interestingly, this also occurred

with the stochastic model, although to a lesser extent. That is, although the stochastic

model included parameters to capture innovations, the time-lagged dynamics were

slightly more biased when such innovations were large. When the number of repeated

measures was increased to seven, this effect was barely noticeable in the stochastic

---

[1] We did not use the relative bias because, when the trajectories are not affected by innovations,

the dynamic error variances and covariances equal zero. Thus, computing the relative bias in

these conditions would imply dividing by zero.

model (range −.04 to .05), although it persisted in the deterministic model (range to

−.10 to .08).

[FIGURE 5]

Latent means and variances

Due to space constraints, the complete numerical results and corresponding

figures are included in the supplemental materials. Here we discuss the most relevant

findings. When the number of repeated measures was three or four, the means and

variances of the additive components were substantially biased, regardless of the model

specification and the proportion of variance due to innovations (ranges −.48 to 0.18 for

the means, and −.01 to 1.03 for variances). Note that the overestimation of the additive

component variances implies an overestimation of the inter-individual differences in the

asymptotes of the trajectories (i.e., the maximum level towards which each individual

tends as time increases). In the stochastic model, all the latent means and variances of

the model were accurately recovered with five or more repeated measures (range −.06

to .07), regardless of the amount of stochasticity due to innovations. In the deterministic

model, however, medium and large proportions of stochasticity (*Ps* = .15 and *Ps* = .25)

led to overestimations of the measurement error variances.

Variability of the estimates

We evaluated the variability (i.e., precision) of the parameter estimates by computing

their empirical standard deviation as $SD = \sqrt{\sum_{k=1}^{K} (\theta_k - \overline{\theta}_{est})^2 / K}$, where θ_k is the

estimated value in a replication *k*, $\overline{\theta}_{est}$ is the average estimate across *K* replications, and

K is the number of replications in a given condition.

Self-feedbacks and couplings

Figure 6 depicts the variability for the self-feedback and coupling estimates across all conditions. Overall, conditions with three repeated measures led to highly varying estimates for the time-lagged parameters (with standard deviations ranging from .18 to 1.49). In the remaining conditions, both model specifications displayed a similar performance. As expected, the most relevant factor for the variability of the parameter estimations was the number of repeated measures, followed by the sample size. Increasing proportions of stochasticity due to innovations led to slightly more variability in the estimates of the stochastic model. However, this effect was small and mostly limited to conditions with three and four repeated measures.

[FIGURE 6]

Latent means and variances

Due to space constraints, the complete numerical results and corresponding figures are included in the supplemental materials. Here we discuss the most relevant findings regarding estimate variability. The latent means and variances displayed a pattern of variability similar to the time-lagged parameters: the variability was lower with increasing number of repeated measures and sample size, but mostly unaffected by the amount of stochasticity due to innovations. In conditions with three and four repeated measures, the additive component means and variances showed high variability (range .05 to .72 for the means and .05 to 1.48 for the variances), especially with low sample sizes. Overall, increasing the proportion of stochasticity due to innovations did not lead to a higher variability of the measurement error variance estimates.

Bias of the standard errors

We examined the accuracy of the standard errors for each parameter in each condition as: $bias = SE_{est} - SD$, where SE_{est} is the estimated standard error and SD is the standard deviation across all replications in a given condition.

Self-feedbacks and couplings

Figure 7 depicts the bias of the standard errors for the self-feedback and coupling parameters across all conditions. In conditions with four or more repeated measures, the proportion of stochasticity due to innovations had little impact on the recovery of the standard errors. Only in the deterministic model, increasing proportions of stochasticity led to more biased standard errors, although this bias became negligible with increasing sample size. In the stochastic model, standard errors were accurately recovered with five repeated measures and 200 individuals, or seven repeated measures and 100 individuals (range −.04 to −.001). In the deterministic model, similar sampling conditions were required to produce standard errors, except when the impact of innovations was large (Ps = .25). In such conditions, the deterministic model required larger samples than its stochastic counterpart to achieve similar levels of accuracy.

[FIGURE 7]

Importantly, when the standard errors were not accurately recovered, they were always underestimated. This means that, in such conditions, the confidence intervals built with the standard errors were narrower than they should be, thus incorrectly inflating the degree of precision around the point estimate.

Latent means and variances

Due to space constraints, the complete numerical results and corresponding figures are included in the supplemental materials. Here we discuss the most relevant findings. Both the deterministic and the stochastic models led to similar results regarding the standard errors of the latent means and variances. Overall, the means and variances of the additive components had negatively biased standard errors, especially with three and four repeated measures. In conditions with three repeated measures, this bias did not decrease with increasing sample size. This result, together with the low accuracy and large variability of the parameter estimates, suggests that the deterministic and stochastic BLCS models may not be tenable with only three repeated measures and the sample sizes examined in the present study.

Discussion

Summary of findings

In this study, we evaluated the ability of the deterministic and stochastic specifications of the Bivariate Latent Change Score model (BLCS) to recover the characteristics of two processes that unfold over time. We examined the performance of the models when the latent trajectories were impacted by varying degrees of stochasticity in the innovations. We also studied this performance across various combinations of sample size and number of repeated measures.

Regarding sampling conditions, the number of repeated measures was the most relevant factor for adequate parameter recovery. In general, including additional measurement occasions had a much larger effect on the quality of the parameter estimates than increasing the sample size. One of the most relevant findings is that both the deterministic and stochastic BLCS models were not tenable with three repeated measures. In this condition, most estimates were substantially biased. Even in conditions with 500

individuals, the parameter estimates were very sensitive to sampling fluctuations—that is, they varied widely across replications. This indicates that samples of such size may not be suitable for the BLCS models presented in this manuscript if only three repeated measures are used.

Another relevant finding is that, in general, the deterministic BLCS model did not always provide adequate estimates when the latent trajectories were affected by stochastic innovations similar to those found in the empirical literature. As expected, some of the deviations in the trajectories produced by innovations were captured as measurement errors, inflating their variance. However, they also led to substantial bias in the self-feedbacks, couplings, and additive component means and variances of the deterministic model. Only when the impact of innovations was null or very small, the deterministic model provided adequate estimates, but it required at least five or more repeated measures and 200 participants.

In contrast, the stochastic BLCS model was capable of: 1) recovering innovation variances as null when they were zero in the population, and 2) capturing innovation variances of different size from small to large. When the proportion of variance in the latent processes due to innovations was null or very small, the sampling requirements for adequate parameter recovery were similar to those of the deterministic model (i.e., 200 participants and at least five repeated measures). However, when this source of variance was larger ($Ps = .15$ and $Ps = .25$), the time-lagged dynamics and the additive component means and variances were slightly more biased. In such conditions, the stochastic model required at least seven repeated measures to achieve an accurate and reliable recovery of the parameters.

Finally, both the deterministic and stochastic BLCS model tended to underestimate standard errors. This could be problematic because: 1) it falsely inflates the degree of precision around the point estimate and 2) it may lead to confidence intervals that are too narrow to cover the true parameter value. Fortunately, this bias was mostly limited to scenarios with few repeated measures and small sample sizes. In conditions with five repeated measures and 200 individuals, or seven repeated measures and 50 individuals, the standard errors were acceptably accurate.

Theoretical and methodological considerations

Several authors have pointed out that LCS models are susceptible to improper solutions and convergence errors, especially when measurement and dynamic errors are simultaneously estimated (e.g., McArdle et al., 2004; Usami et al., 2015; Usami et al., 2019). In this study, we did not find substantial amounts of improper solutions, although they were slightly more frequent in the stochastic model. Improper solutions may appear simply due to sampling fluctuations, and certain combinations of generating parameters may be more prone to improper solutions than others. There are several strategies for dealing with this problem in BLCS models. First, it is usually advisable to standardize the variables of interest with respect to the first measurement occasion (i.e., for all repeated measures, subtract the mean of the first occasion and divide by the standard deviation of the first occasion), because the maximum likelihood algorithm estimates the parameters more easily when they are in similar scales. A second (and compatible) option is using plausible sets of starting values for the parameters. For example, it may be reasonable to fit two univariate LCS models first, and use the resulting parameter estimates as starting values for the BLCS model. Also, researchers can aid the estimation by setting boundaries to restrict the range of possible values that a parameter estimate can take (e.g., setting the lower bound of a variance to 0 to prevent negative estimates).

Several recent studies have expressed the need for comprehensive simulation studies to evaluate the quality of the estimates under different combinations of sample size and number of repeated measures (e.g., Ji & Chow, 2019; Kievit et al., 2018). Our work provides guidelines on the appropriateness of the deterministic and stochastic BLCS model under a broad range of common sampling conditions. From the standpoint of model identifiability, two variables measured three times provide 27 degrees of freedom (15 covariances, 6 variances, and 6 means). The deterministic and stochastic BLCS models include 21 and 24 parameters respectively, therefore both specifications are identified with three repeated measures. However, our findings suggest that the estimates provided by the BLCS model may not be an accurate representation of the underlying process in such conditions, and therefore they should be interpreted with caution.

Importantly, we found that the sampling requirements of the BLCS model differed depending on the amount of stochasticity in the latent processes due to innovations. Based on such findings, we provide the following indications. If the researcher does not expect the processes under study to be affected by innovations, or their expected impact is very small, they can safely use the deterministic model with four repeated measures and 500 individuals, or five repeated measures and 200 individuals. However, it may be more cost efficient to increase the number of repeated measures to seven and recruit only 100 individuals. If measuring ten times is possible, then a sample size of 50 individuals would yield similar results to the previously mentioned conditions. On the other hand, if 15% or more of the individual differences in the latent trajectories are expected to be due to the impact of innovations, the deterministic model will lead to biased estimates. In such scenarios, the stochastic model is more adequate, but it requires larger samples. In order to achieve highly reliable and accurate estimates, the stochastic model may require 500 individuals and five repeated measures, 200 individuals and seven repeated measures, or

50 individuals and ten repeated measures. However, to the best of our knowledge, very few empirical studies have used stochastic models to examine developmental constructs. Moreover, it is very unusual to have a prior expectation of the extent to which any latent construct might be affected by stochastic innovations. In such situations of uncertainty, using the stochastic model with the sampling conditions described above is the safest choice.

It is important to note that the minimum sampling requirements proposed in this manuscript were based on complete versions of the BLCS model including 21 and 24 parameters (for the deterministic and stochastic specifications, respectively). Nevertheless, other specifications are possible depending on the hypothesis of change. For example, in certain empirical scenarios, it may be theoretically plausible to fix some parameters to zero, such as the couplings or the additive component variances. Specifications with fewer parameters may require less demanding sampling conditions to provide adequate estimates, and they may be easier to estimate with three repeated measures. A strategy to choose between different specifications is to sequentially fit nested versions of the BLCS model, starting from a basic restricted version (for example, fixed effects in the additive components), and progressing towards more complex and unconstrained versions. Through sequential comparison of nested models by means of likelihood ratio tests (or other alternative strategies, see Usami et al., 2016), it is possible to check hypothesis about the presence of specific effects and sources of variance. This procedure has been used with deterministic BLCS models applied to a single data set (e.g., Estrada et al., 2019).

Limitations and future directions

In this study, we estimated the stochastic BLCS model as a structural equation model (SEM) with Maximum Likelihood estimation because it is a familiar and accessible approach for empirical researchers. However, BLCS models can be estimated in other frameworks such as state-space models (SSM; Hunter, 2018; Ji & Chow, 2019; Oud & Jansen, 2000), continuous-time modelling (see Voelkle et al., 2012), or using other estimation methods such as Bayesian estimation. SSMs provide an efficient approach to the analysis of intensive longitudinal data, where individuals are measured many times and the computational burden becomes costly for a SEM specification. In developmental settings, however, where individuals are typically measured few times, SSMs should present a performance similar to SEMs. On the other hand, treating time as a continuous variable may provide more accurate estimates if the intervals between occasions are unevenly spaced, which is frequent in longitudinal studies. Finally, Bayesian estimation has been shown to reduce improper solutions and improve the estimation in other dynamic models, such as the STARTS model (Lüdtke et al., 2018), n=1 autoregressive models (Schuurman et al., 2015), and ARMA models (Asparouhov & Muthén, 2020). Future research could extend this work by examining the performance of the BLCS model from the SSM, continuous-time, or Bayesian frameworks.

As described previously, the deterministic and stochastic BLCS models were untenable under conditions with three repeated measures. In our simulation design, conditions with more repeated measures covered a larger portion of the developmental trajectory. Previous research has indicated that the developmental trajectories of certain cognitive abilities may be better estimated when the information provided by the sample is as complete as possible at the region with greatest curvature (Mistler & Enders, 2012; Rhemtulla & Hancock, 2016). One of the reasons of the poor performance of the BLCS

model may be that the curvature of the trajectories is not sufficiently well captured with three repeated measures. An alternative would be to divide the complete time range of the study into three occasions. For example, if we are studying two processes from the ages 11 to 20, measurements could be taken at ages 11, 15.5 and 20, instead of 11, 12, and 13. Future research should examine the performance of the BLCS model under such sampling designs, and investigate whether using three repeated measures is feasible in such scenarios.

Another aspect that requires further research is the inclusion of additional predictors or covariates in the stochastic BLCS. For example, it is possible to include time-invariant predictors affecting the initial conditions (e.g., sex, education, or socio–economic status). Likewise, time-specific influences can be included to (at least partially) account for innovations that we have modelled here as innovation variances (e.g., increases or decreases in anxiety, or physical activity). There are several examples of the use of time-invariant and time-varying predictors in the LCS literature (e.g. Bodenmann et al., 2014; Ghisletta & Lindenberger, 2003; Hertzog et al., 2003; Snitz et al., 2015). In the presence of this type of predictors, the sampling conditions may become different from the ones reported in this study.

As a final note, we evaluated the performance of the stochastic BLCS model with evenly spaced time intervals and without missing data. In this regard, we worked with optimal conditions that are not always found in applied research. For these reasons, we recommend a conservative interpretation of our findings and encourage researchers to use sampling conditions above the minimum requirements recommended in the present work.

Conclusion

Individuals are often exposed to random events (i.e., random shocks) that influence their development over time, resulting in deviations from their expected trajectories. The deterministic BLCS model, which is widely used in developmental research, is not always capable of capturing the dynamics of the latent processes when the trajectories are affected by such random shocks. In contrast, the stochastic Bivariate Latent Change Score model includes innovations at the latent level to account for the impact of these shocks on the system, providing a more accurate representation of developmental processes. Our results suggest that, under the right sampling conditions, the stochastic BLCS model is able to detect latent innovations, distinguish them from other sources of variance, and provide reliable information about the latent dynamics. Based on our findings, researchers should note that:

(1) The complete versions of deterministic and the stochastic specifications of the BLCS model are not tenable with 3 repeated measures if 500 participants or less are recruited.

(2) When the latent trajectories are not affected by random shocks or innovations (or their impact is very small), both the deterministic and stochastic BLCS model require similar sampling conditions to recover the latent dynamics. When the impact of random shocks in the latent trajectories is medium or large, the stochastic model requires larger sample sizes.

(3) We recommend using the deterministic BLCS model only when the researcher expects no (or little) impact of innovations in the latent processes, and the following minimum sampling requirements are met: 4 repeated measures and 500 participants, 5 repeated measures and 200 participants, 7 repeated measures and 100 participants, or 10 repeated measures and 50 participants.

(4) If the researcher suspects that the trajectories are substantially affected by innovations, we recommend estimating the stochastic BLCS model with the following minimum sampling conditions: 5 repeated measures and 500 participants, 7 repeated measures and 200 participants, or 10 repeated measures and 50 participants.

We have shown that, under the right sampling conditions, both the deterministic and stochastic BLCS model are accurate and reliable tools for the recovery of dynamics in developmental data. Based on these findings, we encourage researchers to include stochastic innovations in their BLCS specification if they expect innovations to have an impact in the latent construct under study. We hope the findings in this study will help researchers on the design of future longitudinal studies, and will provide insight about the risks of using BLCS models under adverse sampling conditions.

References

Ahmed, Y., Wagner, R. K., & Lopez, D. (2014). Developmental relations between reading and writing at the word, sentence, and text levels: A latent change score analysis. *Journal of Educational Psychology*, *106*(2), 419–434. https://doi.org/10.1037/a0035692

Asparouhov, T., & Muthén, B. (2020). Comparison of Models for the Analysis of Intensive Longitudinal Data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 275–297. https://doi.org/10.1080/10705511.2019.1626733

Bielak, A. A. M., Gerstorf, D., Kiely, K. M., Anstey, K. J., & Luszcz, M. (2011). Depressive Symptoms Predict Decline in Perceptual Speed in Older Adulthood. *Psychology and Aging*, *26*(3), 576–583. https://doi.org/10.1037/a0023313

Bodenmann, G., Hilpert, P., Nussbeck, F. W., & Bradbury, T. N. (2014). Enhancement of couples' communication and dyadic coping by a self-directed approach: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *82*(4), 580–591. https://doi.org/10.1037/a0036356

Boker, S. M., Neale, M. C., Maes, H. H., Wilde, M. J., Spiegel, M., Brick, T. R., Estabrook, R., Bates, T. C., & Mehta, P. (2018). OpenMx user guide. Retrieved from: https://openmx.ssri.psu.edu/documentation

Cáncer, P. F., Estrada, E., Ollero, M. J., & Ferrer, E. (2021). Dynamical Properties and Conceptual Interpretation of Latent Change Score Models. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.696419

Estrada, E., Ferrer, E., Karama, S., Román, F. J., & Colom, R. (2019). *Time-Lagged Associations Between Cognitive and Cortical Development From Childhood to Early Adulthood*. *55*(6), 1338–1352. http://dx.doi.org/10.1037/dev0000716

Ferrer, E., McArdle, J. J., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2007). Longitudinal models of developmental dynamics between reading and cognition from childhood to adolescence. *Developmental Psychology*, *43*(6), 1460–1473. https://doi.org/10.1037/0012-1649.43.6.1460

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of Reading and IQ Over Time: Empirical Evidence for a Definition of Dyslexia. *Psychological Science*, *21*(1), 93–101. https://doi.org/10.1177/0956797609354084

Finkel, D., Ernsth-Bravell, M., & Pedersen, N. L. (2016). Temporal Dynamics of Motor Functioning and Cognitive Aging. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *71*(1), 109–116. https://doi.org/10.1093/gerona/glv110

Finkel, D., Reynolds, C. A., Emery, C. F., & Pedersen, N. L. (2013). Genetic and environmental variation in lung function drives subsequent variation in aging of fluid intelligence. Behavior genetics, 43(4), 274-285. https://doi.org/10.1007/s10519-013-9600-3

Gerstorf, D., Lövdén, M., Röcke, C., Smith, J., & Lindenberger, U. (2007). Well-being affects changes in perceptual speed in advanced old age: Longitudinal evidence for a dynamic link. *Developmental Psychology*, *43*(3), 705–718. https://doi.org/10.1037/0012-1649.43.3.705

Ghisletta, P., Bickel, J.-F., & Lovden, M. (2006). Does Activity Engagement Protect Against Cognitive Decline in Old Age? Methodological and Analytical Considerations. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *61*(5), P253–P261. https://doi.org/10.1093/geronb/61.5.P253

Ghisletta, P., & Lindenberger, U. (2003). Age-Based Structural Dynamics Between Perceptual Speed and Knowledge in the Berlin Aging Study: Direct Evidence for Ability Dedifferentiation in Old Age. *Psychology and Aging*, *18*(4), 696–713. https://doi.org/10.1037/0882-7974.18.4.696

Ghisletta, P., & McArdle, J. J. (2012). Latent Curve Models and Latent Change Score Models Estimated in R. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(4), 651–682. https://doi.org/10.1080/10705511.2012.713275

Grimm, K. J. (2007). Multivariate longitudinal methods for studying developmental relationships between depression and academic achievement. *International Journal of Behavioral Development*, *31*(4), 328–339. https://doi.org/10.1177/0165025407077754

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, *New Developments and Techniques in Structural Equation Modeling* (pp. 203–246). Erlbaum Publishers.

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent Change Models of Adult Cognition: Are Changes in Processing Speed and Working Memory Associated With Changes in Episodic Memory? *Psychology and Aging*, *18*(4), 755–769. https://doi.org/10.1037/0882-7974.18.4.755

Hunter, M. D. (2018). State Space Modeling in an Open Source, Modular, Structural Equation Modeling Environment. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(2), 307–324. https://doi.org/10.1080/10705511.2017.1369354

Infurna, F. J., & Gerstorf, D. (2013). Linking perceived control, physical activity, and biological health to memory change. *Psychology and Aging*, *28*(4), 1147–1163. https://doi.org/10.1037/a0033327

Ji, L., & Chow, S.-M. (2019). Methodological Issues and Extensions to the Latent Difference Score Framework 1. In E. Ferrer, S. M. Boker, & K. J. Grimm (Eds.), *Longitudinal Multivariate Psychology* (pp. 9–37). Routledge. https://doi.org/10.4324/9781315160542-2

Kail, R. V., & Ferrer, E. (2007). Processing Speed in Childhood and Adolescence: Longitudinal Models for Examining Developmental Change. *Child Development*, *78*(6), 1760–1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

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 (pp. 243–263). American Psychological Association. https://doi.org/10.1037/10409-008

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. Developmental Cognitive Neuroscience, 33, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

King, L. A., King, D. W., McArdle, J. J., Saxe, G. N., Doron-LaMarca, S., & Orazem, R. J. (2006). Latent difference score approach to longitudinal trauma research. *Journal of Traumatic Stress*, *19*(6), 771–785. https://doi.org/10.1002/jts.20188

Liao, J., Muniz-Terrera, G., Head, J., & Brunner, E. J. (2018). Dynamic Longitudinal Associations Between Social Support and Cognitive Function: A Prospective Investigation of the Directionality of Associations. *The Journals of*

Gerontology: Series B, *73*(7), 1233–1243.

https://doi.org/10.1093/geronb/gbw135

Lövdén, M., Ghisletta, P., & Lindenberger, U. (2005). Social Participation Attenuates

Decline in Perceptual Speed in Old and Very Old Age. *Psychology and Aging*,

20(3), 423–434. https://doi.org/10.1037/0882-7974.20.3.423

Lövdén, M., Li, S.-C., Shing, Y. L., & Lindenberger, U. (2007). Within-person trial-to-

trial variability precedes and predicts cognitive decline in old and very old age:

Longitudinal data from the Berlin Aging Study. *Neuropsychologia*, *45*(12),

2827–2838. https://doi.org/10.1016/j.neuropsychologia.2007.05.005

Lüdtke, O., Robitzsch, A., & Wagner, J. (2018). More stable estimation of the STARTS

model: A Bayesian approach using Markov chain Monte Carlo techniques.

Psychological Methods, *23*(3), 570–593. https://doi.org/10.1037/met0000155

Malone, P. S., Lansford, J. E., Castellino, D. R., Berlin, L. J., Dodge, K. A., Bates, J. E.,

& Pettit, G. S. (2004). Divorce and Child Behavior Problems: Applying Latent

Change Score Models to Life Event Data. *Structural Equation Modeling: A*

Multidisciplinary Journal, *11*(3), 401–423.

https://doi.org/10.1207/s15328007sem1103_6

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic

structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom, *Structural equation*

modeling, present and future: A Festschrift in honor of Karl Jöreskog (pp. 7–

46). Scientific Software International.

McArdle, J. J. (2009). Latent Variable Modeling of Differences and Changes with

Longitudinal Data. *Annual Review of Psychology*, *60*(1), 577–605.

https://doi.org/10.1146/annurev.psych.60.110707.163612

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., & Woodcock, R. W. (2002).

Comparative longitudinal structural analyses of the growth and decline of

multiple intellectual abilities over the life span. *Developmental Psychology*,

38(1), 115–142. https://doi.org/10.1037/0012-1649.38.1.115

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for

linear dynamic analyses with incomplete longitudinal data. In L. M. Collins &

A. G. Sayer (Eds.), *New methods for the analysis of change.* (pp. 139–175).

American Psychological Association. https://doi.org/10.1037/10409-005

McArdle, J. J., & Hamagami, F. (2004). Methods for dynamic change hypotheses. In K.

van Montfort, J. Oud, & A. Satorra (Eds.), *Recent Developments on Structural

Equation Models* (Vol. 19, pp. 295–335). Springer Netherlands.

https://doi.org/10.1007/978-1-4020-1958-6

McArdle, J. J., Hamagami, F., Jones, K., Jolesz, F., Kikinis, R., Spiro, A., & Albert, M.

S. (2004). Structural Modeling of Dynamic Changes in Memory and Brain

Structure Using Longitudinal Data From the Normative Aging Study. *The

Journals of Gerontology Series B: Psychological Sciences and Social Sciences*,

59(6), 294–304. https://doi.org/10.1093/geronb/59.6.P294

McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural

equation models*. American Psychological Association.

https://doi.org/10.1037/14440-000

McArdle, J. J., & Prindle, J. J. (2008). A latent change score analysis of a randomized

clinical trial in reasoning training. *Psychology and Aging, 23*(4), 702–719.

https://doi.org/10.1037/a0014349

Mistler, S. A., & Enders, C. K. (2012). Planned missing data designs for developmental research. In B. Laursen, T. D. Little, & N. A. Card (Eds.), Handbook of developmental research methods (pp. 742–754). The Guilford Press.

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10.1007/s11336-014-9435-8

Oud, J. H. L., & Delsing, M. J. M. H. (2010). Continuous Time Modeling of Panel Data by means of SEM. In K. van Montfort, J. H. L. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 201–244). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-11760-2_7

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of sem. *Psychometrika*, *65*(2), 199–215. https://doi.org/10.1007/BF02294374

Peng, P., & Kievit, R. A. (2020). The Development of Academic Achievement and Cognitive Abilities: A Bidirectional Perspective. *Child Development Perspectives*, *14*(1), 15–20. https://doi.org/10.1111/cdep.12352

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental Relations Between Vocabulary Knowledge and Reading Comprehension: A Latent Change Score Modeling Study. *Child Development*, *86*(1), 159–175. https://doi.org/10.1111/cdev.12292

Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. Educational Psychologist, 51, 305–316. http://dx.doi.org/10.1080/00461520.2016.1208094

Sargent-Cox, K. A., Anstey, K. J., & Luszcz, M. A. (2012). The relationship between change in self-perceptions of aging and physical functioning in older adults. *Psychology and Aging*, *27*(3). https://doi.org/10.1037/a0027578

Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., & McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology*, *109*(8), 1120–1140. https://doi.org/10.1037/edu0000193

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n = 1 psychological autoregressive modeling. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01038

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of Reading Disability in Boys and Girls: Results of the Connecticut Longitudinal Study. *JAMA*, *264*(8), 998–1002. https://doi.org/10.1001/jama.1990.03450080084036

Small, B. J., Dixon, R. A., McArdle, J. J., & Grimm, K. J. (2012). Do changes in lifestyle engagement moderate cognitive decline in normal aging? Evidence from the Victoria Longitudinal Study. *Neuropsychology*, *26*(2), 144–155. https://doi.org/10.1037/a0026579

Snitz, B. E., Small, B. J., Wang, T., Chang, C.-C. H., Hughes, T. F., & Ganguli, M. (2015). Do Subjective Memory Complaints Lead or Follow Objective Cognitive Change? A Five-Year Population Study of Temporal Influence. *Journal of the International Neuropsychological Society*, *21*(9), 732–742. https://doi.org/10.1017/S1355617715000922

Usami, S., Hayes, T., & McArdle, J. J. (2015). On the Mathematical Relationship Between Latent Change Score and Autoregressive Cross-Lagged Factor

Approaches: Cautions for Inferring Causal Relationship Between Variables. *Multivariate Behavioral Research*, *50*(6), 676–687. https://doi.org/10.1080/00273171.2015.1079696

Usami, S., Hayes, T., & McArdle, J. J. (2016). Inferring longitudinal relationships between variables: Model selection between the latent change score and autoregressive cross-lagged factor models. Structural Equation Modeling: A Multidisciplinary Journal, 23(3), 331-342. https://doi.org/10.1080/10705511.2015.1066680

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

van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Voelkle, M. C., Gische, C., Driver, C. C., & Lindenberger, U. (2018). The Role of Time in the Quest for Understanding Psychological Mechanisms. *Multivariate Behavioral Research*, *53*(6), 782–805. https://doi.org/10.1080/00273171.2018.1496813

Voelkle, M. C., & Oud, J. H. L. (2015). Relating Latent Change Score and Continuous Time Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(3), 366–381. https://doi.org/10.1080/10705511.2014.935918

Voelkle, M. C., Oud, J. H., Davidov, E., & Schmidt, P. (2012). An SEM approach to continuous time modeling of panel data: relating authoritarianism and anomia. Psychological methods, 17(2), 176. https://doi.org/10.1037/a0027543

Ziegler, M., Cengia, A., Mussel, P., & Gerstorf, D. (2015). Openness as a buffer against cognitive decline: The Openness-Fluid-Crystallized-Intelligence (OFCI) model applied to late adulthood. *Psychology and Aging*, *30*(3), 573–588. https://doi.org/10.1037/a0039493

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

Captions

Table 1. Generating parameters.

Figure 1. Expected trajectories (left panel), with measurement error only (middle panel), and with innovations only (right panel).

Figure 2. Path diagram of a stochastic BLCS model. The parameters capturing dynamic error's variances and covariance are highlighted in bold and red.

Figure 3. Latent trajectories with proportions of variance due to innovations at t=2 of 0, .05, .15 and .25.

Figure 4. Percentage of improper solutions across all conditions for the stochastic BLCS model.

Figure 5. Bias of the self-feedbacks and couplings across all conditions. The population parameter values are $\beta_x = -.35$, $\beta_y = -.25$, $\gamma_x = .1$, and $\gamma_y = .2$.

Figure 6. Standard deviation of the self-feedbacks and couplings across all conditions. The population parameter values are $\beta_x = -.35$, $\beta_y = -.25$, $\gamma_x = .1$, and $\gamma_y = .2$.

Figure 7. Bias of the standard errors of self-feedbacks and couplings across all conditions.