Routledge
Taylor & Francis Group

Multivariate Behavioral Research

ISSN: 0027-3171 (Print) 1532-7906 (Online) Journal homepage: http://www.tandfonline.com/loi/hmbr20

Misspecification in Latent Change Score Models: Consequences for Parameter Estimation, Model Evaluation, and Predicting Change

D. Angus Clark, Amy K. Nuttall & Ryan P. Bowles

To cite this article: D. Angus Clark, Amy K. Nuttall & Ryan P. Bowles (2018) Misspecification in Latent Change Score Models: Consequences for Parameter Estimation, Model Evaluation, and Predicting Change, Multivariate Behavioral Research, 53:2, 172-189, DOI: 10.1080/00273171.2017.1409612

To link to this article: https://doi.org/10.1080/00273171.2017.1409612

Published online: 04 Jan 2018.

MULTIVARIATE BEHAVIORAL RESEARCH
2018, VOL. 53, NO. 2, 172–189
https://doi.org/10.1080/00273171.2017.1409612

Routledge
Taylor & Francis Group

Misspecification in Latent Change Score Models: Consequences for Parameter Estimation, Model Evaluation, and Predicting Change

D. Angus Clark, Amy K. Nuttall, and Ryan P. Bowles

Michigan State University

ABSTRACT

Latent change score models (LCS) are conceptually powerful tools for analyzing longitudinal data (McArdle & Hamagami, 2001). However, applications of these models typically include constraints on key parameters over time. Although practically useful, strict invariance over time in these parameters is unlikely in real data. This study investigates the robustness of LCS when invariance over time is incorrectly imposed on key change-related parameters. Monte Carlo simulation methods were used to explore the impact of misspecification on parameter estimation, predicted trajectories of change, and model fit in the dual change score model, the foundational LCS. When constraints were incorrectly applied, several parameters, most notably the slope (i.e., constant change) factor mean and autoproportion coefficient, were severely and consistently biased, as were regression paths to the slope factor when external predictors of change were included. Standard fit indices indicated that the misspecified models fit well, partly because mean level trajectories over time were accurately captured. Loosening constraint improved the accuracy of parameter estimates, but estimates were more unstable, and models frequently failed to converge. Results suggest that potentially common sources of misspecification in LCS can produce distorted impressions of developmental processes, and that identifying and rectifying the situation is a challenge.

KEYWORDS
Bias; latent change score; longitudinal data analysis; misspecification; model fit

Introduction

Latent change score models (LCS) combine both the autoregressive and growth curve approaches to modeling longitudinal data, taking advantage of each technique's strengths while compensating for some of their respective limitations (McArdle, 2009). Autoregressive models capture the extent to which prior status is related to future status, but fail to provide information on the absolute trajectories of change over time. Growth curve models on the other hand capture general trajectories of change over time, but do not allow for prior status to influence future status. LCS give researchers the opportunity to simultaneously examine both autoregressive processes and general increasing or decreasing trends over time, making them a potentially valuable tool for investigating development. Accordingly, LCS are increasing in popularity across a wide variety of disciplines (Ferrer & McArdle, 2010; Wu, Selig, & Little, 2013), including education (e.g., Curby, Grimm, & Pianta, 2010), clinical psychology (e.g., King, King, McArdle, Shalev, & Doron-LaMarca, 2009), and lifespan development (e.g., McArdle, 2001; McArdle & Prindle, 2008).

Despite their comprehensive nature and flexibility, in practice LCS typically include a number of constraints on certain parameter estimates. Specifically, key parameters related to change over time (autoproportion and basis coefficients) are fixed to equality over time (McArdle, 2001; McArdle & Hamagami, 2001). These constraints do not reflect an inherent assumption of the model, and can be tested, but even when this point is acknowledged, didactic pieces (e.g., Grimm, An, McArle, Zonderman, & Resnick, 2012; Grimm et al., 2016; King et al., 2006; McArdle, 2001; McArdle & Grimm, 2010; McArdle & Hamagami, 2001) and empirical applications (e.g., Curby, Grimm, & Pianta, 2010; Ferrer et al., 2007; Finkel et al., 2009; King et al., 2009; Ghisletta & Lindenberger, 2003) almost always include such specifications without considering less constrained alternatives. As a consequence, invariance over time in certain parameter estimates has effectively become the default in these models. However,

CONTACT D. Angus Clark dangus3@tamu.edu Department of Psychology, Texas A and M University System, 4235 TAMU, College Station, Texas 77840, United States.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A110293 and R324A150063 awarded to Michigan State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. D. Angus Clark is now at the Department of Psychology at Texas A&M University. A portion of this project was presented as a poster at the 2016 Developmental Methods Conference in Whitefish, Montana.

strict invariance along these lines is unlikely in real data.

The purpose of the present study was thus to evaluate the robustness of LCS when invariance is incorrectly imposed over time in the major, change-related parameter estimates using Monte Carlo simulation methods. We specifically focus on the most foundational LCS, the univariate Dual Change Score Model (DCS), to address three main research questions: (1) the impact of incorrect equality constraints across time on parameter estimation bias, (2) the ability of popular fit indices and their commonly used cutoffs to detect misspecified constraints, and (3) the performance of LCS when parameters estimates are not constrained across time. The simulated data used to address these questions are based on an early demonstration of LCS (McArdle, 2001) that focused on the verbal and nonverbal development of elementary school children. As such, the development of verbal ability in childhood is often used in the present study as an example to illustrate various points.

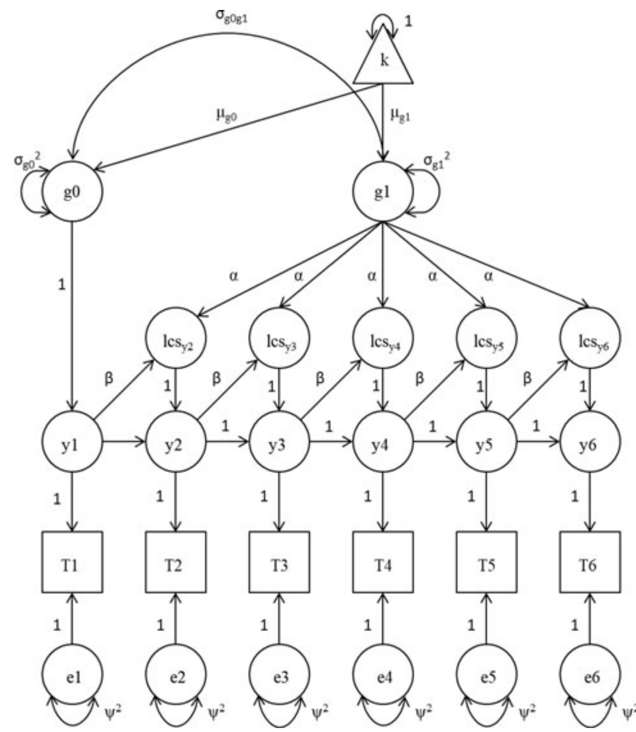

Figure 1. The Dual Change Score Model. T1-T6 = observed variables; y1-y6 = latent scores; e1-e6 = residual variance; lcs_{y2}-$lcsS_{y6}$ = latent change score factors; g0 = initial level factor; g1 = slope factor; k = constant; μ_{g0} = time 1 mean; $\sigma_{g0}{}^2$ = time one variance; μ_{g1} = slope factor mean; $\sigma_{g1}{}^2$ = slope factor variance; σ_{g0g1} = covariance between time one and slope factor; β = autoproportion coefficient; α = basis coefficient; Ψ^2 = residual variance.

Modeling developmental trends with latent change score models

An illustration of a typical LCS is provided in Figure 1. Latent change score models represent growth by breaking down change over time in some outcome construct or constructs (e.g., verbal ability) into a series of latent change score factors that capture differences between adjacent time points, or waves of assessment (lcs_{y2} through lcs_{y6} in Figure 1; McArdle & Grimm, 2010; McArdle & Hamagami, 2001). These latent change score factors are specified to be additive outcomes of two distinct developmental processes: autoproportional growth and constant growth (McArdle & Hamagami, 2001). Autoproportional growth refers to the extent that scores at a prior time point are related to subsequent increases or decreases. For example, children with higher verbal ability one year may make greater gains the next year. Autoproportional growth is represented in LCS via regression paths that flow from one time point to the immediately subsequent latent change score factor (β in Figure 1).

Constant growth refers to general increasing or decreasing trends over time. For example, verbal ability may increase continuously across elementary school. Constant growth is represented in LCS via a latent factor (g1 in Figure 1) that all change score factors load on to. The constant change factor is often compared to the slope factor of latent growth curve models given the conceptual overlap between these factors (modeling increasing or decreasing trends, and intraindividual variability in those trends, over time in observed scores or latent change scores), and the fact that a LCS that omits autoproportional growth processes is equivalent to a latent growth curve model (Grimm, Castro-Schilo, & Davoudzadeh, 2013). That is, when autoproportional paths are omitted from a LCS the model operates as a latent growth curve model with the constant change factor serving the same ultimate function as the slope factor (though trends over time are still modeled on latent change score factors). Given this, we follow the convention of using the term "slope factor" to refer to the constant change factor throughout the manuscript, even though autoproportional growth is included in these models (e.g., McArdle, 2001).

By including both autoproportional paths and the slope factor, LCS capitalizes on the advantages of both autoregressive and growth curve models (McArdle & Hamagami, 2001). That is, these models simultaneously include both general growth trends over time (which are allowed to vary across individuals, as is the case in latent growth curve models), and the degree to which prior levels of the construct are related to future change (an effect which is fixed across individuals). Furthermore, these models have the benefit of directly capturing change between time points, which is often what is of interest to developmental researchers. Indeed, with this information it is possible to both construct trajectories of the target construct over time, and more directly model determinants of change over time.

The dual change score model

The foundational LCS is the DCS (see Figure 1), named for the simultaneous incorporation of the two major change processes described above. More complex LCS (e.g., bivariate and multivariate LCS; Corker, Donnellan, & Bowles, 2013; King et al., 2009) are generally extensions or respecifications of the DCS in one way or another (Grimm et al., 2012). In the standard DCS, a single outcome variable observed at each time point is separated into systematic construct variance (y1–y6 in Figure 1), and residual variance (e1–e6 in Figure 1). The specification of the model is such that the relevant systematic variance may be identified and isolated even when there is just a single observed variable at each time point (McArdle & Nesselroade, 2014). The latent change score factors are derived using the time point specific latent constructs (i.e., the previously identified relevant, systematic variance). There are as many latent change score factors as there are waves of assessment minus one, in order to model the intercept factor (lcs_{y2} through lcs_{y6} in Figure 1). These factors capture differences between the latent variables of adjacent time points, for example, changes in verbal ability between grades 3 and 4.

Importantly, DCS are cumulative models (expected values at one time point are based on the expected values of the previous time points), using a "first differences" (Liker, Augustyniak, & Duncan, 1985) approach to deriving subsequent values from the initial value and difference (McArdle, 2001). This has the benefit of helping to compensate for temporally uneven data collection, assuming a relatively fixed change process (and presuming time is treated as a discrete instead of a continuous variable, which it typically is; McArdle & Nesselroade, 2014). If the observation interval is not constant, latent placeholder variables with values implied by the model (phantom, or placeholder, variables) can be added to maintain a consistent time scale (McArdle, 2001).

Autoproportional growth in the DCS is included via a series of autoproportional regression paths (β in Figure 1) that extend from one time point to the nearest subsequent latent change factor (though alternate specifications are possible; see Grimm, 2012). Constant change is represented with a latent slope factor (g1 in Figure 1) that all latent change score factors load on. The strength of the constant change process for a given change score factor is denoted by a basis coefficient (α in Figure 1), which acts as a factor loading tying the slope factor to the individual change factors. The slope factor is also typically correlated with another latent factor that captures the initial level of the variable under consideration (i.e., scores at time 1; g0 in Figure 1). This covariance denotes the degree to which the constant rate of change is related to participants' starting values.

The autoproportion coefficients and slope factor form the core of the DCS. As the residual variance of each latent change score factor is usually set to 0, change between time points in the DCS is wholly a function of the autoproportion coefficient (multiplied by the prior time point's score), and the slope factor value (multiplied by the basis coefficient; McArdle & Hamagami, 2001). Specifically, the model implied change between two time points is:

$$\Delta_{t,t-1} = (\alpha^* g1) + (\beta^* y_{t-1}) \tag{1}$$

where the values in the first set of parentheses represent the constant change effect (e.g., children's verbal ability increases by a constant rate between all grades included in the study), and the values in the second set of parentheses represent the autoproportion effect (e.g., children with higher levels of verbal ability at one time point make greater increases from one grade to the next). The constant change process can be thought of as setting the baseline rate of change, while the autoproportional process either accentuates or attenuates this steady change effect by serving as an accelerator or a brake on the constant change process (McArdle & Nesselroade, 2014).

The model implied latent change score values of Equation 1 can be used to calculate expected values at each time point in order to obtain a model implied trajectory (McArdle & Hamagami, 2001). For example, how does verbal ability develop, on average, over the course of the study? Thus, in addition to providing insight into both the underlying constant change and autoproportional processes, the DCS provides information on how much change is expected between time points, and how the construct of interest is expected to change over the broader course of the study. The DCS can also be extended to include determinants of change by introducing variables that predict variance in the initial level and slope factors (Malone et al., 2004; Wu et al., 2013). This can be used to address questions such as the extent to which nonverbal ability at grade 1 predicts the constant rate of change in verbal ability across elementary school.

As noted above, applications of the DCS and other LCS models typically include a number of equality constraints on key change parameters over time: all autoproportion coefficients (β in Figure 1) are constrained to equality, and every basis coefficient (α in Figure 1) is fixed to 1 (McArdle, 2001; McArdle & Hamagami, 2001). Although such constraints are not essential components of the model(s) and can be tested and relaxed, instructional and empirical applications (e.g., Curby, Grimm, & Pianta, 2010; Ferrer & McArdle, 2007; Finkel et al., 2009; Ghisletta & Lindenberger, 2003; Grimm et al., 2012; Grimm et al., 2016; King et al., 2006; King et al., 2009; McArdle, 2001; McArdle & Grimm, 2010; McArdle & Hamagami, 2001) include these constraints more often than not, and rarely if ever evaluate less constrained alternatives. To be sure, these constraints on the autoproportion and basis coefficients make the model more parsimonious, and ease the burden on estimation algorithms. Furthermore, given the complexity of these models, thoroughly testing invariance in the major change parameters over time may appear to require a prohibitive amount of data, discouraging such explorations (e.g., Ghisletta & Lindenberger, 2003). However, though these constraints may be reasonable and/or appear to substantive researchers as necessary constraints, strict invariance in the autoproportion and basis coefficients across time is generally unrealistic in real data. As such, these constraints will be incorrect to some degree and thus can potentially introduce bias into the estimation of the developmental parameters of interest.

Present study

To be sure, all models are approximations and therefore at least somewhat misspecified, and misspecification does not necessarily lead to a meaningful amount of parameter bias (MacCallum & Austin, 2000). Yet given the regularity to which time-invariant constraints on the major change parameters are utilized in applications of the DCS and other LCS, it is important to understand the potential impact of these simplifying specifications on the model. The primary aim of the present study was therefore to evaluate the extent to which incorrect impositions of invariance on the major change parameters over time (represented by equal autoproportion paths and unity basis coefficients) lead to biased parameter estimates in the DCS. Both conditional and unconditional DCS were considered (i.e., models with and without external predictors) in order to test whether the inclusion of determinants of change reduces the magnitude of any bias, and if these paths are themselves biased.

Misspecification and bias are less problematic if they are easily detectable, as researchers know to interpret with caution and consider altering their models. If a lack of invariance in the major change parameters over time can be detected, then researchers will know when non-invariant parameter estimates might need to be unconstrained. Accordingly, the second aim of this study was to investigate the ability of popular model fit statistics to detect misspecification and bias in the context of overly constrained DCS.

Of course, given the general implausibility of strict invariance in the autoproportion and basis coefficients over time, the argument could be made that it is sensible to attempt to estimate models that do not include such constraints, at the very least to evaluate their appropriateness. Again, such constraints are not necessary components of the DCS (Grimm et al., 2012), and an extension of the DCS with freely estimated basis coefficients has recently been labeled the triple change score model (McArdle & Nesselroade, 2014; McArdle, Petway, & Hishinuma, 2015). Thus, a third aim of this paper was to examine the performance of DCS when autoproportion and basis coefficients are freely estimated.

The three primary aims of this study were addressed using Monte Carlo simulation methods. Several different population models were created with varying degrees of invariance in the autoproportion and basis coefficients over time, and both unconditional and conditional DCS, with the typical invariance constraints included, were fit to the data. DCS with freely varying autoproportion coefficients and/or basis coefficients were subsequently fit to certain population models. Results are important for understanding the robustness of a theoretically powerful longitudinal model in the face of what are likely common misspecifications.

Method

Data generation

Data were generated in Mplus version 7.4 (Muthen & Muthen, 1998–2015). All population models included six waves of data, and were specified in accordance with the DCS structure in Figure 1. The population model parameters used for the baseline (i.e., invariant) model come from the analysis of verbal ability in McArdle (2001), a seminal demonstration of the DCS that drew on a widely used data set tracking children's cognitive development from 1st through 6th grade. In this model, autoproportion coefficients were constrained to equality across time, and all basis coefficients were set to 1. The population values for each part of the model can be found in Table 2. All datasets were generated with 1000 observations at 6 time points

Table 1. Autoproportion and basis coefficient sets.

	T1 → T2	T2 → T3	T3 → T4	T4 → T5	T5 → T6
Baseline Model					
Autoproportion	.09	.09	.09	.09	.09
Basis Coefficient	1.00	1.00	1.00	1.00	1.00
Autoproportion Sets					
AP-1	.23	.15	.09	.05	.02
AP-2	.09	.175	.225	.165	.115
AP-3	.00	.05	.09	.165	.23
AP-4	.60	.39	.24	.13	.05
AP-5	.60	.45	.30	.37	.25
Basis Coefficient Sets					
BC-1	1.00	1.50	2.00	2.50	3.00
BC-2	3.00	2.50	2.00	1.50	1.00
BC-3	1.00	2.00	3.00	2.00	1.00
BC-4	3.00	2.00	1.00	2.00	3.00
BC-5	3.00	3.50	4.00	4.50	5.00

Note. AP = Autoproportion Set; BC = Basis Coefficient Set. For autoproportion sets, all basis coefficients were 1, for basis coefficient sets, all autoproportion coefficients were .09.

Table 2. Average unconditional model parameter estimates for all autoproportion sets.

	μ_{g0}	σ_{g0}^2	μ_{g1}	σ_{g1}^2	σ_{g0g1}	β_1	β_2	β_3	β_4	β_5	Ψ^2
Baseline Model	20.34	20.82	2.06	.83	.74	.09	.09	.09	.09	.09	12.18
Estimate Mean	20.34	20.74	2.07	.83	.75	.09	.09	.09	.09	.09	12.19
Bias %	.00%	−.38%	.49%	.00%	1.35%	.00%	.00%	.00%	.00%	.00%	.08%
Estimate SD	.17	1.21	.24	.09	.30	.01	.01	.01	.01	.01	.27
Mean SE	.17	1.20	.24	.09	.30	.01	.01	.01	.01	.01	.27
AP-1	20.34	20.82	2.06	.83	.74	.23	.15	.09	.05	.02	12.18
Estimate Mean	20.22	21.57	10.69	5.61	8.55	− .18	− .18	− .18	− .18	− .18	12.36
Bias %	−.59%	3.60%	418.9%	575.9%	1055%	−178.26%	−220.0%	−300.0%	−460.0%	−1000%	1.48%
Estimate SD	.18	1.32	.21	.35	.53	.01	.01	.01	.01	.01	.28
Mean SE	.18	1.32	.21	.34	.51	.01	.01	.01	.01	.01	.28
AP-2	20.34	20.82	2.06	.83	.74	.09	.175	.225	.165	.115	12.18
Estimate Mean	19.48	18.57	3.07	.98	1.77	.13	.13	.13	.13	.13	13.39
Bias %	−4.23%	−10.81%	49.03%	18.07%	139.2%	44.44%	−25.71%	−42.22%	−21.21%	13.04%	9.93%
Estimate SD	.17	1.13	.17	.10	.26	.01	.01	.01	.01	.01	.30
Mean SE	.17	1.12	.18	.10	.26	.01	.01	.01	.01	.01	.30
AP-3	20.34	20.82	2.06	.83	.74	.00	.05	.09	.165	.23	12.18
Estimate Mean	20.36	19.57	− 8.28	3.73	− 8.35	.51	.51	.51	.51	.51	12.33
Bias %	.10%	−6.00%	−502.4%	349.4%	−1228%	—	920.00%	466.7%	209.09%	121.7%	1.23%
Estimate SD	.17	1.10	.32	.33	.58	.01	.01	.01	.01	.01	.28
Mean SE	.17	1.08	.31	.33	.58	.01	.01	.01	.01	.01	.28
AP-4	20.34	20.82	2.06	.83	.74	.60	.39	.24	.13	.05	12.18
Estimate Mean	19.33	18.67	20.07	18.12	17.37	− .16	− .16	− .16	− .16	− .16	15.13
Bias %	−4.97%	−10.33%	874.3%	2083%	2247%	−126.7%	−141.0%	−166.7%	−223.1%	−420.0%	24.22%
Estimate SD	.17	1.26	.18	.90	.90	.002	.002	.002	.002	.002	.34
Mean SE	.18	1.27	.19	.88	.88	.003	.003	.003	.003	.003	.34
AP-5	20.34	20.82	2.06	.83	.74	.60	.45	.30	.37	.25	12.18
Estimate Mean	20.32	20.56	10.34	4.97	8.93	.19	.19	.19	.19	.19	14.23
Bias %	−.10%	−1.25%	401.9%	498.8%	1107%	−68.33%	−57.78%	−36.67%	−48.65%	−24.00%	16.83%
Estimate SD	.17	1.23	.14	.30	.48	.002	.002	.002	.002	.002	.32
Mean SE	.18	1.22	.14	.29	.46	.002	.002	.002	.002	.002	.32

Note. AP = Autoproportion set; SD = Standard Deviation; SE = Standard Error; μ_{g0} = time 1 mean; σ_{g0}^2 = time one variance; μ_{g1} = constant change factor mean; σ_{g0}^2 = constant change factor variance; σ_{g0g1} = covariance between time one and constant change factor; β = autoproportion coefficient; Ψ^2 = residual variance. All basis coefficients fixed to one.

and no missing data, representing an ideal situation for longitudinal data analysis (i.e., consistent data collection and no attrition). For every condition described below, 1000 unique data sets were generated and analyzed.

Study conditions

Several major features of the data/model were systematically varied. The first feature to be manipulated was the pattern of population autoproportion coefficients. In all, five different sets of autoproportion coefficients were used (AP1–AP5; see Table 1 for specific values), as well as the baseline set. The baseline model included equal autoproportion coefficients at each time point; these values come directly from the model presented in McArdle (2001). The five other sets included different autoproportion coefficients at each time point, and were chosen to represent a diverse array of patterns and values[1] while remaining consistent with the metric used in the original study (i.e., percent correct scores), and the general developmental trend of verbal ability in early life (i.e., increasing). AP-1 values were selected by randomly generating 4 values between .01 and .25, that when combined with .09 (the original autoproportion value and fifth number of this set), averaged to between .09 and .12 (final average was .108). These values were then placed in descending order. AP-2 and AP-3 included coefficients of similar (but not exact) magnitude to AP-1 that were placed in increasing and then descending, and descending, order respectively. AP-4 consisted of larger coefficients (to represent a more pronounced autoproportion process) that decreased at a rate comparable to AP-1. AP-5 consisted of similarly large coefficients that did not decrease as dramatically over time. Across conditions, only the autoproportion coefficients varied; the other population parameters remained constant (see Table 2). The trajectories over time implied by these population models can be found in Table 4, and Figure 2. The population trajectories are non-linear and increasing over time, which is consistent with the original data, what would be expected for verbal ability in early life (McArdle, 2001), and one of the oft highlighted advantages of the DCS (i.e., flexibly modeling non-linear change; Grimm et al., 2013).

[1] All autoproportion values reported in the present study were positive. Similar conclusions emerged however when autoproportion values were negative, or there was a mix of positive and negative values.

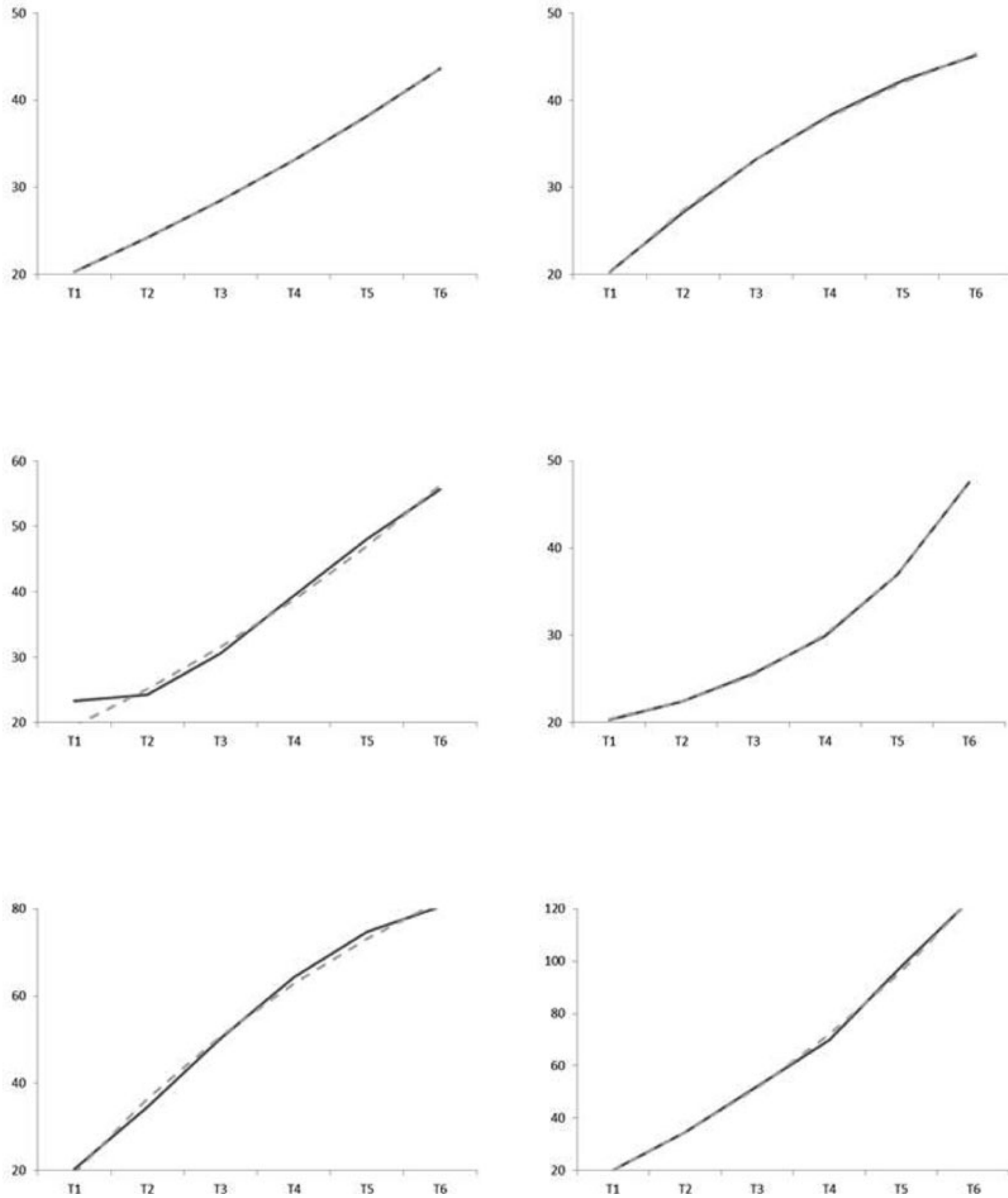

Figure 2. Average population and model implied trajectories for the six autoproportion conditions. Solid line denotes population trajectory, dashed line denotes model implied trajectory. Time point on x-axis, scores on y-axis. All autoproportion conditions presented in order, starting at top left: Baseline, AP-1, AP-2, AP-3, AP-4, AP-5.

Next, in a separate set of simulations, basis coefficients were manipulated while the autoproportion coefficients remained constant. Five different sets of basis coefficients were used (BC1—BC5; see Table 1 for specific values) in addition to the baseline set. In the baseline model, all basis coefficients were set to 1, as is convention with the DCS. In the 5 other sets, basis coefficients were different at each time point. Only the basis coefficients were manipulated across the population models, all other parameters were held constant. Values were chosen to represent a diverse assortment of constant change patterns and magnitudes. BC-1 and BC-2 captured steadily increasing and decreasing growth trends, respectively. BC-3 and BC-4 captured growth trends that both increased and decreased across time. Finally, BC-5 captured a steadily increasing, but more pronounced, growth trend. The specific values and rates of change over time were selected to be reasonable in light of the typical values assigned (i.e., 1), and the mean of the slope factor. The final population model trajectories based on these values can be found in Figure 3. Again, non-linear increasing trends over time are represented, however the non-linearity is less pronounced than in the autoproportion conditions. This is partly the consequence of an invariant, modest autoproportion process in each model (the autoproportion process is largely responsible

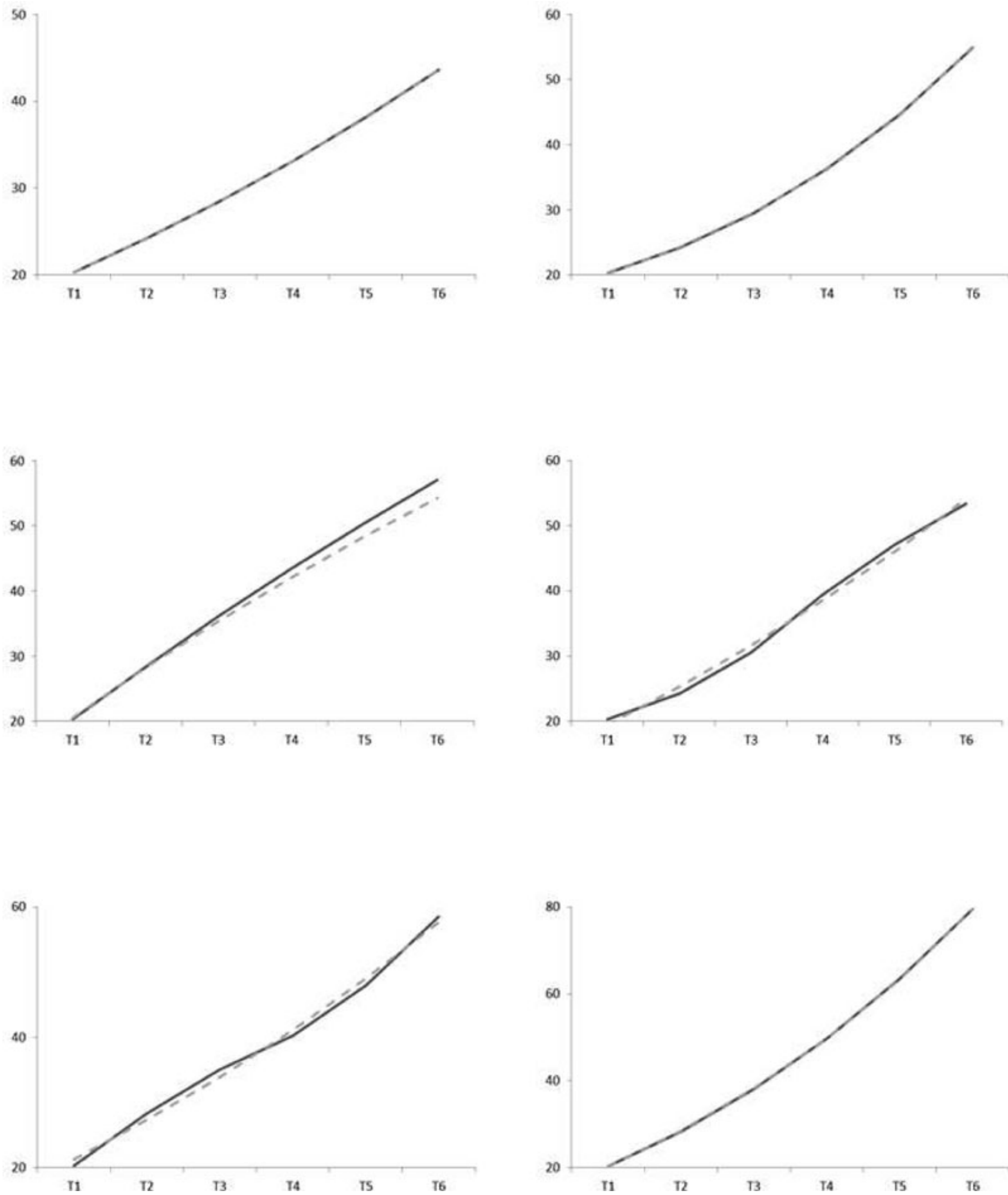

Figure 3. Average population and model implied trajectories for the six basis coefficient conditions. Solid line denotes population trajectory, dashed line denotes model implied trajectory. Time point on x-axis, scores on y-axis. All autoproportion conditions presented in order, starting at top left: Baseline, BC-1, BC-2, BC-3, BC-4, BC-5.

for decelerating and accelerating trends; McArdle & Nesselroade, 2014).

The third data feature to be manipulated was the presence or absence of an external predictor variable for the initial level and slope factors. After first running analyses for each autoproportion and basis coefficient set without external predictor variables included in the population model, an external variable was added to the population model that predicted both the initial level factor, and the slope factor. This external variable was based on the 1st grade nonverbal ability variable from the data used in McArdle (2001; $M = 17.977$ $SD = 8.33$). Three conditional population models were generated for each of the autoproportion and basis coefficient sets, differentiated based on the size of the predictor effects, which represented either weak, moderate, or strong effects (i.e., standardized regression coefficients of .20, .40, and .60, respectively).

Data analytic strategy

For each condition, a standard DCS was fit to each of the 1000 generated data sets. Unconditional models were examined first, followed by conditional models. As the typical DCS includes constrained autoproportion coefficients and unity basis coefficients over time, these models were correctly specified for only the baseline model, and misspecified for every other set. Individual Mplus input files were generated and run for each simulated data set using the Mplus Automation Package (Hallquist & Wiley, 2014) in R (R Development Core Team, 2016). The Mplus Automation package was also used to extract and consolidate parameter and fit information from the individual output files.

For each condition, the average parameter estimates across all replications were compared with the population values. Deviations from the population values were quantified with a percent bias statistic[2] that denotes the degree to which the average estimated value differs from the population value (e.g., a value of 30 indicates that the estimated value is 30% larger than the population value). Following this, the average parameter estimates were used to calculate the average model implied change between each time points, and the average model implied trajectory over the time points. Model implied scores were compared to the corresponding population values. The comparison of individual values was conducted using the percent bias statistic. Additionally, the root mean square error (RMSE) was calculated using the entire series of model implied and population scores. This provides a single value that holistically captures the overall bias in the model implied values over time.

Model fit across the misspecified DCS was evaluated by considering five of the most common structural equation model fit indices (West, Taylor, & Wu, 2012): the chi-square test (χ^2), the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker Lewis index (TLI). Information criteria (e.g., Bayesian information criteria; BIC) were not considered here as the focus was largely on absolute fit (versus comparative fit), and because model fit indices appear to perform better than information criteria in evaluating LCS (Usami, Hayes, & McArdle, 2016). Four statistics for each of the fit indices of interest were calculated. For the χ^2, RMSEA, SRMR, CFI, and TLI, the mean and standard deviation across all replications within a condition was computed. For the χ^2, the percentage of models that demonstrated significant misfit at both the .05 and .01 level was computed. For the RMSEA and SRMR, the percentage of models with values below .08 and .05 was computed (lower values denote better fit). For the CFI and TLI, the percentage of models with values above .90 and .95 was computed (higher values denote better fit). These thresholds represent the commonly invoked standards for "adequate" and "excellent" fitting models (Browne & Cudeck, 1993; Hu & Bentler, 1999; West et al., 2012). The percentage of models evidencing "adequate" and "excellent" fit were calculated excluding models that did not converge.

The effects of loosening constraints on the autoproportion and basis coefficients over time were examined by fitting less constrained models to the AP-1 population model. Although only one population model was used here for the sake of parsimony, similar results emerged when other population models were used, including population models with unequal basis coefficients, and unequal autoproportion and basis coefficients. Two major types of models were fit to the generated data (a slightly altered version was also fit, and is described below). The first matched the population model perfectly such that only the autoproportion coefficients were freely estimated. In the second, both autoproportion and basis coefficients were allowed to freely estimate at all time points (though 1 basis coefficient was fixed to 1 for identification).

Results

The results for each major question are presented in turn. Exploratory follow-up analyses are also briefly reviewed. With the exception of the unconstrained models, no models failed to converge. Non-convergence rates for the unconstrained models are discussed in more detail below.

Misspecified autoproportion path constraints

Parameter estimates

Parameter population values, estimate averages, standard deviations, average estimated standard errors, and percent bias, are presented in Table 2. When the model was correctly specified (baseline model) there was virtually no bias in the parameter estimates. However, parameter estimates were quite biased when the autoproportion coefficients differed across time in the population model. For example, for AP-1, parameter bias ranged from .6% (initial level factor mean) to 1055% (initial level factor-slope factor covariance). Though there was little bias in the estimates of the initial level factor mean and variance, the mean and variance of the slope factor was substantially inflated (by 419% and 576%, respectively). Furthermore, the average autoproportion coefficient estimate (−.18) was well outside the range of any of the population coefficients, and was even opposite in sign (all population coefficients were positive).

Across the unequal autoproportion sets, parameter bias was severe for the parameters most relevant to change over time: the autoproportion coefficient and the parameters related to the slope factor. The initial level factor mean and variance, as well as the residual variance, were relatively unaffected, as would be expected given their relation to the change processes and factors. Typically,

[2] Calculated via the formula 100*((E-P)/P) where E refers to the average estimated value, and P refers to the actual population value

but not always, the direction of the bias was such that the slope factor mean and variance were overestimated, and the autoproportion value was underestimated, often to the point that it was outside the range of all population values. Notably, estimates were biased, but consistent. The standard deviations of parameter estimates across replications were small, and the estimated standard errors accurately reflected this.

Growth trajectories

Model implied change between time points, and the corresponding population values, are presented in Table 3. When the DCS was correctly specified, the average model implied change between time points almost perfectly matched the population change between time points. When the DCS were misspecified, the model implied change between time points still somewhat accurately (i.e., bias generally $< 20\%$), given the degree of parameter bias, captured the true change between time points. Across all 5 sets of autoproportion coefficients, the average bias in model implied change scores was 3.60% (range from 0% to 95%; RMSE from .14 to 2.30).

The average population trajectories over time were accurately captured across all sets of autoproportion coefficients (see Table 4). The trajectory for the correctly specified model evidenced essentially no bias. For the incorrect models, across all sets and time points, bias ranged from only 0.1% to 5% (RMSE from .08 to 1.13; see Figure 2).

Table 3. Average unconditional model implied change for all autoproportion sets.

	T1 → T2	T2 → T3	T3 → T4	T4 → T5	T5 → T6
Baseline Model	3.89	4.24	4.62	5.04	5.49
Model Implied	3.89	4.24	4.62	5.03	5.48
Bias %	.00%	.00%	.00%	−.20%	−.18%
AP-1	6.74	6.12	5.05	3.97	2.90
Model Implied	7.13	5.87	4.84	3.99	3.28
Bias %	5.79%	−4.08%	−4.16%	.50%	13.10%
AP-2	3.89	6.30	8.93	8.57	7.58
Model Implied	5.65	6.40	7.24	8.20	9.29
Bias %	45.24%	1.59%	−18.93%	−4.32%	22.56%
AP-3	2.06	3.18	4.36	7.00	10.56
Model Implied	2.03	3.06	4.61	6.94	10.45
Bias %	−1.46%	−3.77%	5.73%	−.86%	−1.04%
AP-4	14.26	15.56	14.10	10.41	5.79
Model Implied	17.06	14.39	12.15	10.25	8.65
Bias %	19.64%	−7.52%	−13.83%	−1.54%	49.40%
AP-5	14.26	17.63	17.73	27.95	26.54
Model Implied	14.26	17.01	20.29	24.21	28.88
Bias %	.00%	−3.52%	14.44%	−13.38%	8.82%

Note. Change between time points calculated as: $(\mu_{g1} * \alpha) + (\beta_x * T_x)$ where β_x represents the autoproportion coefficient from the earlier to the later time point, and T_x represents the scores at the earlier time point. AP = Autoproportion set.

Table 4. Average unconditional model implied trajectory for all autoproportion sets.

	T1	T2	T3	T4	T5	T6
Baseline Model	20.34	24.23	28.47	33.09	38.13	43.62
Model Implied	20.34	24.23	28.47	33.09	38.12	43.60
Bias %	.00%	.00%	.00%	.00%	−.03%	−.05%
AP-1	20.34	27.08	33.20	38.25	42.22	45.13
Model Implied	20.22	27.35	33.22	38.05	42.04	45.32
Bias %	−.59%	1.00%	.06%	−.52%	−.43%	.42%
AP-2	23.34	24.23	30.53	39.46	48.03	55.62
Model Implied	19.48	25.14	31.53	38.78	46.98	56.27
Bias %	−16.54%	3.76%	3.28%	−1.72%	−2.19%	1.17%
AP-3	20.34	22.40	25.58	29.94	36.94	47.50
Model Implied	20.36	22.39	25.45	30.05	36.99	47.44
Bias %	.10%	−.04%	−.51%	.37%	.14%	−.13%
AP-4	20.34	34.60	50.16	64.26	74.67	80.47
Model Implied	19.33	36.39	50.78	62.93	73.18	81.83
Bias %	−4.97%	5.17%	1.24%	−2.07%	−2.00%	1.69%
AP-5	20.34	34.60	52.24	69.97	97.91	124.45
Model Implied	20.32	34.58	51.59	71.88	96.09	124.96
Bias %	−.10%	−.06%	−1.24%	2.73%	−1.86%	.41%

Note. Average scores at each time point calculated as: $t-1 + t-1 \rightarrow T_{x+1}$ where $t-1$ represents the average score at the immediately prior time point, and $\Delta_{t,t-1}$ represents the average change between the immediately prior time point and the current time point. AP = Autoproportion set.

Model fit

Fit information for all of the models can be found in Table 5. When the models were correctly specified, fit indices uniformly indicated excellent fit. However, fit indices were only sporadically able to detect the misspecification and parameter bias when the models were incorrectly specified. The χ^2 test and RMSEA were most likely to reject misspecified models, but still often indicated that incorrect and substantially biased models were acceptable. The SRMR, CFI, and TLI suggested that all models at least fit the data adequately. In fact, the SRMR and TLI indicated that most ($> 95\%$) models except those for AP-4 fit excellently.

Conditional models

Parameter population values, estimate averages, standard deviations, average estimated standard errors, and percent bias, are presented in Table 6. As results were very similar across the weak, moderate, and strong predictor models, only the results from the strong predictor models are presented. When models were correctly specified there was effectively no bias. When autoproportion values were incorrectly constrained to equality across time, the bias for many parameters was severe. Bias in the autoproportion coefficients ranged from 15% (AP-2) to 900% (AP-1), and bias in the slope factor mean ranged from 50% (AP-2) to 851% (AP-4). The path from the external predictor to the initial level factor was estimated relatively accurately across autoproportion sets (bias

Table 5. Fit information for all autoproportion sets.

	χ^2	RMSEA	SRMR	CFI	TLI
Baseline Model					
M	20.07	.007	.025	1.00	1.00
SD	6.50	.009	.009	.001	.001
% Adequate	99%	100%	100%	100%	100%
% Excellent	94%	100%	100%	100%	100%
AP-1					
M	51.87	.039	.031	.995	.996
SD	12.68	.008	.008	.002	.002
% Adequate	12%	100%	100%	100%	100%
% Excellent	4%	89%	98%	100%	100%
AP-2					
M	394.56	.137	.041	.941	.956
SD	36.46	.007	.006	.006	.004
% Adequate	0%	0%	100%	100%	100%
% Excellent	0%	0%	91%	5%	89%
AP-3					
M	42.12	.032	.033	.996	.997
SD	11.97	.01	.008	.002	.002
% Adequate	39%	100%	100%	100%	100%
% Excellent	18%	98%	96%	100%	100%
AP-4					
M	839.13	.202	.058	.913	.935
SD	52.38	.006	.006	.006	.004
% Adequate	0%	0%	100%	99%	100%
% Excellent	0%	0%	8%	0%	0%
AP-5					
M	632.82	.175	.036	.942	.956
SD	45.80	.007	.007	.004	.003
% Adequate	0%	0%	100%	100%	100%
% Excellent	0%	0%	96%	2%	97%

Note. AP = Autoproportion set; M = Mean; SD = Standard deviation; % Adequate = Percentage of replications that crossed fit thresholds for being deemed adequate (χ^2 $p > .01$, RMSEA $< .08$, SRMR $< .08$, CFI $> .90$, TLI $> .90$); % Excellent = Percentage of replications that crossed fit thresholds for being deemed excellent (χ^2 $p > .05$, RMSEA $< .05$, SRMR $< .05$, CFI $> .95$, TLI $> .95$).

ranged from 2% to 9%), whereas the path to the slope factor was often quite distorted (bias ranged from 29% to 488%). Information regarding the fit of the conditional models is presented in Table 7. Fit indices again did not reliably signal (based on traditional thresholds) that there was often a substantial amount of parameter bias.

Misspecified basis coefficient constraints

Parameter estimates

The results obtained from manipulating the population basis coefficients and fitting overly constrained DCS were analogous to the results obtained when manipulating autoproportion coefficients. As such, these results are only presented in text briefly. When models were misspecified, there was again substantial bias. Bias was most pronounced in the autoproportion coefficient estimate, and estimates related to the slope factor. Bias was minimal in the initial level factor mean and variance, and the residual variance, estimates. Bias in the autoproportion estimate ranged from a low of 5% (BC-3) to a high of 173% (BC-2). Bias in the slope mean ranged from 94% (BC-4) to 338% (BC-2). Bias tended to be smaller in magnitude when there was not a monotonically increasing or decreasing trend in the basis coefficients.

Growth trajectories

For model implied change between adjacent time points, bias ranged from 1% to 50% across all basis coefficient sets and time points. RMSE values for the sets of model implied difference scores ranged from .08 (BC-1) to 1.51 (BC-4). As for the overall trajectory, bias ranged from .04% (BC-1) to 5% (BC-4) across the individual time points. RMSE values for the total model implied trajectories ranged from .05 (BC-1) to 1.5 (BC-2) (see Figure 3).

Model fit

Fit indices generally indicated that the misspecified models were acceptable or excellent. Every fit index considered with the exception of the χ^2 indicated that the DCS fit BC- 1, BC-2, and BC-5 at least adequately, and most often excellently. Only the RMSEA did not indicate an excellent fit for BC-2. BC-3 and BC-4 were the least well fitting, but although the RMSEA routinely rejected these models, the CFI, TLI, and SRMR still indicated at least adequate fit, even though these sets generally yielded the least amount of estimation bias. The χ^2 rejected most models, but it had low power for rejecting models fit to BC- 1 and BC-5, with only 75% and 36% of these models being rejected at the .05 level.

Freely estimating coefficients

Free autoproportion coefficients

Results from the unconstrained models are presented in Table 8; fit information for these models is presented in Table 9. When models with unconstrained autoproportion coefficients were fit to AP-1 (Free AP in Tables 8 and 9), there was little bias in the average parameter estimates across replications, and every marker of model fit indicated that most or all replications fit excellently. However, there was substantial variability in the parameter estimates across replications. For example, the average estimate for the slope mean was 2.09, while the standard deviation was 2.50, a value larger than the average estimate itself. The standard errors accurately reflected this instability, which had the side effect of greatly reducing power to the point that most autoproportion coefficients were non-significant (power for the five coefficients across replications was: 44%, 32%, 16%, 8%, and 6%).

Free autoproportion and basis coefficients

We then considered the least constrained model with freely estimated autoproportion and basis coefficients.

Table 6. Average conditional model parameter estimates for all autoproportion sets.

	μ_{g0}	σ_{g0}^2	μ_{g1}	σ_{g1}^2	σ_{g0g1}	β_1	β_2	β_3	β_4	β_5	Ψ^2	β_{g0}	β_{g1}
Baseline Model	20.34	20.82	2.06	.83	.74	.09	.09	.09	.09	.09	12.18	.45	.085
Estimate Mean	20.35	20.77	2.06	.83	.73	.09	.09	.09	.09	.09	12.18	.45	.085
Bias %	.05%	−.24%	0%	0%	−1.35%	0%	0%	0%	0%	0%	0%	0%	0%
Estimate SD	.39	1.27	.18	.08	.26	.01	.01	.01	.01	.01	.27	.02	.01
Mean SE	.39	1.20	.18	.08	.25	.01	.01	.01	.01	.01	.27	.02	.01
AP-1	20.34	20.82	2.06	.83	.74	.23	.15	.09	.05	.02	12.18	.45	.085
Estimate Mean	20.39	21.82	10.11	5.06	8.12	−.16	−.16	−.16	−.16	−.16	12.40	.44	.28
Bias %	.25%	4.80%	390.8%	509.6%	997.3%	−169.6%	−206.67%	−277.8%	−420%	−900%	1.81%	−2.22%	229%
Estimate SD	.41	1.39	.21	.29	.50	.004	.004	.004	.004	.004	.27	.01	.02
Mean SE	.41	1.32	.22	.28	.47	.004	.004	.004	.004	.004	.28	.02	.01
AP-2	20.34	20.82	2.06	.83	.74	.09	.175	.225	.165	.115	12.18	.45	.085
Estimate Mean	19.49	17.99	3.08	.89	1.98	.132	.132	.132	.132	.132	14.66	.43	.11
Bias %	−4.18%	−13.59%	49.5%	7.23%	167.6%	46.7%	−24.6%	−41.3%	−20.0%	14.8%	20.4%	−4.4%	29.4%
Estimate SD	.38	1.18	.15	.09	.25	.003	.003	.003	.003	.003	.33	.02	.01
Mean SE	.38	1.12	.16	.09	.24	.004	.004	.004	.004	.004	.33	.02	.01
AP-3	20.34	20.82	2.06	.83	.74	.00	.05	.09	.165	.23	12.18	.45	.085
Estimate Mean	20.20	19.34	−7.44	3.14	−7.55	.47	.47	.47	.47	.47	12.34	.46	−.15
Bias %	−.69%	−7.1%	−461%	278%	−1120%	—	840%	422%	185%	104%	1.31%	2.22%	−276%
Estimate SD	.37	1.14	.26	.26	.53	.001	.001	.001	.001	.001	.28	.02	.01
Mean SE	.37	1.07	.25	.25	.50	.001	.001	.001	.001	.001	.28	.02	.01
AP-4	20.34	20.82	2.06	.83	.74	.60	.39	.24	.13	.05	12.18	.45	.085
Estimate Mean	19.57	17.34	19.59	17.08	17.52	−.15	−.15	−.15	−.15	−.15	18.06	.41	.50
Bias %	−3.79%	−16.7%	851%	1958%	2268%	−125%	−138%	−163%	−215%	−400%	48.3%	−8.89%	488%
Estimate SD	.41	1.34	.33	.86	.92	.002	.002	.002	.002	.002	.39	.02	.02
Mean SE	.40	1.29	.33	.82	.86	.002	.002	.002	.002	.002	.40	.02	.02
AP-5	20.34	20.82	2.06	.83	.74	.60	.45	.30	.37	.25	12.18	.45	.085
Estimate Mean	20.43	19.81	10.15	4.62	9.17	.20	.20	.20	.20	.20	16.38	.44	.28
Bias %	.44%	−4.85%	392%	457%	1139%	−66.7%	−55.6%	−33.3%	−46.0%	20.0%	34.5%	−2.22%	299%
Estimate SD	.40	1.29	.19	.28	.48	.002	.002	.002	.002	.002	.36	.02	.01
Mean SE	.39	1.23	.20	.28	.46	.002	.002	.002	.002	.002	.37	.02	.01

Note. AP = Autoproportion set; SD = Standard Deviation; SE = Standard Error; μ_{g0} = time 1 mean; $\sigma_{g0}{}^2$ = time one variance; μ_{g1} = constant change factor mean; $\sigma_{g0}{}^2$ = constant change factor variance; σ_{g0g1} = covariance between time one and constant change factor; β = autoproportion coefficient; Ψ^2 = residual variance; β_{g0} = regression path from external variable to time 1 mean factor; β_{g1} = regression path from external variable to constant change factor. All basis coefficients fixed to one.

Results are presented in Table 8 (Free AP, BC); fit information is presented in Table 9. This model encountered serious estimation difficulties; 773 of 1000 replications failed to converge. The results in Tables 8 and 9 should thus be interpreted cautiously, as they only pertain to those few replications that successfully converged. Of the 227 models that did converge, bias was smaller than in the overly constrained models, but more pronounced than in the Free AP model (e.g., average bias across autoproportion coefficients = 15% versus 0%). The CFI, TLI, and SRMR indicated that these models fit excellently, but the RMSEA and χ^2 rejected most models that converged. Compared to the Free AP model, parameter estimates were more stable across the replications (e.g., standard deviation for the slope mean = .38). However, the estimated standard errors were incredibly large and inaccurate (e.g., average standard error for the slope mean = 90.07). Thus, most models that were fully unconstrained failed to converge, and those that did still evinced problems suggesting a general instability.

In an attempt to improve the performance of the Free AP, BC model and facilitate convergence, one extra constraint was added such that the first two basis coefficients were fixed to 1 instead of just the first (Free AP, BC* Tables 8 and 9). Notably, this constraint accurately reflects the underlying population model. With this model, 204 replications failed to converge out of the total 1000 runs, however, parameter estimates were both biased and unstable. For example, the average slope factor mean was −2.37 (bias = 215%), and the standard deviation was 12.93. Standard errors were sporadic in over-estimating versus under-estimating the actual degree of variation. Despite this parameter bias and instability, most models (>95%) fit excellently according to all markers of fit considered here.

Exploratory follow-up analyses

Two potential issues with the DCS highlighted by the results of the main analyses are that the unjustified application of standard constraints can result in substantial parameter bias, and that popular indices of model fit, or at least their commonly invoked thresholds, do not reliably indicate that parameters are exceptionally biased. Across models, the autoproportion coefficients and parameter estimates associated with the slope factor

Table 7. Fit information for conditional autoproportion sets.

	χ^2	RMSEA	SRMR	CFI	TLI
Baseline Model					
M	24.08	.01	.02	1.00	1.00
SD	6.96	.01	.01	.001	.001
% Adequate	99%	100%	100%	100%	100%
% Excellent	96%	100%	100%	100%	100%
AP-1					
M	74.06	15.77	.02	.995	.995
SD	15.77	.01	.01	.002	.001
% Adequate	1%	100%	100%	100%	100%
% Excellent	0%	74%	100%	100%	100%
AP-2					
M	764.48	.18	.04	.924	.934
SD	51.40	.01	.004	.01	.01
% Adequate	0%	0%	100%	100%	100%
% Excellent	0%	0%	98%	0%	0%
AP-3					
M	54.99	.04	.02	.996	.997
SD	12.42	.001	.01	.002	.001
% Adequate	16%	100%	100%	100%	100%
% Excellent	6%	98%	100%	100%	100%
AP-4					
M	1551.01	.25	.06	.881	.896
SD	67.91	.01	.01	.01	.01
% Adequate	0%	0%	100%	0%	16%
% Excellent	0%	0%	7%	0%	0%
AP-5					
M	1200.76	.22	.04	.916	.926
SD	60.34	.01	.01	.004	.004
% Adequate	0%	0%	100%	100%	100%
% Excellent	0%	0%	97%	0%	0%

Note. AP = Autoproportion set; M = Mean; SD = Standard deviation; % Adequate = Percentage of replications that crossed fit thresholds for being deemed adequate (χ^2 $p > .01$, RMSEA $< .08$, SRMR $< .08$, CFI $> .90$, TLI $> .90$); % Excellent = Percentage of replications that crossed fit thresholds for being deemed excellent (χ^2 $p > .05$, RMSEA $< .05$, SRMR $< .05$, CFI $> .95$, TLI $> .95$).

(mean, variance) tended to have the most serious bias. We therefore examined the correlations between parameter estimates across replications; the correlations between parameters for Baseline Model are presented in Table 10. Although there were several moderate sized correlations, the highest value by far was the correlation between the autoproportion coefficient and the slope factor mean, at $r = -.99$. Similarly large correlations were also evident for the association between the autoproportion coefficient and slope factor mean across the other conditions as well. For AP-1 through AP-5 the correlations between these parameters were −.92, −.97, −.97, −.63, and −.82, respectively. For BC- 1 through BC-5, the correlations between these parameters were −.96, −.90, −.93, −.95, and −.81, respectively.

The correlation between the autoproportion coefficient and slope factor mean implies that the model may have difficulty distinguishing the two growth processes of interest. In essence, this comes down to an issue of multicollinearity; the two predictors of growth over time (i.e., constant and autoproportional processes) are so highly correlated that estimates become unstable. Problems with multicollinearity are often indicative of an inadequate amount of information (Farrar & Glauber, 1967). In this context the amount of available information is largely tied to the number of waves of data. Thus, increasing the number of waves of data may reduce the correlation between the autoproportion coefficient and slope factor mean. To test this, extra waves of data were added to the Baseline Model. When there were 10 waves of data instead of 6, the correlation between the autoproportion coefficient and slope factor mean dropped from −.99 to −.91. When there were 15 waves of data the correlation was −.64. Finally, when there were 20 waves of data the correlation between these two parameters was only −.12. Thus, it took 14 additional waves of data to reduce the correlation between these two parameters to relative triviality.

Discussion

This paper examined the consequences of restrictive constraints on the parameter estimates and fit statistics of dual change score models (DCS), which are the basic and foundational latent change score model (LCS). DCS are typically specified including constraints in which the major growth parameters (autoproportion and basis coefficients) are invariant across time. In real data these restrictions are likely to be inaccurate, yet they offer the advantage of more parsimonious and easy to estimate models. The results of the current study show that when invariance does not hold in the data, but is still imposed, parameter estimates may become exceptionally biased. Further, fit statistics are unreliable indicators regarding the degree of misspecification and parameter bias. Although this potentially suggests freely estimating all parameters to attempt to assess the appropriateness of such constraints across time, this approach has its own corresponding limitations and pitfalls discussed below. We thus caution readers against reaching such conclusions from the present results.

Summary

If either autoproportion or basis coefficients varied over time, constraining them to equality introduced a substantial amount of bias. Bias was most pronounced in estimates of the autoproportion coefficient, and parameters related to the slope factor (i.e., slope factor mean, variance, and covariance with the initial level factor). As such, bias was most prevalent in the parameters that are most relevant for capturing change over time: in other words, the parameters that tend to be of the most substantive interest. Notably, the estimate of the autoproportion coefficient often fell well outside the range of the population model autoproportion coefficients.

Table 8. Results for unconstrained models.

	μ_{g0}	σ_{g0}^2	μ_{g1}	σ_{g1}^2	σ_{g0g1}	α_1	α_2	α_3	α_4	α_5	β_1	β_2	β_3	β_4	β_5	Ψ^2
Free AP	20.34	20.82	2.06	.83	.74	1.00	1.00	1.00	1.00	1.00	.23	.15	.09	.05	.02	12.18
Estimate Mean	20.34	20.78	2.09	1.12	.76	1.00	1.00	1.00	1.00	1.00	.23	.15	.09	.05	.02	12.17
Bias %	0%	−.20%	1.5%	35%	2.7%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	−.10%
Estimate SD	.18	1.23	2.50	.56	2.27	.00	.00	.00	.00	.00	.12	.09	.08	.07	.06	.27
Mean SE	.18	1.22	2.52	.58	2.28	.00	.00	.00	.00	.00	.12	.09	.08	.07	.06	.27
Free AP, BC	20.34	20.82	2.06	.83	.74	1.00	1.00	1.00	1.00	1.00	.23	.15	.09	.05	.02	12.18
Estimate Mean	20.35	20.71	1.57	.82	.39	1.00	1.21	1.14	1.09	1.15	.25	.16	.10	.06	.03	12.14
Bias %	.05%	−.50%	−24%	−1.0%	−47%	0%	21%	14%	9.0%	15%	8.7%	6.7%	11%	20%	50%	−.30%
Estimate SD	.17	1.22	.38	.45	.39	.00	.72	.59	.62	.71	.02	.04	.03	.02	.02	.29
Mean SE	.18	1.22	79.71	2.99	81.12	.00	3.59	7.16	9.94	13.39	3.92	3.49	2.93	2.51	2.48	.27
Free AP, BC*	20.34	20.82	2.06	.83	.74	1.00	1.00	1.00	1.00	1.00	.23	.15	.09	.05	.02	12.18
Estimate Mean	20.35	20.28	− 2.36	8.80	− 3.29	1.00	1.00	.34	.66	1.25	.45	.31	.34	.10	− .07	12.52
Bias %	.05%	−2.6%	−214%	960%	−544%	0%	0%	−66%	−34%	25%	96%	107%	278%	100%	−450%	2.8%
Estimate SD	.18	1.65	12.93	16.85	12.45	.00	.00	2.57	1.63	2.54	.64	.48	1.60	.51	.65	.93
Mean SE	.18	1.21	19.16	24.10	18.46	.00	.00	1.80	2.08	7.34	.94	.71	1.16	.81	1.96	.28

Note. Free AP = models estimated with freely estimating autoproportion coefficients; Free AP, BC = models estimated with freely estimating autoproportion and basis coefficients; Free AP, BC* = models estimated with freely estimating autoproportion and mostly freely estimating basis coefficients; SD = Standard Deviation; SE = Standard Error; μ_{g0} = time 1 mean; σ_{g0}^2 = time one variance; μ_{g1} = constant change factor mean; σ_{g0}^2 = constant change factor variance; σ_{g0g1} = covariance between time one and constant change factor; β = autoproportion coefficient; α_1 = basis coefficient; β = autoproportion coefficient; Ψ^2 = residual variance.

For example, the estimated autoproportion coefficient for AP-1 was −.18, which is larger in magnitude and of opposite sign than all the population coefficients; it is nowhere near the average of the population coefficients as might be expected. Furthermore, estimates were both biased *and* stable; across replications the DCS tended to be consistent in its inaccurate estimates. This low variability in estimates indicates that misspecified models will reliably over or under estimate many parameters.

In addition to being large in magnitude, bias was somewhat unpredictable in direction. Although with most autoproportion sets the slope mean was overestimated and the autoproportion values were underestimated, with AP-3 the slope factor mean was underestimated, and the autoproportion value was overestimated. One distinguishing feature of AP-3 was that autoproportion values increased monotonically, whereas in many other sets they decreased. The direction of bias may thus partly be a function of the pattern of autoproportion coefficients in the population. This is not an especially useful insight however as in practical applications the population values are unknown, and moreover, actual autoproportion values are unlikely to follow a strict increasing or decreasing pattern. As an illustration of the problem this poses, when the final autoproportion value of AP-3 had its sign switched to negative (i.e., increasing autoproportion values, then a sharp decrease), the overall magnitude of bias remained consistent with the original analysis, but the direction of the bias flipped.

The impact of overly constraining models was also examined when external variables related to the latent

Table 9. Fit information for unconstrained models.

Autoproportion Set	χ^2	RMSEA	SRMR	CFI	TLI
Free AP					
M	16.10	.007	.023	1.00	1.00
SD	5.88	.01	.008	.001	.001
% Adequate	99%	100%	100%	100%	100%
% Excellent	94%	100%	100%	100%	100%
Free AP, BC					
M	101.50	.08	.04	.99	.98
SD	31.13	.02	.01	.01	.01
% Adequate	6%	22%	100%	100%	100%
% Excellent	6%	7%	100%	100%	100%
Free AP, BC*					
M	15.87	.01	.02	1.00	1.00
SD	16.37	.02	.01	.001	.001
% Adequate	96%	98%	100%	100%	100%
% Excellent	92%	97%	100%	100%	100%

Note. Free AP = models estimated with freely estimating autoproportion coefficients; Free AP, BC = models estimated with freely estimating autoproportion and basis coefficients; Free AP, BC* = models estimated with freely estimating autoproportion and mostly freely estimating basis coefficients M = Mean; SD = Standard deviation; % Adequate = Percentage of replications that crossed fit thresholds for being deemed adequate (χ^2 $p > .01$, RMSEA $< .08$, SRMR $< .08$, CFI $> .90$, TLI $> .90$); % Excellent = Percentage of replications that crossed fit thresholds for being deemed excellent (χ^2 $p > .05$, RMSEA $< .05$, SRMR $< .05$, CFI $> .95$, TLI $> .95$).

Table 10. Correlations between DCS parameters across 1000 replications.

Parameter	μ_{g0}	σ_{g0}^2	μ_{g1}	σ_{g1}^2	σ_{g0g1}
μ_{g0}					
σ_{g0}^2	.09				
μ_{g1}	− .35	− .09			
σ_{g1}^2	− .20	.02	.49		
σ_{g0g1}	− .26	− .26	.72	.16	
β	.33	.09	− .99	− .49	− .73

Note. Correlations based on parameter estimates from Baseline Model. μ_{g0} = time 1 mean; σ_{g0}^2 = time one variance; μ_{g1} = constant change factor mean; σ_{g0}^2 = constant change factor variance; σ_{g0g1} = covariance between time one and constant change factor; β = autoproportion coefficient.

factors were included, as these paths could help reduce bias, and are often of substantive interest (e.g., To what extent does nonverbal ability predict the constant change effect in the development of verbal ability?). The addition of these covariates did in fact reduce bias, but only slightly, even when the extra variable was strongly related to the factors (e.g., bias reduced from 419% to 391% for the constant change mean in AP-1). Furthermore, the new regression paths themselves were often biased to a non-trivial degree (especially the paths to the slope factor; see Table 6), implying that the role of potential determinants of change might also become distorted. Still, it is notable that bias decreased as the strength of the association between the external variable and the factors increased (e.g., bias of 419%, 410%, 399%, and 391% for no, weak, moderate, and strong predictors for the slope factor mean in AP-1). It is also worth pointing out the interesting contradiction whereby adding a predictor made parameters *less* biased, but fit indices *more* likely to reject the model. Together these findings lend credence to the idea that external variables can help to improve estimates. However, this insight may not be practical, as the bias remained large even with strong predictors, and strong predictor variables are not common in psychological research (Meyer et al., 2001).

Not all parameters were biased in the face of incorrect equality constraints. The initial level factor mean and variance, as well as the residual variance, were accurately captured generally. This is unsurprising given that these parameters are not directly related to the misspecified change processes. However, the estimated models, although biased, also rather accurately captured the average change between adjacent time points, and the overall average trajectory over time. Incorrect models were able to accomplish this in spite of the misspecification and biased parameter estimates.

This fact may partly explain why popular indices of model fit and their commonly used cutoff values generally indicated that the misspecified and biased models fit the data adequately or excellently. Indeed, the examination of several individual replications from the different conditions revealed that models were overall able to reproduce both the observed means, *and* the observed variance/covariance matrices rather accurately (e.g., the average bias in the implied variance-covariance matrix for three replications from AP-1 was 1.55%, 1.89%, and 1.29%). However, across conditions and replications there was a consistent trend such that the observed means were reproduced more accurately than the variances and covariances (e.g., the average bias in the implied means for the same three replications from AP-1 was 0.31%, 0.51%, 0.26%). This highlights that overall these models are adept at accommodating misspecification to reproduce the observed data, however this effectiveness is especially pronounced in the mean structure. Most models fit well or excellent by conventional standards, but to the degree that there was misfit (as most models did evince some misfit), it was more likely to come from the covariance, as opposed to the mean, structure.

To be sure, some leniency in fit statistics can be a virtue, especially as many fit statistics were developed to counter the tendency of the chi-square test to pick up on trivial misfit, especially with larger sample sizes. As such, fit statistics should not necessarily indicate a major issue when there is some misspecification, as misspecification will not always have major consequences. However, parameter bias in most of the misspecified models was severe, which makes the performance of the fit statistics in this context more troubling. Parameter estimates could be 2 to 5 time larger in absolute magnitude than they should have been, yet most models fit the data well or very well by several indices. Notably, the degree of parameter bias was not consistently related to the performance of the fit statistics. For example, AP-2 was associated with the least amount of parameter bias, yet models were more likely to be rejected for this set than were models based on AP-1, which was associated with much more bias. Of course, fit statistics primarily capture the ability of the model to reproduce the observed data, and as highlighted above, most models achieved this aim quite well in spite of their problematic estimates. Reinforcing this notion, even though AP-2 was associated with the least amount of parameter bias, it was associated with some of the most pronounced bias in the reproduced trajectories (see Table 4). Thus, fit statistics will generally be more attuned with models' ability to accurately represent overall change trajectories than their ability to accurately represent the underlying change processes, which necessitates their cautious application when using LCS.

Overall, these results are concerning as they indicate that misspecified models with substantial parameter bias may appear to provide a good fit to the data by conventional standards. Follow-up analyses revealed a very strong correlation between the autoproportion path and slope factor mean that may help explain these issues, as well as the fact that although parameter estimates were biased, the population trajectory was mostly recovered accurately. Given the high degree of correspondence, these key parameters may compensate for one another in reproducing the population trajectory. That is, the most relevant growth parameters appear capable of accommodating misspecification in their counterpart in order to accurately capture the population trajectory. Increasing the number of waves of assessment succeeded in reducing the correlation between parameters, which could address the problems identified here, making this a potentially fruitful avenue for future research.

Although this ostensibly implies that researchers should attempt to freely estimate the autoproportion and basis coefficients, the present findings indicate many potential difficulties with this strategy. When only the parameters that were unequal across time were freed, models on average correctly recovered the parameter values, but estimates were quite variable across replications, making it more difficult to interpret the results from any one analysis. Further, as it is unrealistic to assume that the parameters that are unequal across time are known, and that only one growth process is time-varying, it would likely be most defensible to begin by estimating every parameter without imposing invariance over time. These models tended to recover the parameters accurately without much variability across replications, but they were difficult to estimate, and often produced biased standard errors. Indeed, the majority of fully estimating models often failed to converge, suggesting a substantial degree of instability. Given the rather low rates of convergence, and biased standard errors, even models that successfully estimate all parameters may not be the most trustworthy. As such, convergence by itself cannot be used to indicate a reliable model. Again, this may be an issue stemming from the difficulty of attempting to disentangle two highly correlated change processes without a substantial amount of data.

Interestingly, adding a single extra constraint helped increase the rate of convergence and the accuracy of the standard errors, but even when accurate this single "stabilizing" constraint introduced a substantial amount of bias, and actual parameter instability. All of the more unconstrained models generally fit the data excellently (often near perfectly), which is especially problematic in the latter scenario in which a single, seemingly innocuous, constraint introduces substantial bias. Indeed, it was often possible to go from a nonconverging model to a near perfectly fitting, but very biased, model just by fixing one extra basis coefficient to 1. This occurred even when the basis coefficient was 1 in the population, suggesting that even population-congruent constraints can lead to excellent fitting but biased models when partially constrained DCS are estimated.

Thus, less constrained models may be just as untrustworthy as constrained models, an issue compounded by potential instability in estimation. As such, to the extent they are useful, unconstrained models likely best serve as tools for assessing the plausibility of constraints rather than an end unto themselves. That is, given the issues encountered here, to the extent such models are employed, they should likely be utilized cautiously. Parameters in unconstrained and constrained models can be compared, as can fit statistics. Further, the residuals of models with varying degrees of constraint can be examined to assess how constraints contribute to the reproduction of the mean and covariance structures. Of course, these unconstrained models may fail to converge, or provide unstable estimates, especially if a minimal number of constraints are added to improve rates of convergence. Altogether then, the results based on unconstrained models suggest that evaluating the plausibility of constraints across time presents many challenges, and that when these constraints appear invalid, it may still be difficult to obtain a trustworthy model. In such instances, different analytic approaches may be necessary.

Implications

Latent change score models are conceptually powerful, and offer developmental researchers a flexible tool for investigating change over time without the substantive limitations inherent to autoregressive or growth curve models. The present study however suggests that LCS are not without major pitfalls, and should be applied cautiously in many contexts, particularly when invariance in the change parameters across time is unrealistic. To be sure, the present study only focused on one particular type of LCS, the DCS, but the DCS serves as the underlying model of other more complex LCS, such as the bivariate dual change score model (King et al., 2009).

We briefly considered three potential ways to address the issues we identified with LCS models. Including external predictors of the components of the model, for one, could reduce bias while providing the scientifically meaningful insights that are often of primary interest. However, our findings indicated that the ability of such variables to reduce bias and improve the performance of fit statistics is modest at best, and including them may actually exacerbate the problem as the regression paths themselves may also be considerably biased. Alternatively, unconstraining parameters to assess the feasibility of such constraints can result in accurate parameter estimates. However, these models are more difficult to estimate, and parameter estimates may be quite unstable and the standard errors inaccurate, making these models difficult to interpret and evaluate. Furthermore, even a single constraint to assist estimation may result in substantially biased estimates. Including more waves of data reduced the correlation between parameters in the present study, and may thus address some of the issues identified here; more work is needed to explore this possibility. Our present results suggest it may take many more waves of data to achieve this level of stability than are commonly available in longitudinal studies.

It is worth reiterating that despite the bias in the individual parameter estimates, the misspecified models rather effectively captured the change between time points in the latent change factors, and the overall trajectory across time. This second characteristic is of dubious usefulness if the actual parameters underlying the trajectory cannot be trusted, as then these models provide little about change beyond what can be gained by simply examining the observed means and standard deviations at each time point. That is, if the parameters cannot be trusted, the model becomes descriptive rather than explanatory, and therefore of limited scientific value. The model implied latent change scores were also generally well captured though, and these components of the model are potentially more useful as outcome or predictor variables in advanced investigations of developmental processes. More work is needed to test the extent to which the latent change factors can confidently be used in investigations however.

Finally, the results here reemphasize the point that the commonly applied cutoff values for fit statistics are not universal, and should not be strictly adhered to across analyses. Across replications, fit statistics commonly indicated that substantially biased models fit the data excellently. The fit cutoffs usually used are primarily based on investigations of simple confirmatory factor analytic models (e.g., Hu & Bentler, 1999), and there is evidence that these standards do not apply to all types of models (e.g., Fan & Sivo, 2007). The current results add to this collection of evidence and indicate that the typical rules of thumb for guiding model selection will not always apply to LCS. As such, when estimating LCS researchers should treat fit information cautiously, especially until the functioning of fit indices in the context of latent change score modeling is better understood.

Limitations and future directions

We note several limitations with our study. Notably, the current study included no attrition, and data collected at every necessary time point (LCS models generally require evenly spaced time intervals). These conditions are likely not the conditions faced by most developmental researchers. Future work should more thoroughly examine the role of attrition and unequal time interval spacing (i.e., the inclusion of phantom variables). Considering that problematic trends were observed with ideal data, it is likely that these additional real world concerns will only exacerbate the challenges we identified (e.g., with missing waves it is impossible to unconstrain all parameters, and the results here suggest that even one extra constraint can introduce substantial bias).

Further, although it is a strength that the population model used here was based on real data (i.e., a realistic set of parameters), it may be that the results here do not necessarily generalize to other population models. However, the primary goal here was to illustrate potential issues with this model as a caution, not test all population values and patterns. Indeed, despite the potential limitations on generalizability, the current results are useful for showing that, at the very least, under *some* circumstances (reasonable circumstances as well, again, given the origins of the population model) DCS are likely to suffer from the shortcomings identified here. Given that population values are generally unknown, the limited knowledge provided here is thus still useful for engendering a justifiable caution in developmental researchers.

Future work should thus build on the present findings by examining the functioning of more advanced LCS. To be sure, the DCS may be more popular as a building block for more complex models than as a standalone model. The bivariate dual change score model, for instance, includes two parallel DCS that are synched with a series of cross-lagged coupling parameters (McArdle & Hamagami, 2001). Two simultaneously misspecified DCS could greatly increase the amount of bias present; however, the presence of another variable, paired up with the coupling parameters, may help to stabilize estimates. Furthermore, the coupling parameters themselves are often constrained to equality across time, and the ramifications of these constraints would also need to be investigated. These questions require additional research.

It is also worth noting that the current study focused on absolute model fit, however future work should consider relative/comparative fit. Although fit statistics may struggle to identify misspecification and bias in the absolute sense, they may be more effective in comparing more progressively constrained models. This will require improving the feasibility and integrity of more unconstrained models, however.

In this vein, methods that may make unconstrained models more computationally tractable and trustworthy must be considered. One possibility is using Bayesian SEM estimation techniques instead of the more typical maximum likelihood approach (Kaplan & Depaoli, 2012). The use of informative priors, for example, could ease the computational burden of more unconstrained models. To date however, there is little work on Bayesian estimation and LCS. More reliably estimated unconstrained models would in part make it more feasible to evaluate the appropriateness of constraints across time.

Conclusion

Latent change score models represent a flexible, modern approach to rigorously analyzing change and development of psychological constructs over time. These models are potentially powerful tools, combining the conceptual strengths of both autoregressive and growth curve models, that can provide numerous insights about developmental processes beyond what can typically be gained from more basic models. In this study, we found that imposing the parameter invariance over time that is typically introduced in these models can lead to a distorted picture of the underlying developmental processes. Furthermore, we found that model fit statistics do not generally indicate that anything has gone awry in the modeling process. Precautions can be taken in an attempt to avoid these pitfalls (e.g., including predictors, freeing parameters), but such safeguards provide limited protection, and may backfire under several circumstances.

Article information

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

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

Funding: This work was supported by Grants *R305A110293 and R324A15006* from the *Institute of Education Sciences, U.S. Department of Education*.

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

Acknowledgments: The authors would like to thank the reviewers for their comments on prior versions of this manuscript. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions or the Institute of Education Sciences is not intended and should not be inferred.

References

Browne, M. W., & Cudeck (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: SAGE Punblications, Inc.

Corker, K. S., Donnellan, M. B., & Bowles, R. P. (2013). The development of achievement goals throughout college: Modeling stability and change. *Personality and Social Psychology Bulletin*, *39*(11), 1407–1417. doi:10.1177/0146167213494243

Curby, T. W., Grimm, K. J., & Pianta, R. C. (2010). Stability and change in early childhood classroom interactions during the first two hours of a day. *Early Childhood Research Quarterly*, *25*, 373–384. doi:10.1016/j.ecresq.2010.02.004

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. doi:10.1080/00273170701382864

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, *49*(1), 92–107. doi: 10.2307/1937887

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science*, *19*(3), 149–154. doi:10.1177/0963721410370300

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. doi: 10.1037/0012-1649.43.6.1460

Finkel, D., Reynolds, C. A., McArdle, J. J., Hamagami, F., & Pedersen, N. L. (2009). Genetic variance in processing speed drives variation in aging of spatial and memory abilities. *Developmental psychology*, *45*(3), 820–834. doi: 820.10.1037/a0015332

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. doi: 10.1037/0882-7974.18.4.696

Grimm, K. J. (2012). Intercept centering and time coding in latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*, 137–151. doi:10.1080/10705511.2012.634734

Grimm, K. J., An, Y., McArle, J. J., Zonderman, A. B., & Resnick, S. M. (2012). Recent changes leading to subsequent changes: Extensions of multivariate latent difference score models. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(2), 268–292. doi:10.1080/10705511.2012.659627

Grimm, K. J., Castro-Schilo, L., & Davoudzadeh, P. (2013). Modeling intraindividual change in nonlinear growth models with latent change scores. *The Journal of Gerontopsychology and Geriatric Psychiatry*, *26*(3), 153–162. doi:10.1024/1662-9647/a000093

Grimm, K. J., Mazza, G. L., & Mazzocco, M. M. (2016). Advances in methods for assessing longitudinal change. *Educational Psychologist*, *51*(3–4), 342–353. doi: 10.1080/00461520.2016.1208569

Hallquist, M., & Wiley, J. (2014). *MplusAutomation: Automating mplus model estimation and interpretation*. R package version 0.6–3. Retrieved from http://CRAN.R-project.org/package=MplusAutomation

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A multidisciplinary Journal*, *6*(1), 1–55. doi: 10.1080/10705519909540118

Kaplan, D., & Depaoli, S. (2012). Bayesian structural equation modeling. In *Handbook of structural equation modeling* (pp. 650–673). New York, NY: The Guilford Press.

King, D. W., King, L. A., McArdle, J. J., Shalev, A. Y., & Doron-LaMarca, S. (2009). Sequential temporal dependencies between symptoms of depression and posttraumatic stress disorder: An application of bivariate latent difference score structural equation modeling. *Multivariate Behavioral Research*, *44*(4), 437–464. doi: 10.1080/00273170903103308

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. doi: 10.1002/jts.20188

Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). Panel data and models of change: A comparison of first difference and conventional two-wave models. *Social Science Research*, *14*, 80–101.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226. doi: 10.1146/annurev.psych.51.1.201

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*, *11*(3), 401–423. doi: 10.1207/s15328007sem1103_6.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In S. DuToit, R. Cudeck, & D. Sorboom (Eds.), *Structural equation modeling: Present and future* (pp. 341–380). Lincolnwood, IL: Scientific Software International, Inc.

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*, 577–605. doi: 10.1146/annurev.psych.60.110707.163612

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* (pp. 245–274). New York, NY: Springer Science and Business Media.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural model 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). Washington, DC: American Psychological Association.

McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models*. Washington, DC: American Psychological Association.

McArdle, J. J., Petway, K. T., & Hishinuma, E. S. (2015). IRT for growth. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling* (pp. 435–456).New York, NY: Routledge.

McArdle, J. J., & Prindle, J. J. (2008). A latent change score analysis of a randomized clinical trial in reasoning training. *Psychology of Aging*, *23*(4), 702–719. doi: 10.1037/a0014349

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*(2), 128–165. doi: 10.1037//0003-066X.56.2.128

Muthen, L. K., & Muthen, B. O. (1998-2015). *Mplus user's guide*. (Seventh Edition). Los Angeles, CA: Muthen & Muthen.

R Core Team. (2016). *R: A language and environment for statistical computing*. (3.3.1) [computer program]. Vienna, Austria: R Foundation for Statistical Computing.

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. doi: 10.1080/10705511.2015.1066680

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In *Handbook of structural equation modeling* (pp. 209–231). New York, NY: The Guilford Press.

Wu, W., Selig, J. P., & Little, T. D. (2013). Longitudinal data analysis. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (pp. 387–410). New York, NY: Oxford University Press.