
Multivariate Behavioral Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hmbr20

# Regime-Switching Bivariate Dual Change Score Model

Sy-Miin Chow [a], Kevin J. Grimm [b], Guillaume Filteau [c], Conor V. Dolan [d] & John J. McArdle [e]

[a] The Pennsylvania State University
[b] University of California, Davis
[c] University of North Carolina at Chapel Hill
[d] University of Amsterdam
[e] University of Southern California

Published online: 26 Jul 2013.

To cite this article: Sy-Miin Chow, Kevin J. Grimm, Guillaume Filteau, Conor V. Dolan & John J. McArdle (2013) Regime-Switching Bivariate Dual Change Score Model, Multivariate Behavioral Research, 48:4, 463-502, DOI: 10.1080/00273171.2013.787870

To link to this article: http://dx.doi.org/10.1080/00273171.2013.787870

*Multivariate Behavioral Research*, 48:463–502, 2013

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273171.2013.787870

Routledge
Taylor & Francis Group



# Regime-Switching Bivariate Dual Change Score Model

Sy-Miin Chow
*The Pennsylvania State University*

Kevin J. Grimm
*University of California, Davis*

Guillaume Filteau
*University of North Carolina at Chapel Hill*

Conor V. Dolan
*University of Amsterdam*

John J. McArdle
*University of Southern California*

Mixture structural equation model with regime switching (MSEM-RS) provides one possible way of representing over-time heterogeneities in dynamic processes by allowing a system to manifest qualitatively or quantitatively distinct change processes conditional on the latent "regime" the system is in at a particular time point. Unlike standard mixture structural equation models such as growth mixture models, MSEM-RS allows individuals to transition between latent classes over time. This class of models, often referred to as regime-switching models in the time series and econometric applications, can be specified as regime-switching mixture structural equation models when the number of repeated measures involved is not large. We illustrate the empirical utility of such models using one special case—a regime-switching bivariate dual change score model in which two growth

Correspondence concerning this article should be addressed to Sy-Miin Chow, 118 Henderson Building, The Pennsylvania State University, University Park, PA 16801. E-mail: symiin@psu.edu

processes are allowed to manifest regime-dependent coupling relations with one another. The proposed model is illustrated using a set of longitudinal reading and arithmetic performance data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 study (ECLS-K; U.S. Department of Education, National Center for Education Statistics, 2010).



Stepwise or stagewise theories conceptualize change processes as discontinuous progressions through distinct, "categorical" phases. One of the earliest and perhaps most prominent theoretical examples of such processes in psychology is Piaget's 1969 theory of human cognitive development and related extensions (Piaget & Inhelder, 1969; Dolan, Jansen, & van der Maas, 2004; Hosenfeld, 1997; van der Maas & Molenaar, 1992). Other examples of such unidirectional stagewise change processes include Kohlberg's (Kohlberg & Kramer, 1969) hypothesis that moral reasoning develops through stages of increasing moral sophistication, Fukuda and Ishihara's 1997 work on the emergence of human sleep and wakefulness rhythm during the first 6 months of life, and Van Dijk and Van Geert's 2007 study on discrete shifts in early language development.

Longitudinal regime-switching models provide one possible way of testing stagewise theories of change. Within- and between-person heterogeneities in change are represented in such models by allowing individuals to be characterized by different longitudinal models contingent on the latent class or "regime" they are in at a particular time point (Hamilton, 1989; Kim & Nelson, 1999). In addition, individuals are allowed to transition between different latent classes or "regimes" over time. These models thus provide a methodological framework for testing and extending theories concerning how individuals may manifest quantitatively and/or qualitatively distinct dynamics during different stages of a change process. One distinct utility of these models is that they provide not only ways of representing the distinct change trajectories associated with each stage but also a way to test whether the progression through stages unfolds sequentially (e.g., from earlier stages to later stages in an irreversible manner) or in other constrained forms. For instance, in the time series and econometric literature, researchers have proposed regime-dependent autoregressive processes wherein the transition between regimes is modeled as a first-order Markov-switching process that allows an individual to transition freely between any of the hypothesized regimes (e.g., Hamilton, 1989). Other alternatives include models which posit that the switching between regimes is governed by deterministic thresholds (e.g., as in threshold autoregressive models; Tong & Lim, 1980) or past values of a system (e.g., as in self-exciting threshold autoregressive models; Tiao & Tsay, 1994). Thus, such models enrich conventional ways of conceptualizing stagewise processes by offering more ways to represent changes *within* as well as *between* stages.

In cases involving longitudinal panel data, regime-switching models can be implemented using existing structural equation modeling software such as Mplus (Asparouhov & Muthén, 2011; B. O. Muthén & Asparouhov, 2011; Nylund-Gibson, Muthén, Nishina, Bellmore, & Graham, under review) and Open-Mx (Boker et al., 2011). In particular, it has been well established that the structural equation modeling framework is not well suited for handling the type of intensive repeated measures data typically used in applications of regime-switching models in the time series and econometric literature—namely, data with a much larger number of time points than participants (Chow, Ho, Hamaker, & Dolan, 2010; Hamaker, Dolan, & Molenaar, 2003). However, as we elucidate and illustrate in the present article, formulating the longitudinal processes within regimes as structural equation models (SEMs) has some computational advantages in cases involving longitudinal panel data.

SEMs with regime–switching properties can regarded as a longitudinal extension of mixture structural equation model (MSEM; Dolan, 2009; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997b; L. K. Muthén & Muthén, 2001; B. O. Muthén, 2002). We refer to them herein as mixture structural equation models with regime switching (MSEMs-RS). In standard MSEMs, individuals may be characterized by a distinct measurement and/or structural model contingent upon the latent classes to which they belong. As discussed elsewhere (Lubke, 2005; Tueller & Lubke, 2010), MSEMs subsume factor mixture models such as growth mixture models (Nagin & Land, 1993) and other related variations (e.g., Dolan & van der Maas, 1998; Lubke, 2005; B. O. Muthén & Shedden, 1999; Yung, 1997) by allowing the regime dependency to extend to structural (specifically, regression) relations among the latent variables. MSEM-RS further expands MSEM by allowing individuals to transition between latent classes over time.

Pertinent applications of longitudinal MSEM-RS include the work of Dolan, Schmittmann, Lubke, and Neale (2005), who presented a regime-switching latent growth curve model and allowed individuals to switch between different growth curve processes. Schmittmann, Dolan, van der Maas, and Neale (2005) illustrated the added flexibility attained by representing the change processes within regimes by means of covariance structure models (or more broadly, SEMs). Kaplan (2008) demonstrated the use of Mplus to fit models with Markov switching properties. Among the different variations considered was the well-known mover-stayer model.

In this article, we present a regime-switching bivariate dual change score model to capture two learning processes that fluctuate between a coupled and a decoupled phase. Using a set of longitudinal reading and arithmetic performance data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) study (ECLS-K; U.S. Department of Education, National Center for Education Statistics, 2010), we seek to evaluate the emergence and disintegration of coupling relations between children's reading and arithmetic learning processes from kindergarten through middle school. The proposed model is formulated as an MSEM-RS. In addition to illustrating how SEM software such as Mplus can be used to implement the proposed regime-switching model, we also present a simulation study to examine how sample size, regime stability, and differences in regime dynamics may affect estimation results associated with the proposed model.

## MIXTURE STRUCTURAL EQUATION MODEL WITH REGIME SWITCHING (MSEM-RS)

Structural equation modeling (Goldberger & Duncan, 1973) is a statistical technique for representing multivariate relationships among observed and latent variables. An SEM is composed of a measurement model that relates the observed indicators to a set of underlying latent variables and a structural model that specifies the interrelationships among the latent variables. One assumption of traditional SEM is that every individual in the sample conforms to the same statistical model. However, it is likely that in some cases, the parameters of these models may show systematic heterogeneities across subpopulations. When such heterogeneities can be explained completely by means of known covariates, a multiple-group analysis can be performed to allow the modeling parameters to vary over groups. However, when the grouping variable that distinguishes the subpopulations is unobserved (i.e., latent), multiple-group analysis cannot be performed in a straightforward manner. In this case, MSEM (Arminger & Muthén, 1998; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997b; Vermunt & Magidson, 2005; Yung, 1997) may be employed.

MSEM is an extension of SEM that allows for heterogeneities in the means and covariance structures of an SEM model conditional on a series of latent, unobserved groups, typically referred to as latent *classes* and denoted herein as $S_i$ for individual $i$ $(i = 1, \ldots, n)$. The measurement model for an individual $i$ in class $k$, denoted here as $\mathbf{y}_{i|S_i=k}$, can be expressed as

$$\mathbf{y}_{i|S_i=k} = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \qquad \boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_k), \tag{1}$$

where $\mathbf{y}_i$ is a $p \times 1$ vector of observed variables for individual $i$ who is in class $k$, $\boldsymbol{\nu}_k$ is a $p \times 1$ vector of intercepts, $\boldsymbol{\Lambda}_k$ is a $p \times w$ matrix of factor loadings, $\boldsymbol{\eta}_i$ is a $w \times 1$ vector of latent variables, and $\boldsymbol{\epsilon}_i$ is a $p \times 1$ vector of measurement errors. The subscript $k$ highlights the class-specific nature of the parameters. It is possible to employ mixed (including both continuous and categorical) manifest responses in MSEM, but we focus only on continuous indicator variables in this article.

Conditional on class $k$, a structural model serves to relate the latent variables to one another, with

$$\boldsymbol{\eta}_{i|S_i=k} = \boldsymbol{\alpha}_k + \boldsymbol{\beta}_k \boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \qquad \boldsymbol{\zeta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_k), \tag{2}$$

where $\boldsymbol{\alpha}_k$ is a $w \times 1$ vector of regime-dependent intercepts, $\boldsymbol{\beta}_k$ is a $w \times w$ matrix of regime-dependent regression effects among the latent variables, and $\boldsymbol{\zeta}_i$ is a $w \times 1$ vector of disturbances.

The model-implied means and covariance structures of the observed data can be, respectively, obtained as

$$\begin{aligned} E(\mathbf{y}_i|S_i = k) &\triangleq \boldsymbol{\mu}_k = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k(\mathbf{I} - \boldsymbol{\beta}_k)^{-1}[\boldsymbol{\alpha}_k] \quad \text{and} \\ Cov(\mathbf{y}_i|S_i = k) &\triangleq \boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k(\mathbf{I} - \boldsymbol{\beta}_k)^{-1}\boldsymbol{\Psi}_k(\mathbf{I} - \boldsymbol{\beta}_k')^{-1}\boldsymbol{\Lambda}_k' + \boldsymbol{\Omega}_k. \end{aligned} \tag{3}$$

Defining $\boldsymbol{\theta}$ as the vector of all unknown parameters, the conditional likelihood for individual $i$ in latent class $k$ is

$$f(\mathbf{y}_i|S_i = k, \boldsymbol{\theta}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}_k|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_k)\right], \tag{4}$$

and the overall (marginal) likelihood can then be obtained by summing across classes the conditional likelihood function for each class weighted by the probability of the class. Correspondence between MSEM and other finite mixture models such as latent class analysis and latent profile analysis has been described elsewhere (Bartholomew & Knott, 1999; Dolan et al., 2004; Gibson, 1959; Lazarsfeld & Henry, 1968; B. O. Muthén, 2002).

More recently, SEM statistical packages, such as Mplus, allow users to fit longitudinal extensions of MSEMs and estimate how individuals transition among different latent classes over time (Kaplan, 2008; B. O. Muthén & Asparouhov, 2011; Nylund-Gibson et al., under review). In the resultant MSEM-RS framework, the initial class (or regime) probabilities can be represented using a multinomial regression model as

$$\Pr(S_{i1} = k|\mathbf{x}_{i1}, \boldsymbol{\theta}) \triangleq \pi_{k,i1} = \frac{\exp(a_{k1} + \mathbf{b}_{k1}'\mathbf{x}_{i1})}{\sum_{s_1=1}^{K_1} \exp(a_{s1} + \mathbf{b}_{s1}'\mathbf{x}_{i1})}, \tag{5}$$

where $S_{it}$ represents individual $i$'s class membership at time $t$ and $\mathbf{S}_i = [S_{i1}, \ldots, S_{iT}]'$ is now a class membership vector indicating person $i$'s entire class history from Time 1 to $T$, $K_1$ denotes the number of classes at Time 1, $a_{k1}$ is the logit intercept at Time 1, and $\mathbf{x}_{i1}$ is a vector of fixed covariates for person $i$ at Time 1 with corresponding logit slopes in $\mathbf{b}_{k1}$ for predicting the initial class probabilities at $t = 1$. Typically, $a_{K1}$ is set to 0 so the last class is used as the reference class.

A multinomial logistic regression model is then used to describe each individual $i$'s transition in class membership from time $t$-1 to time $t$ as

$$\Pr(S_{it} = k | S_{i,t-1} = j, \mathbf{x}_{it}, \boldsymbol{\theta}) \triangleq \pi_{jk,it} = \frac{\exp(a_{kt} + \mathbf{b}'_{kt}\mathbf{x}_{it})}{\sum_{st=1}^{K_t} \exp(a_{st} + \mathbf{b}'_{st}\mathbf{x}_{it})}, \quad (6)$$

where $K_t$ denotes the number of classes at time $t$, highlighting the possibility that latent classes may emerge or diminish over time. $\pi_{jk,it}$ is individual $i$'s transition probability of moving from class $j$ at time $t$-1 to class $k$ at time $t$; $a_{kt}$ is the logit intercept for the $k$th class at time $t$; and $\mathbf{x}_{it}$ is a vector of fixed covariates for predicting the transition probabilities with an associated vector of logit slopes, $\mathbf{b}_{kt}$. Included in $\mathbf{x}_{it}$ are $K_t - 1$ binary constants reflecting the deviation in log odds of switching into latent class $k$ at time $t$ from latent class $j$ (i.e., $S_{i,t-1} = j$, $S_{it} = k$) as compared with switching into $S_{it} = k$ from the reference class, that is, $S_{i,t-1} = K_{t-1}$. Note that no binary constants are included in $\mathbf{x}_{i1}$ in Equation 5 because at Time 1, there are no transition probabilities to be predicted, only class probabilities.

For identification purposes, $a_{K_t} = 0$ and $\mathbf{b}_{K_t} = \mathbf{0}$ for each $t$, which has the effect of setting the logits of the last latent class, defined as $\text{logit}(\pi_{jK_t,it}) = \log(\pi_{jK_t,it}/\pi_{jK_t,it})$, $j = 1, \ldots, K_{t-1}$), to 0. This also has the effect of setting each row of person $i$'s transition probability matrix at time $t$, defined as

$$\boldsymbol{\pi}_{it} = \begin{bmatrix} \pi_{11,it} & \pi_{12,it} & \cdots & \pi_{1K_t,it} \\ \pi_{21,it} & \pi_{22,it} & \cdots & \pi_{2K_t,it} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{K_{t-1}1,it} & \pi_{K_{t-1}2,it} & \cdots & \pi_{K_{t-1}K_t,it}, \end{bmatrix}, \quad (7)$$

to sum to 1. In other words, given that an individual was in a particular latent class at time $t$-1, he or she would have to transition into one of the $K_t$ possible latent classes at time $t$. The transition probability matrix may potentially vary over persons and time points because of the use of person-specific and time-varying covariates in $\mathbf{x}_{it}$ to predict the transition probabilities.

Depending on a researcher's model specification, the transition between two regimes can be unidirectional or bidirectional in nature. One popular class of constraints imposed in the context of many stagewise theories dictates that the transition between regimes unfolds in a unidirectional manner. That is, the transition has to occur from one regime to a later regime in a sequential manner and transition in the reverse direction is rare or not allowed. In the case of a two-regime model, a sequential transition pattern can be implemented by freeing $\pi_{12,it}$ and setting $\pi_{21,it}$ to zero. In cases where researchers expect that an individual can transition from Regime 1 to Regime 2 and vice versa, this can be implemented by freeing both $\pi_{12,it}$ and $\pi_{21,it}$. For instance, in the study of cognitive development, a child may revert to an earlier cognitive stage when asked to solve less familiar problems—a phenomenon known as "horizontal decalage" in Piagetian terminology (Mercer, 1998). In the structural equation modeling framework, it is also plausible to allow later class membership to condition on the entire past history of class membership, namely, $\mathbf{S}_i^{t-1} = \{S_{i1}, S_{i2}, \ldots, S_{i,t-1}\}$ (as implemented, e.g., in Mplus; Asparouhov & Muthén, 2011). In the present context, we only allow an individual's class membership at time $t$ to depend on his or her class membership at the previous time point and other covariates of interest.

Maximum likelihood (ML) estimates of the parameters in the MSEM-RS can be obtained via the Expectation-Maximization (EM; Dempster, Laird, & Rubin, 1977; Everitt & Hand, 1981; Titterington, Smith, & Makov, 1985) algorithm. The estimation details have been documented elsewhere (see, e.g., Asparouhov & Muthén, 2011; B. O. Muthén & Shedden, 1999), but they are included here for didactic purposes.

We define our notations as follows. We denote $\mathbf{S}_i = [\mathbf{S}_{i1}, \ldots, \mathbf{S}_{iT}]'$ as person $i$'s vector of class history and it can take on a total of $h = 1, \ldots\ K$ possible patterns of values, with $\mathbf{S}_h$ representing the $h$th possible configuration of class membership values from time 1 to time $T$.[1] $\mathbf{x}_i = \{\mathbf{x}_{i1}, \ldots \mathbf{x}_{iT}\}$ is person $i$'s array of time-varying covariates for all time points and the complete data set is $\mathbf{Z} = \{\mathbf{Y}, \mathbf{S}\} = \{\mathbf{y}_{it}, S_{it}; i = 1, \ldots n; t = 1, \ldots T\}$. The complete data log likelihood function can be written as

$$\log[L(\boldsymbol{\theta}|Z)] = \sum_{i=1}^{n}\sum_{h=1}^{K} I_{(\mathbf{S}_i=\mathbf{S}_h)}\left[\log f(\mathbf{y}_i|\mathbf{S}_i = \mathbf{S}_h, \boldsymbol{\theta}) + \log \Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{x}_i, \boldsymbol{\theta})\right],$$

During the E step of the EM algorithm, expectation of the complete data log likelihood function is taken with respect to $\mathbf{S}_i$ conditional on the current parameter estimates $\boldsymbol{\theta}^q$ and observed data as

$$\begin{aligned} E\{\log[L(\boldsymbol{\theta}|\mathbf{Z})]|\mathbf{Y}, \mathbf{x}, \boldsymbol{\theta}^{\mathbf{q}}\} = \sum_{i=1}^{n}\sum_{h=1}^{K} &\Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}^q)\left[\log f(\mathbf{y}_i|\mathbf{S}_i = \mathbf{S}_h, \boldsymbol{\theta})\right. \\ &\left. + \ \log \Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{x}_i, \boldsymbol{\theta})\right], \end{aligned} \quad (8)$$

[1]For instance, with two classes at each time point and three time points, there is a total of $2^3 = 8$ possible ways in which individual's $i$ class membership may vary over time. Thus, $h$ indexes one specific pattern of class history out of eight possible patterns.

where an analog to $\log f(\mathbf{y}_i|\mathbf{S}_i = \mathbf{S}_h, \boldsymbol{\theta})$ was defined in Equation 4 for MSEM with only a single class indicator as opposed to the present case with $T$ indicators to indicate class membership for all time points. $\Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{x}_i, \boldsymbol{\theta})$ can be derived using Equations 5 and 6; $\Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\theta}^q)$ is the posterior probability of individual $i$'s entire class membership history given all observed data for person $i$ and the current parameter estimates in $\boldsymbol{\theta}^q$. It is computed by means of the Bayes theorem as

$$\Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{y}_i, \mathbf{x}_{it}, \boldsymbol{\theta}^q) = \frac{f[\mathbf{y}_i|\mathbf{S}_i = \mathbf{S}_h, \mathbf{x}_i, \boldsymbol{\theta}^q]\Pr(\mathbf{S}_i = \mathbf{S}_h|\mathbf{x}_i, \boldsymbol{\theta}^q)}{\sum_{l=1}^{K} f[\mathbf{y}_i|\mathbf{S}_i = \mathbf{S}_l, \boldsymbol{\theta}^q]\Pr(\mathbf{S}_i = \mathbf{S}_l|\mathbf{x}_i, \boldsymbol{\theta}^q)}. \tag{9}$$

Thus, the E step is performed to compute the expected complete log likelihood function in Equation 8. In the Maximization step, updated parameter estimates are obtained by maximizing the expected complete log likelihood function in Equation 8 with respect to the parameters. The E and M steps are then iterated sequentially until some predefined convergence criteria have been met.

## REGIME-SWITCHING BIVARIATE DUAL CHANGE SCORE MODEL

The Regime-Switching Bivariate Dual Change Score Model proposed in this article is an extension of the bivariate dual change score (BDCS) model presented by McArdle and Hamagami (2001) as well as the latent class extension considered by McArdle and Grimm (2010), which allowed for between-class differences in modeling parameters but not within-person transition between regimes. According to the BDCS model, over–time fluctuations in the scores of two observed processes, denoted here as $y_{1,it}$ and $y_{2,it}$, are driven by measurement error as well as changes that occur at the true score level, namely, at the level of two underlying latent variables, $\eta_{1,it}$ and $\eta_{2,it}$. For instance, in our illustrative example, the two manifest processes are students' scores on an arithmetic task and a verbal performance task. The latent variables (i.e., $\eta_{1,it}$ and $\eta_{2,it}$.), in contrast, correspond to the students' true arithmetic and verbal scores at the latent level.

The observed variables, $y_{1,it}$ and $y_{2,it}$, are used to identify their corresponding latent true scores, $\eta_{1,it}$ and $\eta_{2,it}$, through a measurement model hypothesized to be regime-invariant. Let $\mathbf{y}_{it} = [y_{1,it} \;\; y_{2,it}]'$ and $\boldsymbol{\eta}_{it} = [\eta_{1,it} \;\; \eta_{2,it}]'$. This measurement model can be expressed as

$$\mathbf{y}_{it} = \boldsymbol{\eta}_{it} + \boldsymbol{\epsilon}_{it}, \qquad \boldsymbol{\epsilon}_{it} \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}[\omega_{11}, \omega_{22}]),$$

where $\boldsymbol{\epsilon}_{it}$ is a $2 \times 1$ vector of measurement errors for the two observed variables.

At $t = 1$, interindividual differences in initial performances for the $k$th class are modeled as

$$\begin{bmatrix} \eta_{1,i1} \\ \eta_{2,i1} \\ \alpha_{1,i} \\ \alpha_{2,i} \end{bmatrix} = \begin{bmatrix} \mu_{\eta_{1k}} \\ \mu_{\eta_{2k}} \\ \mu_{\alpha_{1k}} \\ \mu_{\alpha_{2k}} \end{bmatrix} + \begin{bmatrix} \zeta_{\eta_1,i} \\ \zeta_{\eta_2,i} \\ \zeta_{\alpha_1,i} \\ \zeta_{\alpha_2,i} \end{bmatrix} \tag{10}$$

$$\begin{bmatrix} \zeta_{\eta_1,i} \\ \zeta_{\eta_2,i} \\ \zeta_{\alpha_1,i} \\ \zeta_{\alpha_2,i} \end{bmatrix} \sim \mathrm{N}\left( \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \psi_{11} & & & \\ \psi_{21} & \psi_{22} & & \\ \psi_{31} & \psi_{32} & \psi_{33} & \\ \psi_{41} & \psi_{42} & \psi_{43} & \psi_{44} \end{bmatrix} \right),$$

where $\eta_{1,i1}$ and $\eta_{2,i1}$ are the initial levels of the two latent processes of interest at time 1. These initial levels are a function of the regime-specific group average initial levels, $\mu_{\eta_{1k}}$ and $\mu_{\eta_{2k}}$, and person-specific deviations, $\zeta_{\eta_1,i}$ and $\zeta_{\eta_2,i}$. That is, at $t = 1$, the different hypothesized classes are assumed to have different average initial levels on the two processes of interest, while each individual within each regime also shows some person-specific deviations in initial levels. $\alpha_{1,i}$ and $\alpha_{2,i}$ are person-specific constant change terms that govern the amounts of linear changes manifested by the two processes of interest. These constant change terms are functions of the regime-dependent group average slopes, $\mu_{\alpha_{1k}}$ and $\mu_{\alpha_{2k}}$, and person-specific deviation terms, $\zeta_{\alpha_1,i}$ and $\zeta_{\alpha_2,i}$. Thus, the different regimes are also allowed to show different average levels of constant change on the two processes of interest.

In the absence of further structural relationships at the latent level, the model-implied change trajectories specified thus far are identical to those seen in a bivariate linear growth mixture model wherein different latent classes are stipulated to show different average initial levels and average linear slopes over time. However, the constraints described here render the proposed model distinct from the linear growth mixture model. Beginning at time $t = 2$, (for $t = 2, \ldots, T$), the true scores for person $i$ at time $t$ are a function of the person's true scores at time $t$-1 plus latent changes as

$$\boldsymbol{\eta}_{it} = \boldsymbol{\eta}_{i,t-1} + \Delta\boldsymbol{\eta}_{it},$$

where $\Delta\boldsymbol{\eta}_{it}$ is a $2 \times 1$ vector of latent change in $\boldsymbol{\eta}$ between time $t-1$ and time $t$. The latent change vector, $\Delta\boldsymbol{\eta}_{it}$, is assumed to conform to a regime-specific change process, expressed for regime $S_{it} = k$ as

$$\Delta\boldsymbol{\eta}_{it} = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_k \boldsymbol{\eta}_{i,t-1},$$

where $\boldsymbol{\alpha}_i = [\alpha_{1,i}\ \alpha_{2,i}]'$, with vector elements as defined in Equation 11 and

$$\boldsymbol{\beta}_k = \begin{bmatrix} \beta_{11_k} & \beta_{12_k} \\ \beta_{21_k} & \beta_{22_k} \end{bmatrix}.$$

$\beta_{11_k}$ and $\beta_{22_k}$ are the regime-specific autoproportion (or self-feedback) effect of $\eta_{1,i,t-1}$ on $\eta_{1,it}$ and of $\eta_{2,i,t-1}$ on $\eta_{2,it}$, respectively. These parameters are related to the autoregression parameters in the time series literature.[2] If the self-feedback parameters are negative, these parameters, together with the person-specific constant (i.e., linear) slopes in $\boldsymbol{\alpha}_i = [\alpha_{1,i}\ \alpha_{2,i}]'$, govern the asymptotes of the growth processes. When the constant change term is positive, a larger negative value of the self-feedback parameter signals lesser "growth" when previous levels of the process were high and consequently, a lower asymptote. Furthermore, $\beta_{21_k}$ and $\beta_{12_k}$ are the coupling, or in time series terminology, *cross–regression* parameters reflecting the effect of $\eta_{1,i,t-1}$ on $\eta_{2,it}$ and of $\eta_{2,i,t-1}$ on $\eta_{1,it}$, respectively. In this article, we imposed theory-informed constraints on the number of regimes expected in the proposed regime-switching BDCS model as well as other between-regime differences. These constraints are described in the context of our empirical illustrative example.

## EMPIRICAL ILLUSTRATION

### Data Descriptions

The Early Childhood Longitudinal Study, Class of 1998–99 (ECLS-K; U.S. Department of Education, National Center for Education Statistics, 2010) is a longitudinal study on children's early school experiences beginning with kindergarten and following children through middle school. The sample included data from more than 21,000 children. For our empirical illustration, we used a subsample of the data set ($n = 2{,}369$) with longitudinal assessments of the children's reading and arithmetic skills across seven waves: fall and spring of kindergarten and first grade and the spring of third, fifth, and eighth grades. The key dependent variables used for model-fitting purposes were adaptively scaled estimates of the children's latent reading and arithmetic abilities estimated using item response theory (IRT) model. The participants' IRT-scaled ability estimates are plotted in Figure 1a–1b.

The ECLS-K data have been previously analyzed using nonlinear growth curve models such as the Gompertz curve (Grimm & Ram, 2011; Grimm & Ram, 2010) and other variations of mixture latent Markov models (Kaplan,

[2]Specifically, one plus the self-feedback parameter for a process results in the lag-1 autoregression coefficient of the process on itself. For details, see the formulation detailed in Chow, Ho, Hamaker, and Dolan (2010).

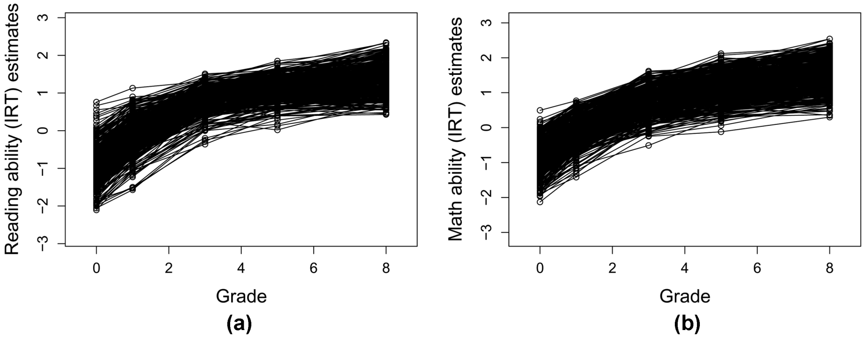


FIGURE 1 A plot of the participants' item response theory (IRT)-scaled ability estimates for (a) a reading task and (b) an arithmetic task.



2008). None of these previous applications focused on between-regime differences in the coupling between two academic learning processes and how their coupling relations emerged and diminished over time. Specifically, we used the BDCS as an alternative linear difference equation that is capable of giving rise to Gompertz-like learning curves. We considered two growh processes simultaneously, namely, the children's acquisition of reading and arithmetic skills from kindergarten through eighth grade. Pertinent questions included whether there was evidence for a bidirectional coupling relation (as reflected in the cross-regression terms) between these two learning processes and whether we could identify latent regimes during which these children show distinct coupling relations. Model fitting was performed using Mplus using the default estimation option of ML-EM with robust standard error estimation (L. K. Muthén & Muthén, 2001). Mplus script for fitting one of the proposed two-regime models, sample simulated data, and the corresponding output can be found on the first author's website at www.personal.psu.edu/quc16/Sy-Miin%27s%20website/soft.htm. The Akaiki information criterion (AIC), Bayesian information criteria (BIC), and sample-size adjusted BIC were used for model selection purposes.

Researchers have long speculated on possible shifts in coupling relations among multiple cognitive constructs throughout the life span. The investment hypothesis (Cattell, 1963, 1971; Horn & Cattell, 1966), for instance, postulates that during school years, there is a pronounced influence of fluid intelligence on crystallized intelligence that enables children to use the former to acquire knowledge/skills that later crytallize into domain-specific skill sets (e.g., quantitative skills, academic knowledge, perceptual-motor skills). Other researchers (Ferrer & McArdle, 2004; Ferrer & Widaman, 2008) further evaluated whether and how the coupling relationships among fluid intelligence, crystallized intelligence, and academic performance showed developmental discontinuities. Ferrer and

McArdle (2004) observed that the influence of fluid reasoning (cognition) on achievement was the strongest from ages 5 to 10, less strong from ages 11 to 15, and weaker from ages 16 to 24. Then, using a different longitudinal data set, Ferrer and Widaman (2008) reported a similar finding: that the coupling from cognition to reading was stronger in childhood than in adolescence.

Much of the attention in early education has focused on early reading and language skills at the expense of other skills, such as mathematics. For example, children spent more than twice as much time on language and literacy skills than math skills in prekindergarten classrooms. On average children spent 17% of the day on language and literacy skills compared with 8% on math-related activities (Early et al., 2010). This discrepancy continues into third grade and through fifth grade. Third-grade students were found to spend approximately 48% of their time on literacy and language arts compared with 24% of their time on mathematics (National Institute of Child Health and Human Development Early Child Care Research Networks, 2005). Fifth-grade students were shown to spend an average of 37% of their time on literacy skills compared with 25% on math-related skills (Pianta, Belsky, Houts, & Morrison, 2007). The focus on early reading skills is primarily based upon the notion that learning to read is a necessary foundation to build other academic skills, such as mathematics (Snow, Burns, & Griffin, 1998), and nonacademic skills, such as self-confidence and social skills (McArdle & Hamagami, 2001). Furthermore, children with poor reading skills are more likely to drop out of school and tend to be limited to low-paying jobs through their lives (U.S. Department of Education, 2003).

Recent research has suggested that early mathematics skills should be considered a key developmental precursor. Duncan et al. (2007) analyzed six large-scale longitudinal national studies to understand associations between early school readiness indicators and later achievement. They found that early mathematics skills were the strongest predictor of later mathematics and reading skills. Even though reading and mathematics achievement are known to be positively related (Aiken, 1971; Monroe & Englehart, 1931), the underlying cause of this positive relationship remains unclear. Wrigley, Saunders, and Neuhaus (1958) proposed that reading and mathematics were not directly related but appeared related because of their relationships with general intelligence. Muscio (1962), on the other hand, believed that mathematics achievement depended on verbal ability in addition to general intelligence. More recently this link has been discussed in multiple early developmental theories (Carey, 2004; Gelman & Butterworth, 2005; Mix, Huttenlocher, & Levine, 2002) where language is proposed to shape the development of number concepts and seen as having a causal influence on at least some aspects of numeracy (Carey, 2004), which is assumed to be a prerequisite of more advanced mathematical achievement. This notion that reading skills are a precursor in the learning of mathematics is compatible with network theories of intelligence (van der Maas et al., 2006) and has gained empirical support, albeit to varying degrees, across studies. Cummins, Kintsch, Reusser, and Weimer (1988) found that correct responses to algebraic word problems were associated with an accurate recall of the problem structure upon completion of the problem. Errors, in contrast, were found to be correct solutions to miscomprehended problems, leading Cummins and colleagues to stress the importance of comprehension when solving word problems. Grimm (2006) fit latent difference score models to reading and mathematics data collected from the Chicago Public Schools and found that reading had a small positive effect on subsequent changes in mathematics skills related to data interpretation and problem solving but was unrelated to subsequent changes in computation. Thus, the nature of this association may depend on the type of mathematics studied. To date, no studies have examined whether and how the association between reading and mathematics skills evolves over time in early school years.

We were interested primarily in identifying regimes with distinct change dynamics by grade level. To do this, the participants' ability estimates in kindergarten and the first grade were averaged across two academic semesters to yield only one measurement occasion (as opposed to two) per school year. Doing so yielded a total of nine time points, four of which were missing. Using the BDCS model with regime switching, we sought to evaluate the existence and nature of the coupling relations between reading and arithmetic performance (e.g., unidirectional as opposed to bidirectional) and how such relations might vary from kindergarten to middle school. In particular, we began by considering two models of increasing complexity: (a) a one-regime BDSC model in which the self–feedback and coupling parameters linking the two learning processes (i.e., reading and arithmetic performance scores) were freed to vary and (b) a two-regime model in which one regime evidenced no coupling between the reading and arithmetic learning processes (denoted as *R1*, the *Decoupled Learning* regime) whereas the second regime showed freely estimated coupling between the reading and arithmetic learning processes (denoted as *R2*, the *Coupled Learning* regime). The average initial levels and constant slopes of the two processes, as well as the self-feedback parameters, were also freed to vary across the two regimes. The participants were allowed to transition between the two regimes freely. That is, no constraint was placed to restrict the individuals' transition between regimes.

If the two-regime model was preferred over the one-regime model, we used several person-specific covariates to predict the participants' probabilities of being in a particular regime at Time 1 in the same model. The covariates included gender, hours spent on nonschool reading, family poverty status (0 = below poverty threshold; 1 = at or above the poverty threshold), and parents' highest education level. The last covariate was contrast coded to yield two categorical predictors for contrasting (a) children whose parents did not complete high school compared with those whose parents completed high school or beyond high school education (this is denoted herein as the *HSvsNoHS* effect) and (b) children whose parents completed high school or some college education but did not obtain a formal college degree compared with those whose parents possessed at least a college degree (this is denoted herein as the *COLLEGE* effect). These covariates were included as regressors in the multinomial regression logistic model shown in Equation 5 to predict individuals' probabilities of being in two latent classes or regimes at $t = 1$.

It is worth noting that from kindergarten to first grade (i.e., from $t = 1$ to 2), the two predefined regimes were only different from each other in terms of their average initial level and average constant slope (see Equation 11). Between-regime differences in reading-arithmetic coupling strengths would start emerging from $t = 2$ and beyond. Thus, the regimes that existed at $t = 1$ may be conceptually different from those observed at other time points. Thus, we also considered a third model wherein we allowed the transition probabilities to change at $t = 2$ and estimated two sets of transition probabilities: (a) the probabilities that governed the transition between regimes from $t = 1$ to $t = 2$ and (b) the probabilities that governed the transition between regimes after $t = 2$. This allowed us to examine the probabilities of children with differing initial levels at $t = 1$ to transition into the *Coupled* and *Decoupled* learning regimes at $t = 2$ (i.e., in first grade) and the relative time length over which individuals stayed in each of these regimes during the remaining study span. In addition, gender was used as a predictor of individual differences in these transition probability patterns. Doing so helped provide insights into whether there were gender differences in the process through which students acquire arithmetic and reading skills from kindergarten to middle school.

The two-regime model served to extend some of the conjectures gleaned from Ferrer and colleagues' studies concerning how the coupling relations between cognition and academic performance might attenuate from childhood to adolescence. Although the variables and the constructs they represented were not a direct extension of those from Ferrer and colleagues' studies (2004; 2008), investigating the emergence and disintegration of the coupling relations between verbal and arithmetic abilities is important in its own right. For instance, such investigation helps shed light on how these two abilities might begin to emerge into two distinct domains that have fueled myriad gender difference studies (Marsh, 1989; Skaalvik & Rankin, 1990).

## Empirical Results

We first compared the fit of the one-regime model to the two-regime model by means of AIC, BIC, and sample size-adjusted BIC. In fitting the one-regime model, three covariances of the between-person difference components (namely, the covariance between deviations in initial reading level and deviations in the constant slope in arithmetic skills, $\psi_{41}$; the covariance between the deviations in initial arithmetic level and deviations in constant slope in reading, $\psi_{32}$; and the covariance between the two deviations in slopes, $\psi_{42}$) were estimated to be close to zero and their inclusion in the model led to a singular numerical Hessian matrix. Thus, these three covariance terms were fixed at zero in all subsequent models. Overall, the two-regime model was observed to show an improved fit compared with the one-regime model as indicated by the AIC, BIC, and adjusted BIC. With 18 free parameters, the one-regime model showed an AIC of −1,244.15, BIC of −1,140.29, and a sample-size adjusted BIC of −1,197.48. This is in contrast to a two-regime model with no covariate, which was associated with an AIC of −2,567.31, BIC of −2,417.29 and a sample-size adjusted BIC of −2,499.90, with 26 parameters.

After identifying the number of regimes by means of information criterion-based indices, we then considered a series of models with covariates within which the original two–regime model was nested. Wald tests as well as likelihood ratio tests were used where appropriate (Jedidi, Jagpal, & DeSarbo, 1997a, 1997b; Yung, 1997), to further refine the regime-switching BDCS model. When covariates were included to predict initial class membership, all of the associated logit slopes, except for that of the effect of time spent reading outside of school, were statistically different from zero. This covariate was omitted in all subsequent models.

We then allowed the initial transition probability matrix to assume different values from the rest of the study span to test whether and how initial between-regime differences in initial levels and constant slopes might translate into later between-regime differences in coupling between the reading and arithmetic learning processes. The two-regime model with covariates was nested within this model and a follow-up likelihood ratio test was used to assess whether the transition probabilities showed sudden shifts after kindergarten. Freeing up the transition probabilities that governed the transition between regimes from time $t = 1$ and $t = 2$ compared with the rest of the study period led to a substantial improvement in fit ($\Delta\chi^2 = 207$ with 2 degrees of freedom). The two resultant transition probability matrices, as discussed later, were characterized by relatively distinct transition patterns, further validating the need to relax these invariance constraints.

Parameter estimates from the best–fitting model are summarized in Table 1.[3] Based on the obtained estimates, the two regimes identified at Time 1 (i.e., during kindergarten), may be regarded as a high (Regime 1) versus low performing

[3]We also considered a three-regime model, but the two-regime model with covariates already took approximately 1.46 hr to run on an Intel(R) Core(TM) 2.39 GHz 2-quad computer with 3.00 GB of RAM when specified with 10 sets of random starting values in the initial stage and 3 random sets of starting values in the final stage. The three-regime model also did not converge. Due to the lack of clear theoretical reason for positing a three-regime model over other alternative models with fewer regimes, we did not consider the three-regime model further.

TABLE 1
Parameter Estimates Obtained From Fitting the Two-Regime Bivariate Dual Change Score (BDCS) Model (R1 = Decoupled Learning Regime; R2 = Coupled Learning Regime)

| *Parameters* | *Meanings* | *Estimates (SE)* |
|---|---|---|
| R1 $\mu_{\eta_{Reading}}$ | Regime 1 group initial mean for reading[a] | −0.74 (0.02) |
| R2 $\mu_{\eta_{Reading}}$ | Regime 2 group initial mean for reading[a] | −1.34 (0.02) |
| R1 $\mu_{\eta_{Arithmetic}}$ | Regime 1 group initial mean for arithmetic[a] | −0.65 (0.01) |
| R2 $\mu_{\eta_{Arithmetic}}$ | Regime 2 group initial mean for arithmetic[a] | −1.17 (0.02) |
| R1 $\mu_{\alpha_{Reading}}$ | Regime 1 group constant slope for reading | 0.61 (0.01) |
| R2 $\mu_{\alpha_{Reading}}$ | Regime 2 group constant slope for reading | 0.43 (0.01) |
| R1 $\mu_{\alpha_{Arithmetic}}$ | Regime 1 group constant slope for arithmetic | 0.55 (0.004) |
| R2 $\mu_{\alpha_{Arithmetic}}$ | Regime 2 group constant slope for arithmetic | 0.39 (0.01) |
| R1 $\beta_{11}$ | Regime 1 reading self-feedback parameter | −0.22 (0.01) |
| R2 $\beta_{11}$ | Regime 2 reading self-feedback parameter | −0.74 (0.01) |
| R1 $\beta_{22}$ | Regime 1 arithmetic self-feedback parameter | −0.22 (0.01) |
| R2 $\beta_{22}$ | Regime 2 arithmetic self-feedback parameter | −0.22 (0.02) |
| R1 $\beta_{12}$ | Regime 1 arithmetic → reading coupling (cross-regression) | = 0 |
| R1 $\beta_{21}$ | Regime 1 reading → arithmetic coupling (cross-regression) | = 0 |
| R2 $\beta_{12}$ | Regime 2 arithmetic → reading coupling (cross-regression) | 0.29 (0.01) |
| R2 $\beta_{21}$ | Regime 2 reading → arithmetic coupling (cross-regression) | −0.11 (0.01) |
| $\omega_{11}$ | Measurement error variance for reading scores | 0.02 (0.000) |
| $\omega_{22}$ | Measurement error variance for arithmetic scores | 0.02 (0.001) |
| $\psi_{11}$ | Variance in initial level, reading | 0.14 (0.01) |
| $\psi_{22}$ | Variance in constant slope, reading | 0.01 (0.000) |
| $\psi_{33}$ | Variance in initial level, arithmetic | 0.09 (0.003) |
| $\psi_{44}$ | Variance in constant slope,arithmetic | 0.01 (0.000) |
| $\psi_{21}$ | Cov(initial level reading, constant slope reading) | 0.01 (0.001) |
| $\psi_{31}$ | Cov(initial level arithmetic, initial level reading) | 0.08 (0.004) |
| $\psi_{32}$ | Cov(initial level arithmetic, constant slope reading) | = 0 |
| $\psi_{41}$ | Cov(initial level reading, constant slope arithmetic) | = 0 |
| $\psi_{42}$ | Cov(constant slope reading, constant slope arithmetic) | = 0 |
| $\psi_{43}$ | Cov(initial level arithmetic, constant slope arithmetic) | 0.01 (0.001) |
| $a_{11}$ | Logit intercept for being in regime 1 at $t = 1$ | −1.78 (0.25) |
| $b_{11,poverty}$ | Regression slope of poverty status on logodds of regime 1 | 1.84 (0.19) |
| $b_{11,readinghours}$ | Regression slope of reading hours on logodds of regime 1 | = 0 |
| $b_{11,HSvsNoHS}$ | Regression slope of *HSvsNoHS* on logodds of regime 1 | 2.23 (0.29) |
| $b_{11,COLLEGE}$ | Regression slope of *COLLEGE* on logodds of regime 1 | 1.05 (0.09) |
| $b_{11,gender}$ | Regression slope of gender on logodds of regime 1 | 0.48 (0.15) |
| $a_{12}$ | Logit intercept of transitioning into R1 at time 2 from R2 at time 1 | = 0 |
| $b_{12,0}$ | Deviation in logodds of staying in R1 relative to $a_{12}$ | 1.93 (0.27) |
| $b_{12,gender}$ | Deviation in logodds of staying in R1 due to gender | = 0 |
| $a_{1t}$, $t = 3, \ldots 9$ | Logit intercept of being in R1 at time $t$ | −1.80 (0.13) |
| $b_{1t,0}$, $t = 3, \ldots 9$ | Deviation in logodds of staying in R1 relative to $a_{1t}$ | = 0 |
| $b_{1t,gender}$, $t = 3, \ldots 9$ | Deviation in logodds of staying in R1 relative to $a_{1t}$ due to gender | = 0 |

[a]Given that we allowed the transition probability matrix at $t = 2$ to differ from subsequent transition probability matrix at later time points, the two regimes that existed at Time 1 were not necessarily the same as the two regimes (i.e., the *Coupled* and *Decoupled* regimes) extracted from other later time points.

class (Regime 2). That is, the latter was characterized by lower average initial scores and average constant slopes than the other class on both the reading and arithmetic tasks. Four covariates were identified to be significant predictors of this initial class membership, including gender of the students, poverty status, parental high school education, and parental college education. Compared with male students, female students were 1.62 times more likely to be in the high as opposed to low performance regime in kindergarten. Relative to children from households that were living below the poverty threshold, children from households that were living at or above the poverty threshold were 6.31 times more likely to be in the high as opposed to low performance regime in kindergarten. Parental high school education (i.e., the effect of *HSvsNoHS*) led to the most pronounced difference in predicting regime membership: compared with children whose parents did not complete high school, children whose parents completed high school or above high school education were 9.27 times more likely to be in the high than in the low performance regime. In addition, children whose parents possessed at least a college degree were 2.85 times more likely to be in the high than in the low performance regime compared with children whose parents did not complete college education formally (but had at least a high school diploma or its equivalent).

Across both regimes, significant between-individual differences were found in the initial levels and constant slopes of the two tasks. The positive covariances among some of the individual difference components indicated that children who started out with higher initial levels on one task tended to show higher initial levels on the other task. In addition, children showing higher initial levels on a task tended to show higher constant slopes on the task.

No gender differences were found in the transition probability patterns. Based on the estimates from the best-fitting two-regime BDCS model, the transition probability matrix that governed the initial switching of regimes, $\boldsymbol{\pi}_{i2}$, is given by

$$\boldsymbol{\pi}_{i2} = \begin{array}{r} \\ \text{High performance class at } t = 1 \\ \text{Low performance class at } t = 1 \end{array} \begin{array}{c} \begin{array}{cc} \text{Decoupled at } t = 2 & \text{Coupled at } t = 2 \end{array} \\ \begin{bmatrix} .87(.03) & .13(.03) \\ =.50 & =.50 \end{bmatrix} \end{array}, \text{ for all } i. \quad (11)$$

That is, children who started in the high performance class in kindergarten had a relatively high probability (.87) of transitioning into the *Decoupled* regime in first grade whereas those who started out in the low performance class had an even odd of moving either into the *Decoupled* or the *Coupled* regime.[4]

[4]The logit intercept parameter, $a_{12}$, was not significantly different from zero. Setting this parameter to zero in the best-fitting model led to an equal probability of transitioning from the low performance class into any of the two regimes at $t = 2$.

Given the time invariance constraints on the group constant slope parameters and the finding that most individuals from the high–performing class at $t = 1$ transitioned into the *Decoupled* regime at $t = 2$, individuals in the *Decoupled* regime also manifested higher average initial levels and constant slopes on the two tasks compared with individuals who were in *Coupled* regime.

The self-feedback parameters for individuals' reading and arithmetic learning trajectories were negative and significantly different from zero in both of the regimes. However, the self-feedback parameter for reading was substantially higher (i.e., less negative) in the *Decoupled Learning* regime than in the *Coupled Learning* regime. When the self-feedback parameter is negative, this indicates that when a participant's reading at the previous time point was high, his or her reading score tends not to increase as much on the next occasion. Thus, the negative self-feedback parameter "pulls" the individual toward an asymptote, the level of which is defined jointly by this parameter and the person-specific constant (linear) slope, $\alpha_{1,i}$. If coupling is present, the asymptote of a change process also depends in part on the strength at which it is coupled to another change process and the corresponding level of the other process.

The coupling (cross-regression) parameters between the reading and arithmetic learning processes were indeed estimated to be statistically different from zero in the *Coupled* regime in both directions (i.e., both in the direction of reading → arithmetic as well as arithmetic → reading coupling). In particular, a positive coupling weight was observed in the direction from arithmetic ability to reading ability whereas a negative coupling was observed in the direction from reading ability to arithmetic ability. Thus, arithmetic scores at one occasion were positively associated with increases in reading scores at the next occasion, so the higher an individual's arithmetic score at one time, the more positive the change in reading at the next measurement occasion. However, a kind of "antagonistic" lagged influence was found in the other direction; that is, an especially high (low) reading score on the previous occasion was associated with a smaller (greater) increase in arithmetic score on the next occasion. In contrast, the coupling parameters of the *Decoupled* were not significantly different from zero even when they were freed to vary.

Predicted trajectories simulated using parameter estimates from the best-fitting two-regime BDCS model (see Figure 2a and b) assuming no switching between regimes can help to further clarify some of the differences between regimes. The corresponding vector field plots shown in Figure 2c and d portray the predicted levels of the reading and arithmetic scores at time $t$ for an individual with the group average initial levels and constant slope, but a range of different scores at time $t$-1. The plots indicate that regardless of the differences in values at time $t$-1, an individual with average group-based parameters was predicted to eventually converge to the regime-specific predicted asymptotes over time. For the *Decoupled Learning* regime, the asymptote for the arithmetic

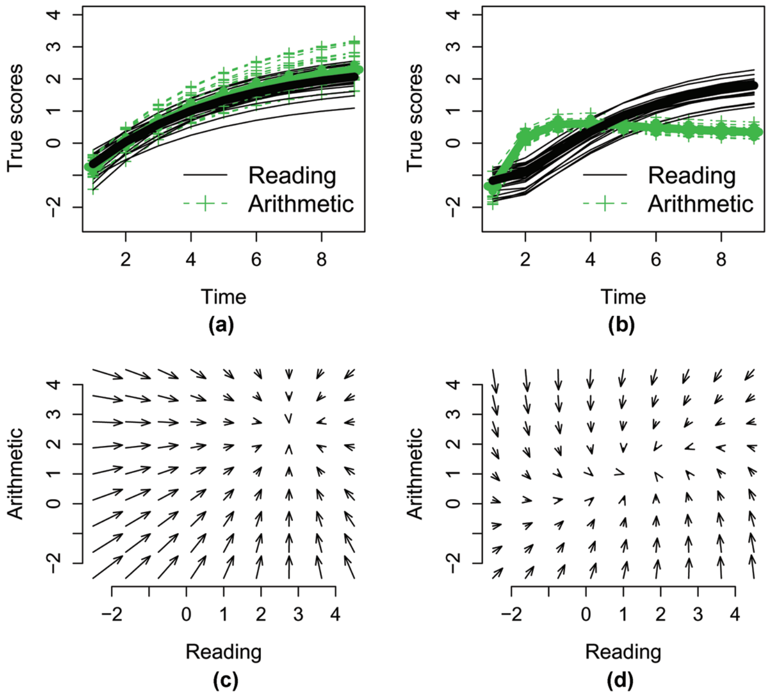


FIGURE 2 (a) & (b) Predicted trajectories simulated using parameter estimates obtained from the best-fitting two-regime Bivariate Dual Change Score (BDCS) model with regime switching; (c) & (d) corresponding vector field plots simulated using different starting values of reading and arithmetic scores. (a) & (c) *Decoupled Learning* regime (R1). (b) & (d) *Coupled Learning* regime (R2) (color figure available online).

task was predicted to assume an approximate value of 2.0 and for the reading task an asymptote of approximately 3.0. For the *Coupled Learning* regime, the arithmetic asymptote was predicted to fall around a value of 1.0 and the reading task a value of 1.5. Overall, the *Coupled Learning* regime was characterized by lower asymptotes. Thus, an individual who was predicted to start out in the high-performing and *Decoupled Learning* regimes who then switched into the *Coupled Learning* regime was expected to show sharp rises in reading and arithmetic skills in early school years that unfolded in a decoupled manner. After the first grade, however, the rises were projected to slow down abruptly to a lower predicted asymptote (particularly for the arithmetic task) during later school years. If the individual remained in the *Coupled Learning* regime, subsequent rises in reading scores may actually lead to decreases in arithmetic scores due to the negative coupling from reading to arithmetic scores.

The transition probability matrix that governed the switching of regimes throughout the remaining study span, $\boldsymbol{\pi}_{it}$, is given by

$$\boldsymbol{\pi}_{it} = \begin{array}{r} \\ \text{Decoupled (R1) at } t-1 \\ \text{Coupled (R2) at } t-1 \end{array} \begin{array}{c} \begin{array}{cc} \text{Decoupled (R1) at } t & \text{Coupled (R2) at } t \end{array} \\ \begin{bmatrix} .14(.02) & .86(.02) \\ .14(.02) & .86(.02) \end{bmatrix} \end{array}, \quad \text{for all } i \text{ and } t = 3, \ldots, 9. \tag{12}$$

The transition probability patterns beyond the first grade were somewhat distinct from the initial transition probabilities shown in Equation 11. In contrast to the *Coupled Learning* regime, which showed relatively high stability (i.e., individuals tended to stay in this regime once they transitioned into it), individuals tended *not* to stay in the *Decoupled Learning* regime once they transitioned into it. A plot of the posterior probabilities of being in the *Coupled Learning* regime from a random subsample of 500 participants is shown in Figure 3a. A plot of



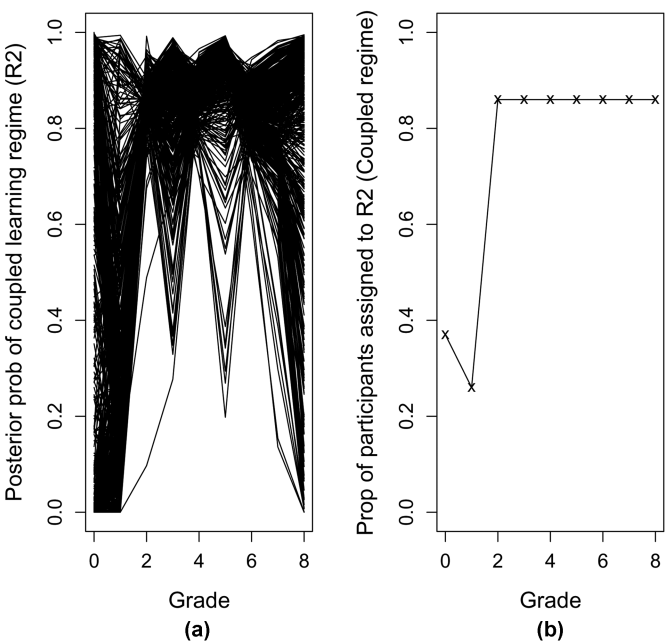


FIGURE 3 (a) A plot of the posterior probability (Prob) of being in the *Coupled Learning* regime (R2) at each time point from a random subsample of 500 participants; (b) a plot of the proportion (Prop) of participants with higher posterior probability of being in the *Coupled Learning* regime than in the *Decoupled Learning* regime at each time point.

the total proportion of participants being assigned to this regime based on their highest posterior probabilities is shown in Figure 3b. As revealed by the plots, the *Coupled Learning* regime tended to be more prevalent during the later school years (i.e., after the first grade), with most children having transitioned into the *Coupled Learning* regime by the second grade. Thus, the acquisition of reading and arithmetic skills took place in more of a coupled fashion starting from the second grade. Some children, however, continued to show signs of transition between the two regimes.



## SIMULATION STUDY

### Simulation Designs

The purpose of the simulation study was to evaluate the performance of the ML-EM approach in recovering the true values of the parameters; their associated *SE*s, and relatedly, individuals' true underlying classes. Several factors that affect estimation results in MSEM models have been considered by other researchers (e.g., Dolan & van der Maas, 1998; Henson, Reise, & Kim, 2007; Lubke & Muthén, 2007; Tueller & Lubke, 2010; Yung, 1997). Some of the well-established findings include the choice of fit indices to help extract the appropriate number of latent classes or regimes (e.g., sample-adjusted BIC was concluded as the recommended measure of fit in a number of studies; Henson, Reise, & Kim, 2007; Nylund, Asparouhov, & Muthén, 2007); the extent of class separation needed to enable satisfactory parameter recovery (e.g., Dolan & van der Maas, 1998; Lubke & Muthén, 2007; Yung, 1997); and issues raised by unbalanced class sizes when the smaller classes are coupled with low sample sizes (Tueller & Lubke, 2010), particularly in regard to class assignment. Other known factors that affect parameter recovery in general (e.g., higher reliability of the indicators of choice and the number of time points) are also applicable but not considered here. Instead, we focused on evaluating aspects that pertained to the dynamics of the underlying processes and the transition between regimes over time.

The best-fitting model from empirical model fitting was used as a basis to construct our simulation model. To mirror the parameter space of our empirical illustration, the population values of all the time-invariant parameters were chosen to closely approximate those obtained from empirical model fitting. Two sample sizes were considered with the same number of measurement occasions, namely, with $T = 5$, $n = 1{,}000$, and $T = 5$, $n = 400$. The first condition was selected to provide a large-sample comparison that is more realistic in empirical settings than the uncharacteristically large sample size in our empirical illustration (with five nonmissing time points and 2,369 participants). The second condition consisted of an even smaller number of participants to mirror sample sizes commonly seen in empirical studies. Besides using the parameter values that closely mirror those observed in our empirical study as true data generation values, we also included a second condition with different transition probability values and a third condition with different regime-specific coupling parameters. Varying the values of the transition probabilities has the effect of altering the stability of the two regimes. As shown in previous simulation studies involving time series data (Yang & Chow, 2010), a larger proportion of the data tends to come from regimes with higher staying probabilities. That is, regimes with low staying probabilities tend to be fleeting and may lead to a scarcity of realizations from these regimes to enable accurate recovery of regime-specific parameters. In addition, varying dynamic parameters such as the coupling parameters can lead to different degrees of separation between regimes. Whereas the effect of regime separation and pertinent guidelines have been provided in the context of MSEMs (e.g., Dolan & van der Maas, 1998; Yung, 1997), the consequences of altering transition probability values and regime separation (and hence, effect size) in the context of MSEM-RS with longitudinal panel data were yet to be evaluated.

To summarize, we considered 2 sample sizes $\times$ 3 sets of parameter values (one set to mirror the parameter values of the empirical study, a second set to evaluate the effect of a different set of transition probability values, and a third set to examine the effect of different coupling parameters and consequently, effect size). This yielded a total of six simulation conditions. For each condition, 500 Monte Carlo replications were conducted.

We only considered the case of balanced initial class size, with initial class probability set to .5 for all conditions. To simplify notation, we refer to this parameter in our results as $\pi_1$. With two regimes in total, only one element of each row of the $2 \times 2$ transition probability matrix can be freely estimated. We chose to estimate the parameters $\pi_{it,11}$ and $\pi_{it,22}$. Because these parameters were specified to be invariant over time and individuals in the present simulation study, the person and time indices were omitted in all subsequent descriptions to ease presentation. We considered two possible sets of values for the transition probability matrix. The first set of transition probabilities entailed relatively high stability of staying within each regime once individuals had transitioned into it:

$$\mathbf{P} = \begin{bmatrix} \pi_{11} = .9 & 1 - \pi_{11} = .1 \\ 1 - \pi_{22} = .3 & \pi_{22} = .7 \end{bmatrix}. \tag{13}$$

The second set of transition probability values, in contrast, showed relatively high probabilities of transitioning out of each of the regimes, with

$$\mathbf{P} = \begin{bmatrix} \pi_{11} = .6 & 1 - \pi_{11} = .4 \\ 1 - \pi_{22} = .5 & \pi_{22} = .5 \end{bmatrix}. \tag{14}$$

Yang and Chow (2010) showed in a simulation study involving time series data that having a reasonably large probability of staying within regimes helped improve the accuracy of parameter and standard error estimation. In particular, such transition probability values help ensure that a sufficient number of time points is available from each regime, thereby providing enough information to distinguish between the regimes, especially when the separation between regimes is small.

For the third set of parameter values, we changed the coupling parameters in the second regime from $\beta_{12} = .29$ and $\beta_{21} = -.10$ to $\beta_{12} = .5$ and $\beta_{21} = -.3$, respectively. This choice of coupling parameters led to a sharper rise in the second learning process, followed by a gradual decline after an early asymptote was reached, due largely to being negatively coupled to the first process. These renewed dynamics resulted in a *smaller* multivariate Hosmer's measure of distance between the two regimes (Dolan & van der Maas, 1998; Hosmer, 1974), given by

$$\max_{h \in \{1,2\}} [(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)' \boldsymbol{\Sigma}_h^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]^{1/2}. \tag{15}$$

In our case, the second set of coupling parameters led to a multivariate Hosmer's distance of 2.33 compared with 5.63 with the first set of coupling parameters. In this way, the second set of coupling parameters yielded a smaller separation between the two regimes and can be regarded as having a smaller effect size. To avoid the issue of "label switching" (McLachlan & Peel, 1995; Tueller & Lubke, 2010), namely, the issue that the labels for different regimes are switched when no explicit constraints are imposed to distinguish among the regimes, we imposed the confirmatory constraint that the coupling parameters in the first regime were equal to zero (i.e., the first regime was the *Decoupled* regime).

The true values of other parameters used to generate the data are summarized in Tables 2–7. The root mean squared error (RMSE) and bias were used to quantify the performance of the ML point estimator. The empirical *SE* of a parameter (i.e., the standard deviation of the estimates of a particular parameter across all Monte Carlo runs) was used as the "true" standard error. As a measure of the relative performance of the *SE* estimates, we also included the average relative deviance of an *SE* estimate of an estimator (aRDSE), namely, the difference between an *SE* estimate and the true *SE* over the true *SE*, averaged across Monte Carlo runs.

Ninety-five percent confidence intervals (CIs) were constructed for each of the $N = 500$ Monte Carlo samples in each condition by adding and subtracting 1.96**SE* estimate in each replication to the parameter estimate from the replication. The coverage performance of a CI was assessed with its empirical coverage rate, namely, the proportion of 95% CIs covering $\theta$ across the 500 Monte Carlo replications. In addition, we computed the proportions of 95% CIs not covering

TABLE 2
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model With $T = 5$, $n = 1{,}000$, and High Probabilities of Staying Within Regimes

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −0.77 | 0.04 | −0.04 | 0.083 | 0.023 | −0.73 | 0.88 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.62 | 0.00 | 0.00 | 0.008 | 0.007 | −0.09 | 0.99 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.31 | 0.03 | 0.03 | 0.081 | 0.022 | −0.72 | 0.90 | 1.00 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.45 | 0.01 | 0.01 | 0.016 | 0.008 | −0.51 | 0.92 | 1.00 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.68 | 0.03 | −0.03 | 0.063 | 0.020 | −0.69 | 0.89 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.55 | 0.00 | −0.00 | 0.007 | 0.007 | −0.09 | 0.98 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.14 | 0.03 | 0.03 | 0.065 | 0.018 | −0.72 | 0.90 | 1.00 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.39 | 0.00 | 0.00 | 0.014 | 0.007 | −0.51 | 0.91 | 1.00 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | 0.00 | 0.005 | 0.006 | 0.20 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.004 | 0.005 | 0.27 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.76 | 0.01 | −0.01 | 0.030 | 0.035 | 0.17 | 0.99 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.21 | 0.01 | 0.01 | 0.037 | 0.047 | 0.27 | 0.99 | 1.00 |
| R2 $\beta_{12}$ | 0.29 | 0.31 | 0.02 | 0.02 | 0.035 | 0.041 | 0.16 | 0.99 | 1.00 |
| R2 $\beta_{21}$ | −0.10 | −0.11 | 0.01 | −0.01 | 0.031 | 0.040 | 0.28 | 0.99 | 0.80 |
| $\pi_{2,11}$ | 0.90 | 0.89 | 0.01 | −0.01 | 0.013 | 0.015 | 0.19 | 1.00 | 1.00 |
| $\pi_{2,22}$ | 0.70 | 0.70 | 0.00 | −0.00 | 0.020 | 0.025 | 0.27 | 1.00 | 1.00 |
| $\pi_{1,1}$ | 0.50 | 0.49 | 0.01 | −0.01 | 0.035 | 0.027 | −0.24 | 0.97 | 1.00 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.000 | 0.001 | 0.27 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.000 | 0.000 | 0.31 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.13 | 0.00 | 0.00 | 0.024 | 0.009 | −0.63 | 0.89 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.002 | 0.001 | −0.38 | 1.00 | 1.00 |
| $\psi_{33}$ | 0.09 | 0.09 | 0.00 | 0.00 | 0.018 | 0.006 | −0.64 | 0.90 | 1.00 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.00 | −0.00 | 0.001 | 0.001 | −0.27 | 1.00 | 1.00 |
| $\psi_{21}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.007 | 0.002 | −0.69 | 0.91 | 1.00 |
| $\psi_{31}$ | 0.08 | 0.09 | 0.01 | 0.01 | 0.021 | 0.007 | −0.68 | 0.90 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.005 | 0.002 | −0.67 | 0.90 | 1.00 |
| $\psi_{41}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.006 | 0.002 | −0.72 | 0.90 | 0.20 |
| $\psi_{32}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.006 | 0.002 | −0.72 | 0.90 | 0.20 |
| $\psi_{42}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.001 | 0.001 | −0.56 | 1.00 | 0.10 |
| Average | | | 0.01 | 0.00 | 0.024 | 0.015 | −0.23 | 0.95 | 0.99 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) $= \sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias $= \frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/Type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

TABLE 3
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model With $T = 5$, $n = 1{,}000$, and Low Probabilities of Staying Within Regimes

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −0.98 | 0.25 | −0.25 | 0.102 | 0.023 | −0.77 | 0.24 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.56 | 0.06 | −0.06 | 0.033 | 0.011 | −0.67 | 0.32 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.27 | 0.07 | 0.07 | 0.213 | 0.091 | −0.57 | 0.57 | 0.99 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.69 | 0.25 | 0.25 | 0.860 | 0.049 | −0.94 | 0.65 | 0.98 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.86 | 0.21 | −0.21 | 0.088 | 0.022 | −0.75 | 0.24 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.49 | 0.06 | −0.06 | 0.030 | 0.010 | −0.67 | 0.30 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.11 | 0.06 | 0.06 | 0.187 | 0.097 | −0.48 | 0.56 | 0.99 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.46 | 0.07 | 0.07 | 0.121 | 0.055 | −0.54 | 0.63 | 0.99 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.006 | 0.008 | 0.31 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.006 | 0.007 | 0.24 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.78 | 0.03 | −0.03 | 0.022 | 0.030 | 0.35 | 0.93 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.041 | 0.051 | 0.25 | 1.00 | 1.00 |
| R2 $\beta_{12}$ | 0.29 | 0.32 | 0.03 | 0.03 | 0.026 | 0.035 | 0.33 | 0.94 | 1.00 |
| R2 $\beta_{21}$ | −0.10 | −0.10 | 0.01 | 0.01 | 0.033 | 0.042 | 0.26 | 1.00 | 0.65 |
| $\pi_{2,11}$ | 0.60 | 0.59 | 0.01 | −0.01 | 0.031 | 0.030 | −0.02 | 0.96 | 1.00 |
| $\pi_{2,22}$ | 0.50 | 0.51 | 0.01 | 0.01 | 0.051 | 0.053 | 0.04 | 0.96 | 1.00 |
| $\pi_{1,1}$ | 0.50 | 0.78 | 0.28 | 0.28 | 0.193 | 0.074 | −0.62 | 0.52 | 1.00 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.27 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.000 | 0.001 | 0.27 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.19 | 0.06 | 0.06 | 0.031 | 0.011 | −0.66 | 0.26 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.02 | 0.01 | 0.01 | 0.003 | 0.001 | −0.50 | 0.55 | 0.99 |
| $\psi_{33}$ | 0.09 | 0.14 | 0.05 | 0.05 | 0.023 | 0.008 | −0.67 | 0.25 | 1.00 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.002 | 0.001 | −0.41 | 0.99 | 0.99 |
| $\psi_{21}$ | 0.01 | 0.04 | 0.03 | 0.03 | 0.009 | 0.003 | −0.71 | 0.24 | 1.00 |
| $\psi_{31}$ | 0.08 | 0.14 | 0.07 | 0.07 | 0.027 | 0.008 | −0.71 | 0.24 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.03 | 0.02 | 0.02 | 0.007 | 0.002 | −0.68 | 0.25 | 1.00 |
| $\psi_{41}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.009 | 0.002 | −0.74 | 0.24 | 0.86 |
| $\psi_{32}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.008 | 0.002 | −0.72 | 0.24 | 0.86 |
| $\psi_{42}$ | 0.00 | 0.01 | 0.01 | 0.01 | 0.003 | 0.001 | −0.61 | 0.65 | 0.76 |
| Average | | | 0.06 | 0.01 | 0.083 | 0.028 | −0.31 | 0.64 | 0.98 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) = $\sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias = $\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/Type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

TABLE 4
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model With $T = 5$, $n = 1{,}000$, High Probabilities of Staying Within Regimes and Higher Coupling Strengths in Regime 2

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −0.93 | 0.20 | −0.20 | 0.115 | 0.029 | −0.75 | 0.95 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.56 | 0.06 | −0.06 | 0.035 | 0.012 | −0.65 | 0.95 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.22 | 0.12 | 0.12 | 0.129 | 0.063 | −0.52 | 0.96 | 1.00 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.49 | 0.05 | 0.05 | 0.061 | 0.028 | −0.55 | 0.97 | 1.00 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.85 | 0.20 | −0.20 | 0.100 | 0.021 | −0.79 | 0.94 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.51 | 0.04 | −0.04 | 0.026 | 0.010 | −0.63 | 0.95 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.01 | 0.16 | 0.16 | 0.121 | 0.036 | −0.70 | 0.95 | 1.00 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.40 | 0.01 | 0.01 | 0.032 | 0.020 | −0.37 | 0.99 | 1.00 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.004 | 0.005 | 0.40 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.004 | 0.008 | 0.94 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.77 | 0.01 | −0.01 | 0.019 | 0.024 | 0.27 | 1.00 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.26 | 0.04 | −0.04 | 0.023 | 0.031 | 0.38 | 0.99 | 1.00 |
| R2 $\beta_{12}$ | 0.50 | 0.53 | 0.03 | 0.03 | 0.020 | 0.028 | 0.36 | 1.00 | 1.00 |
| R2 $\beta_{21}$ | −0.30 | −0.28 | 0.02 | 0.02 | 0.018 | 0.027 | 0.44 | 1.00 | 1.00 |
| $\pi_{2,11}$ | 0.60 | 0.62 | 0.02 | 0.02 | 0.037 | 0.029 | −0.21 | 0.99 | 1.00 |
| $\pi_{2.22}$ | 0.50 | 0.52 | 0.02 | 0.02 | 0.031 | 0.031 | −0.02 | 0.99 | 1.00 |
| $\pi_{1,1}$ | 0.50 | 0.66 | 0.16 | 0.16 | 0.162 | 0.081 | −0.50 | 0.98 | 1.00 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.000 | 0.001 | 0.15 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.000 | 0.001 | 0.44 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.19 | 0.06 | 0.06 | 0.034 | 0.010 | −0.70 | 0.94 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.003 | 0.002 | −0.48 | 0.98 | 1.00 |
| $\psi_{33}$ | 0.09 | 0.13 | 0.04 | 0.04 | 0.023 | 0.007 | −0.69 | 0.94 | 1.00 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.003 | 0.001 | −0.50 | 1.00 | 1.00 |
| $\psi_{21}$ | 0.01 | 0.03 | 0.02 | 0.02 | 0.011 | 0.003 | −0.72 | 0.94 | 1.00 |
| $\psi_{31}$ | 0.08 | 0.14 | 0.06 | 0.06 | 0.029 | 0.008 | −0.74 | 0.94 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.03 | 0.02 | 0.02 | 0.009 | 0.002 | −0.76 | 0.94 | 1.00 |
| $\psi_{41}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.010 | 0.002 | −0.76 | 0.94 | 0.98 |
| $\psi_{32}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.009 | 0.002 | −0.78 | 0.94 | 0.98 |
| $\psi_{42}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.003 | 0.001 | −0.69 | 0.98 | 0.97 |
| Average | | | 0.05 | 0.01 | 0.040 | 0.020 | −0.26 | 0.97 | 1.00 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) = $\sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias = $\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/Type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

TABLE 5
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model With $T = 5$, $n = 400$, and High Probabilities of Staying Within Regimes

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −0.81 | 0.08 | −0.08 | 0.116 | 0.036 | −0.69 | 0.78 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.61 | 0.01 | −0.01 | 0.017 | 0.012 | −0.29 | 0.93 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.28 | 0.06 | 0.06 | 0.109 | 0.037 | −0.67 | 0.80 | 1.00 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.45 | 0.01 | 0.01 | 0.027 | 0.014 | −0.48 | 0.86 | 1.00 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.71 | 0.06 | −0.06 | 0.088 | 0.033 | −0.63 | 0.79 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.55 | 0.00 | −0.00 | 0.017 | 0.012 | −0.29 | 0.95 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.12 | 0.04 | 0.04 | 0.084 | 0.032 | −0.62 | 0.81 | 1.00 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.40 | 0.01 | 0.01 | 0.022 | 0.013 | −0.43 | 0.85 | 1.00 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.008 | 0.010 | 0.21 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.007 | 0.009 | 0.30 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.76 | 0.01 | −0.01 | 0.047 | 0.059 | 0.27 | 0.99 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.21 | 0.01 | 0.01 | 0.062 | 0.078 | 0.27 | 0.97 | 0.79 |
| R2 $\beta_{12}$ | 0.29 | 0.30 | 0.01 | 0.01 | 0.055 | 0.069 | 0.27 | 0.98 | 0.99 |
| R2 $\beta_{21}$ | −0.10 | −0.11 | 0.01 | −0.01 | 0.053 | 0.067 | 0.26 | 0.99 | 0.44 |
| $\pi_{2,11}$ | 0.90 | 0.89 | 0.01 | −0.01 | 0.028 | 0.028 | −0.00 | 0.98 | 1.00 |
| $\pi_{2,22}$ | 0.70 | 0.70 | 0.00 | −0.00 | 0.037 | 0.043 | 0.17 | 0.99 | 1.00 |
| $\pi_{1,1}$ | 0.50 | 0.50 | 0.00 | −0.00 | 0.072 | 0.051 | −0.29 | 0.96 | 1.00 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.27 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.30 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.14 | 0.01 | 0.01 | 0.031 | 0.015 | −0.53 | 0.80 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.002 | 0.002 | −0.35 | 1.00 | 1.00 |
| $\psi_{33}$ | 0.09 | 0.10 | 0.01 | 0.01 | 0.023 | 0.011 | −0.53 | 0.80 | 1.00 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.00 | 0.00 | 0.002 | 0.001 | −0.19 | 1.00 | 1.00 |
| $\psi_{21}$ | 0.01 | 0.02 | 0.01 | 0.01 | 0.009 | 0.004 | −0.62 | 0.82 | 0.99 |
| $\psi_{31}$ | 0.08 | 0.10 | 0.02 | 0.02 | 0.027 | 0.011 | −0.58 | 0.79 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.02 | 0.01 | 0.01 | 0.007 | 0.003 | −0.58 | 0.82 | 1.00 |
| $\psi_{41}$ | 0.00 | 0.01 | 0.01 | 0.01 | 0.008 | 0.003 | −0.65 | 0.80 | 0.34 |
| $\psi_{32}$ | 0.00 | 0.01 | 0.01 | 0.01 | 0.008 | 0.003 | −0.64 | 0.81 | 0.32 |
| $\psi_{42}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.002 | 0.001 | −0.51 | 1.00 | 0.14 |
| Average | | | 0.01 | 0.00 | 0.037 | 0.025 | −0.21 | 0.91 | 0.97 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) = $\sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias = $\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

TABLE 6
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model with $T = 5$, $n = 400$ and Low Probabilities of Staying Within Regimes

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −1.00 | 0.27 | −0.27 | 0.066 | 0.038 | −0.43 | 0.17 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.55 | 0.07 | −0.07 | 0.025 | 0.016 | −0.36 | 0.24 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.25 | 0.09 | 0.09 | 0.206 | 0.105 | −0.49 | 0.53 | 0.99 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.65 | 0.21 | 0.21 | 0.761 | 0.056 | −0.93 | 0.64 | 0.98 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.88 | 0.23 | −0.23 | 0.057 | 0.033 | −0.41 | 0.17 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.48 | 0.07 | −0.07 | 0.020 | 0.014 | −0.34 | 0.22 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.11 | 0.06 | 0.06 | 0.196 | 0.093 | −0.53 | 0.54 | 0.99 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.46 | 0.07 | 0.07 | 0.117 | 0.047 | −0.60 | 0.61 | 0.98 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.010 | 0.012 | 0.28 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.009 | 0.012 | 0.31 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.78 | 0.03 | −0.03 | 0.038 | 0.049 | 0.28 | 0.96 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | 0.00 | 0.067 | 0.082 | 0.23 | 0.98 | 0.80 |
| R2 $\beta_{12}$ | 0.29 | 0.32 | 0.03 | 0.03 | 0.043 | 0.056 | 0.30 | 0.96 | 1.00 |
| R2 $\beta_{21}$ | −0.10 | −0.10 | 0.00 | 0.00 | 0.055 | 0.067 | 0.23 | 0.99 | 0.38 |
| $\pi_{2,11}$ | 0.60 | 0.58 | 0.02 | −0.02 | 0.046 | 0.050 | 0.09 | 0.98 | 1.00 |
| $\pi_{2,22}$ | 0.50 | 0.51 | 0.01 | 0.01 | 0.080 | 0.092 | 0.15 | 0.95 | 0.99 |
| $\pi_{1,1}$ | 0.50 | 0.83 | 0.33 | 0.33 | 0.161 | 0.088 | −0.46 | 0.42 | 0.98 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.28 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.31 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.20 | 0.07 | 0.07 | 0.022 | 0.017 | −0.21 | 0.20 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.02 | 0.01 | 0.01 | 0.003 | 0.002 | −0.14 | 0.77 | 0.99 |
| $\psi_{33}$ | 0.09 | 0.14 | 0.05 | 0.05 | 0.017 | 0.013 | −0.23 | 0.23 | 0.99 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.002 | 0.002 | 0.02 | 1.00 | 0.99 |
| $\psi_{21}$ | 0.01 | 0.04 | 0.03 | 0.03 | 0.006 | 0.005 | −0.23 | 0.16 | 0.99 |
| $\psi_{31}$ | 0.08 | 0.15 | 0.07 | 0.07 | 0.018 | 0.014 | −0.26 | 0.18 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.03 | 0.02 | 0.02 | 0.005 | 0.004 | −0.23 | 0.16 | 0.99 |
| $\psi_{41}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.005 | 0.004 | −0.28 | 0.16 | 0.96 |
| $\psi_{32}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.005 | 0.004 | −0.26 | 0.17 | 0.96 |
| $\psi_{42}$ | 0.00 | 0.01 | 0.01 | 0.01 | 0.002 | 0.002 | −0.16 | 0.78 | 0.83 |
| Average | | | 0.07 | 0.01 | 0.078 | 0.037 | −0.13 | 0.62 | 0.96 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) = $\sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias = $\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/Type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

TABLE 7
Summary Statistics of the Parameter Estimates (Across 500 Monte Carlo Replications) for the Two-Regime Bivariate Dual Change Score (BDCS) Model With $T = 5$, $n = 400$, High Probabilities of Staying Within Regimes and Higher Coupling Strengths in Regime 2

| | *True* $\theta$ | *Mean* $\hat{\theta}$ | *RMSE* | *Bias* | *SD* $\hat{\theta}$ | *Mean* $\widehat{SE}$ | *RDSE* | *Coverage of 95% CIs* | *Power/Type I Error* |
|---|---|---|---|---|---|---|---|---|---|
| R1 $\mu_{\eta_1}$ | −0.73 | −0.98 | 0.25 | −0.25 | 0.075 | 0.040 | −0.47 | 0.19 | 1.00 |
| R1 $\mu_{\alpha_1}$ | 0.62 | 0.54 | 0.08 | −0.08 | 0.026 | 0.016 | −0.39 | 0.26 | 1.00 |
| R2 $\mu_{\eta_1}$ | −1.34 | −1.27 | 0.07 | 0.07 | 0.219 | 0.102 | −0.53 | 0.55 | 0.99 |
| R2 $\mu_{\alpha_1}$ | 0.44 | 0.60 | 0.17 | 0.17 | 0.610 | 0.044 | −0.93 | 0.59 | 0.99 |
| R1 $\mu_{\eta_2}$ | −0.65 | −0.88 | 0.23 | −0.23 | 0.056 | 0.029 | −0.48 | 0.17 | 1.00 |
| R1 $\mu_{\alpha_2}$ | 0.55 | 0.49 | 0.06 | −0.06 | 0.027 | 0.015 | −0.44 | 0.37 | 1.00 |
| R2 $\mu_{\eta_2}$ | −1.17 | −1.07 | 0.10 | 0.10 | 0.195 | 0.078 | −0.60 | 0.46 | 1.00 |
| R2 $\mu_{\alpha_2}$ | 0.39 | 0.43 | 0.04 | 0.04 | 0.077 | 0.044 | −0.43 | 0.68 | 0.99 |
| R1 $\beta_{11}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.007 | 0.008 | 0.27 | 1.00 | 1.00 |
| R1 $\beta_{22}$ | −0.22 | −0.22 | 0.00 | −0.00 | 0.010 | 0.012 | 0.19 | 1.00 | 1.00 |
| R2 $\beta_{11}$ | −0.75 | −0.77 | 0.02 | −0.02 | 0.031 | 0.039 | 0.27 | 0.98 | 1.00 |
| R2 $\beta_{22}$ | −0.22 | −0.25 | 0.03 | −0.03 | 0.041 | 0.050 | 0.21 | 0.96 | 1.00 |
| R2 $\beta_{12}$ | 0.50 | 0.53 | 0.03 | 0.03 | 0.034 | 0.044 | 0.29 | 0.95 | 1.00 |
| R2 $\beta_{21}$ | −0.30 | −0.28 | 0.02 | 0.02 | 0.035 | 0.043 | 0.20 | 0.97 | 1.00 |
| $\pi_{2,11}$ | 0.60 | 0.61 | 0.01 | 0.01 | 0.047 | 0.045 | −0.04 | 0.94 | 1.00 |
| $\pi_{2,22}$ | 0.50 | 0.51 | 0.01 | 0.01 | 0.051 | 0.053 | 0.04 | 0.96 | 1.00 |
| $\pi_{1,1}$ | 0.50 | 0.76 | 0.26 | 0.26 | 0.180 | 0.088 | −0.51 | 0.53 | 0.99 |
| $\omega_{11}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.28 | 1.00 | 1.00 |
| $\omega_{22}$ | 0.02 | 0.02 | 0.00 | −0.00 | 0.001 | 0.001 | 0.25 | 1.00 | 1.00 |
| $\psi_{11}$ | 0.13 | 0.20 | 0.07 | 0.07 | 0.025 | 0.017 | −0.30 | 0.22 | 1.00 |
| $\psi_{22}$ | 0.01 | 0.02 | 0.01 | 0.01 | 0.002 | 0.002 | −0.13 | 0.82 | 0.99 |
| $\psi_{33}$ | 0.09 | 0.14 | 0.05 | 0.05 | 0.016 | 0.011 | −0.30 | 0.21 | 1.00 |
| $\psi_{44}$ | 0.01 | 0.01 | 0.01 | 0.01 | 0.002 | 0.002 | −0.13 | 0.99 | 0.99 |
| $\psi_{21}$ | 0.01 | 0.04 | 0.03 | 0.03 | 0.007 | 0.005 | −0.35 | 0.20 | 0.99 |
| $\psi_{31}$ | 0.08 | 0.15 | 0.07 | 0.07 | 0.019 | 0.013 | −0.34 | 0.19 | 1.00 |
| $\psi_{43}$ | 0.01 | 0.03 | 0.02 | 0.02 | 0.005 | 0.003 | −0.34 | 0.19 | 1.00 |
| $\psi_{41}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.006 | 0.004 | −0.33 | 0.22 | 0.92 |
| $\psi_{32}$ | 0.00 | 0.02 | 0.02 | 0.02 | 0.006 | 0.003 | −0.40 | 0.18 | 0.94 |
| $\psi_{42}$ | 0.00 | 0.01 | 0.01 | 0.01 | 0.002 | 0.001 | −0.26 | 0.80 | 0.78 |
| Average | | | 0.06 | 0.01 | 0.069 | 0.031 | −0.18 | 0.63 | 1.00 |

*Note.* True $\theta$ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{m=1}^{N}\hat{\theta}_m$, where $\hat{\theta}_m$ = estimate of $\theta$ from the $m$th Monte Carlo replication; RMSE (root mean squared error) = $\sqrt{\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)^2}$; rBias = relative bias = $\frac{1}{N}\sum_{m=1}^{N}(\hat{\theta}_m - \theta)/\theta$; *SD* = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; aRDSE (average relative deviance of an SE estimate) = average relative deviance of $\widehat{SE}$ = (Mean $\widehat{SE} - SE)/SE$; coverage = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain the true $\theta$; power/Type I error = proportion of 95% confidence intervals (CIs) across the Monte Carlo runs that contain 0.

zero across the 500 Monte Carlo replications. For parameters whose true values were different from zero in the population, these proportions served as power estimates; for parameters whose true values were equal to zero (including $\psi_{41}$, $\psi_{32}$, and $\psi_{42}$), these coverage rates were taken as a measure of Type I error rate.

We were also interested in examining the effects of the aforementioned factors on class classification. We assigned each person's data at each time point to the regime with the highest posterior probability, computed using Equation 9. In the presence of only two regimes, this amounted to assigning an individual to a particular regime at a particular time point when the corresponding posterior probability was higher than .5. To assess the accuracy of the classification results, we computed the proportion of correct regime classification (across all persons and time points) within each condition and compared these proportions across conditions.



## Simulation Results

Statistical properties of the ML estimator across all conditions are summarized in Tables 2–7 in conjunction with the true parameter values used for data generation. To facilitate comparison across conditions, we included the average RMSE, bias, and 95% coverage rate for each condition across all but the parameters whose true population values were equal to zero in Tables 2–7. In general, the point estimates were characterized by low RMSEs and biases across all conditions. Consistent with the simulation results evidenced in other studies involving regime-switching models with large $T$, the condition with high probabilities of staying within regimes yielded lower RMSEs and biases when averaged across parameters than the condition with low probabilities of staying within regimes. Lowering the separation between regimes by increasing the coupling strengths between the two processes also led to a slight increase in the average RMSE and bias of the point estimates. The extent of increase in biases was slightly less than that observed when the transition probabilities were altered. Parameters that were noticeably affected were the group average initial level and constant slope parameters (or in other words, the fixed effects parameters), the initial class probability (e.g., $\pi_1$), and the variance and covariance parameters.

The point estimates remained largely unbiased within the condition with high regime stability even when $n$ was reduced from 1,000 to 400 (see Table 5 compared with Table 2), although the RMSEs tended to be slightly higher, especially among the group average initial level and constant slope parameters, as well as the variance-covariance parameters. We did not observe a disproportionate increase in biases of the point estimates in the other two parameter conditions (i.e., the conditions with lower regime stability and higher coupling strengths) when sample size was reduced. The accuracy of dynamic parameters such as $\beta_{11}$, $\beta_{12}$, and the transition probability parameters were relatively unaffected by the reduction in the number of participants. Based on past simulation studies, our speculation was that these parameters may be affected more by factors such as the number of time points available from each participant and its interaction effects with the ranges of the dynamic parameters and the transition probability parameters.

As expected, the condition with larger sample ($n = 1{,}000$) was observed to exhibit greater precision (in the sense of smaller average *SE* estimates and relatedly, smaller true *SE*s) compared with the $n = 400$ condition. The *SE* estimates provided a reasonable approximation to the true variability of parameters such as the dynamic and transition probability parameters. However, the *SE* estimates generally underestimated the true variability of the group average initial level, the group average constant slope parameters, and some of the variance and covariance parameters. Larger discrepancies were observed between the *SE* estimates and the Monte Carlo *SE*s in the conditions with lower regime stability and to a lesser extent, the conditions with high coupling strengths. Surprisingly, smaller discrepancies in *SE* estimation were actually observed in the smaller sample conditions compared with the $n = 1{,}000$ conditions. The smaller discrepancies in *SE* estimation in the finite sample conditions may be attributable to the strengths of the robust *SE* estimator in "adjusting" *SE* estimates in the presence of outliers in finite samples.

For most conditions, the coverage rates of the 95% CIs were close but slightly lower than the .95 nominal rate for most parameters. Lower coverage rates (namely, compared with the nominal rate of .95) were observed for some the variance–covariance and initial class probability parameters. Substantially lower coverage rates were observed for the parameters in conditions involving low probabilities of staying within regimes, and when high coupling strengths were paired with a smaller sample size. Overall, due in part to the robust *SE* estimates in the finite sample conditions, coverage rates were relatively unaffected by sample size, except when a smaller sample size was paired with low separation between regimes (i.e., in the conditions with high coupling strengths). Larger sample sizes did, however, improve the coverage rates of fixed effects parameters such as the group average intercepts and constant slope parameters. In addition, high type I error rates were observed for the parameters $\psi_{41}$, $\psi_{32}$, and $\psi_{42}$ in the conditions with low regime staying probabilities and low separation between regimes. This may be related to the inflated biases in the fixed effects parameters in these conditions.

Evaluation of the posterior probability estimates indicated that these estimates were able to capture some but not all of the within-person shifts in regimes over time. Relatively high proportions of regime classification errors were observed (see Table 8) when these posterior probability estimates were used for regime classification purposes. Greater inaccuracies were evidenced when the time spent staying in a particular regime was brief and when the regime separation was

TABLE 8
Average Proportions of Incorrect Regime Classification Across 500 Monte Carlo Replications

| *Regime Properties* | *n = 400* | *n = 1,000* |
|---|---|---|
| High staying probabilities | 0.19 | 0.17 |
| Low staying probabilities | 0.33 | 0.32 |
| High coupling strengths | 0.29 | 0.27 |



small. Slight improvements in classification accuracy were observed in the $n = 1{,}000$ conditions compared to the $n = 400$ conditions. This is likely attributable to the increased accuracy of the point estimates of the parameters in the large sample condition. Overall, the less satisfactory classification results associated with regime switching models are consistent with previous simulation results reported elsewhere with MSEM models (e.g., Tueller & Lubke, 2010). The inclusion of covariates may aid the estimation of initial class membership and transition probabilities, thereby improving regime classification accuracy. Still, the general appropriateness of regime–switching models when used for classification and forecasting purposes needs to be evaluated with caution.

## DISCUSSION

In this article, we illustrated the utility of one particular MSEM-RS model—the regime-switching BDCS model—in representing multivariate processes with distinct dynamics during different "phases" of the data. The two-regime BDCS model was found to provide a better fit to the ECLS–K data than a one-regime model. Two regimes were identified: a *Coupled Learning* regime with statistically significant coupling parameters and a *Decoupled Learning* regime with no coupling between the reading and arithmetic learning processes. The *Decoupled Learning* regime was predicted to show higher asymptotes on the reading and arithmetic tasks than the *Coupled Learning* regime.

Based on the posterior probabilities from model estimation, we found that the *Coupled* regime was more prevalent in later school years. Most children were found to have transitioned into this regime after the second grade. Most individuals who transitioned into the *Decoupled* regime at first grade came from the "high-performing class" with higher initial reading and arithmetic skills and higher constant slopes at kindergarten. Those who then switched into the *Coupled* regime after the second grade were expected to show sharp rises in reading and arithmetic skills in early school years that slowed down abruptly to lower predicted asymptotes during later school years. This indicates that there may be more performance heterogeneities in early school years and a one–regime model (with coupling parameters freely estimated) may be adequate for describing the learning dynamics in later school years.

The nature of the coupling parameters in the *Coupled* regime with a positive coupling coefficient from arithmetic to changes in reading and a negative coupling coefficient from reading to changes in arithmetic was inconsistent with the finding reported by Grimm (2006) that a positive effect was observed from reading to changes in mathematics, with a near zero effect from mathematics to changes in reading. However, the results presented here are in line with the work of Duncan et al. (2007), who reported that early mathematics skills are a strong predictor of subsequent changes in reading. A better understanding of how reading may serve as a negative predictor of subsequent changes in mathematics can be hastened by considering the large amount of time students spend on acquiring literacy skills compared with mathematics skills throughout elementary school. In particular, the more time students spend on acquiring literacy skills, the less time is available for mathematics instruction. Thus, the heavy emphasis placed on early literacy skills can also have a detrimental effect on mathematics skills. Furthermore, the findings by Grimm were based on mathematics skills related to data interpretation and problem solving. However, findings not presented in Grimm suggested that reading skills had a negative effect on scores from a mathematics concepts and estimation test. Thus, reading skills may have positive effects on subsequent changes in certain aspects of mathematics where reading skills are necessary; however, reading may have negative impacts on other areas of mathematics.

Results from our simulation study showed that under the sample size configurations considered in this study, parameters from the regime-switching BDSC model can be recovered well using the EM estimation procedure in Mplus. Point estimates generally showed low biases when the differences in dynamics across regimes are pronounced and the choice of transition probabilities ensures that enough data points are available from each regime for estimation purposes. This latter finding was consistent with the simulation results reported elsewhere for regime-switching models with time series data (Yang & Chow, 2010). In terms of separation between regimes, we found that the point estimates showed satisfactory accuracy even with a relatively small multivariate Hosmer's distance (i.e., 2.33). Lower accuracy in *SE* estimation and lower coverage rates were observed, however, among some of the parameters, particularly the group average intercept and constant slope parameters, the variance-covariance parameters, and the initial class probabilities.

The proposed regime-switching BDCS model is but one possible way of representing between-person, over-time heterogeneities in dynamic parameters. Alternatively, the constant slopes, self-feedback, and coupling parameters may be freed to vary continuously over time as in other time-varying parameter models (e.g., Chow, Zu, Shifren, & Zhang, 2011; Ferrer & McArdle, 2004; Ferrer & Widaman, 2008). It would also be interesting to investigate whether process noise and/or measurement errors structure show regime-dependent properties. A three-regime model in which the cross-regression parameters were specified to be zero (an independent regime), positive (a coactivated regime), or negative (an antagonistic regime) is another interesting extension.

By allowing between-regime differences to be manifested either in the structural or the measurement model within each regime or class, MSEM-RS differs in its emphasis from another class of well-known longitudinal models of discrete changes—the hidden Markov model (Elliott, Aggoun, & Moore, 1995) or the related latent transition model (Collins & Wugalter, 1992; Lanza, Patrick, & Maggs, 2010). In MSEM-RS, latent classes can be identified using mixed observed responses (Asparouhov & Muthén; B. O. Muthén, 2001, 2002). In this way, MSEM-RS is based on a more flexible measurement model than latent transition models (Collins & Wugalter, 1992), which are typically identified using categorical manifest variables (Cho, Cohen, Kim, & Bottge, 2010; Lanza, Patrick, & Maggs, 2010). In addition, in MSEM-RS, the emphasis may be placed on between-regime differences in how latent variables are implicated in a longitudinal change process. Thus, the changes *within* each regime may be continuous in nature, even though the shifts *between* regimes or classes are discrete. In contrast, in hidden Markov models and other related variations noted earlier, no constraints are imposed on how the process may vary over time within a regime. As a result, emphasis is often placed exclusively on between-regime differences in levels (e.g., sudden shifts in means) and at times, variance–covariance components. Thus, MSEM-RS models are more suited to representing processes wherein, in addition to the progression or shifts through discrete phases or stages, the changes that unfold within stages are also of interest to the modeler.

Applications of regime-switching models in the time series and econometric literature are based largely on intensive repeated measures data (e.g., Hamilton, 1989; Kim & Nelson, 1999; Tong & Lim, 1980; Tiao & Tsay, 1994; Yang & Chow, 2010). As in other applications of regime-switching models within the structural equation modeling framework (e.g., Dolan, Schmittmann, Lubke, & Neale, 2005; Schmittmann, Dolan, van der Maas, & Neale, 2005), we evaluated a set of longitudinal panel data within the MSEM-RS framework. Consistent with the modeling conventions in the SEM framework, longitudinal processes in MSEM-RS are specified by including repeated measurement occasions of $\mathbf{y}_{it}$ and $\boldsymbol{\eta}_{it}$ in $\mathbf{y}_i$ and $\boldsymbol{\eta}_i$, respectively. When $T$ is small, this modeling framework provides great flexibility in allowing the number of latent classes to vary for each time point and for latent class membership to depend on all previous class membership information. The latter property allows each individual's class membership to manifest higher order Markov switching characteristics as opposed to first-order Markov switching as assumed in other regime-switching applications.

In longitudinal data with large $T$, SEM-based approaches are not computationally feasible, especially when $T > n$ (Chow, Ho, Hamaker, & Dolan, 2010; Hamaker, Dolan, & Molenaar, 2003, see, e.g.,). Even when $T < n$, the increase in computational costs when $T$ is large still poses some clear estimation difficulties. In particular, all exact likelihood approaches (e.g., the EM algorithm implemented in Mplus) requires the storage of the entire regime history (i.e., the regime in which each individual resides at each $t$) for estimation purposes. Storage of such regime membership information can prove computationally intractable when the number of time points involved is large.[5] One prominent feature of the estimation techniques used in the time series/state-space literature (e.g., Hamilton, 1994; Kim & Nelson, 1999) lies in the use of different procedures to "collapse" or take weighted averages of the previous regime history of a system to avoid the need to store the entire regime history. This alleviates considerable computational burden and helps make estimation involving intensive repeated measures data feasible. The SEM formulation and the associated estimation techniques do not utilize such approximations. Thus, they show better statistical properties over other approximation-based approaches when $T$ is small. For instance, when the EM approach was used, we noticed considerable gain in the precision of the transition probability parameters compared with other simulations using approximation-based approaches (e.g., Yang & Chow, 2010). However, when $T$ is large, time series and/or state-space estimation techniques offer a practical alternative.

As in any simulation study, it is important to emphasize that our simulation results depend in part on the choices of the parameter values used in the data generation process. In particular, drawing from estimates observed in the empirical study, many of the variance parameters were set to near-zero values. In addition, we used a balanced initial class sample size in our simulations. An unbalanced initial class sample size coupled with unbalanced transition probabilities would further reduce the sample size associated with the less prevalent regime, thus leading to even greater biases in estimating parameters that are unique to this regime.

We utilized a commercial SEM software package, Mplus, for the modeling work undertaken in this study. The proposed regime-switching BDCS model can also be estimated using OpenMx (Boker et al., 2011). Some examples

[5]For instance, with only two regimes and $T = 9$, there are $2^9 = 512$ ways in which the participants' regime history is manifested over time. The increase in computational time was substantial when the number of time points in our proposed regime-switching BDCS model was increased from five to nine. In the former, a single Monte Carlo replication with five sets of random starting values in the initial stage and two sets of random starting values in the second stage took approximately 1 min to run. When the number of time points was increased to nine (with missing values), a single replication with 10 random starting values in the initial stage and 3 sets of random starting values in the second stage took approximately 1.5 hr to run.

implementing similar models using the Mx software have been reported by other researchers (e.g., Dolan, Schmittmann, Lubke, & Neale, 2005; Schmittmann, Dolan, van der Maas, & Neale, 2005). Other statistical programs that can handle matrix operations, such as Matlab, R (R Development Core Team, 2009), SAS/IML (SAS Institute Inc., 2008), GAUSS (Aptech Systems Inc., 2009), and OxMetrics (Doornik, 1998), may also be used. Kim and Nelson (1999), for instance, provided some GAUSS codes for fitting the models considered in their book. The msm (Jackson, 2011) and depmix packages (Visser, 2007) in R, as well as the PROC LTA procedure in SAS (Lanza & Collins, 2008), can all be used to fit hidden Markov models and/or latent transition models, but these programs cannot be used to fit MSEM–RS wherein the means and covariance structures of each regime/class are governed by a specified structural equation model.

Despite some of the promises of MSEM-RS models, some limitations remain. As noted earlier, the increase in computational costs with increase in the number of time points poses some practical challenges. In addition, relatively high classification errors were observed when the posterior class probabilities were used for classification purposes. This is a major issue associated with using regime-switching models for forecasting purposes. That is, because the forecast values from a regime-switching model are generally regime-dependent, it only takes a small amount of errors in regime classification to yield forecast values that deviate substantially from the true trajectories (Dacco & Satchell, 1999). Thus, modifications may need to be considered when regime-switching models are used for classification and forecasting purposes.

Model identification is another key issue that warrants more attention from researchers. When the number of regimes is large or the distinctions between regimes are not pronounced, researchers may want to impose more than the necessary number of constraints to aid estimation. Appropriate constraints to avoid the well-known problem of "label switching" (McLachlan & Peel, 1995; Tueller & Lubke, 2010), particularly in simulation studies, are also helpful. When multiple regimes exist, multiple local maxima are prone to arise in the likelihood expression. To reduce the sensitivity of the estimation results to starting values, Mplus allows users to specify multiple sets of starting values and provides some simple diagnostics for checking convergence status. Currently, the amount of perturbation imposed on the modeling parameters in Mplus is governed by one control parameter. In future versions, it may be helpful to provide the users with the option of specifying different magnitudes of perturbation to different parameters.

## ACKNOWLEDGMENT

Funding for this study was provided by a grant from the National Science Foundation (BCS-0826844).

## REFERENCES


Aiken, L. R., Jr. (1971). Verbal factors and mathematics learning: A review of research. *Journal for Research in Mathematics Education, 2,* 304–313.

Aptech Systems Inc. (2009). GAUSS (Version 10) [Computer software manual]. Black Diamond, WA: Aptech Systems Inc.

Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika, 63,* 271–300.

Asparouhov, T., & Muthén, B. O. (2011). *C on C and X* (Mplus technical appendix). Los Angeles, CA: Muthén & Muthén.

Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). London, UK: Griffin.

Boker, S. M., Neale, H., Maes, H., Wilde, M., Spiegel, M., Brick, T., ... Fox, J. (2011). OpenMx: An open source extended structural equation modeling framework. *Psychometrika, 76,* 306–317.

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedulus, 133,* 59–68.

Cattell, Raymond B. (1963). The interaction of hereditary and environmental influences. *The British Journal of Statistical Psychology, 16,* 191–210.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Cho, S., Cohen, A. S., Kim, S., & Bottge, B. (2010). Latent transition analysis with a mixture item response theory measurement model. *Applied Psychological Measurement, 34,* 483–504.

Chow, S.-M., Ho, M.-H. R., Hamaker, E. J., & Dolan, C. V. (2010). Equivalences and differences between structural equation and state-space modeling frameworks. *Structural Equation Modeling, 17,* 303–332.

Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research, 46,* 303–339.

Collins, L. M., & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 28* 131–157.

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology, 20,* 405–438.

Dacco, R., & Satchell, S. (1999). Why do regime-switching models forecast so badly? *Journal of Forecasting, 18,* 1–16.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39,* 1–38.

Dolan, C. V. (2009). Structural equation mixture modeling. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 568–592). Thousand Oaks, CA: Sage.

Dolan, C. V., Jansen, B. R., & van der Maas, H. L. J. (2004). Constrained and unconstrained multivariate normal finite mixture modeling of piagetian data. *Multivariate Behaviorial Research, 39,* 69–98.

Dolan, C. V., Schmittmann, V. D., Lubke, G. H., & Neale, M. C. (2005). Regime switching in the latent growth curve mixture model. *Structural Equation Modeling, 12,* 94–119.

Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63,* 227—253.

Doornik, J. A. (1998). *Object-oriented matrix programming using Ox 2.0*. London, UK: Timberlake Consultants Press.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Early, D. M., Iruka, I. U., Ritchie, S., Barbarin, O. A., Winn, D. C., Crawford, G. M., ... Pianta, R. C. (2010). How do pre-kindergarteners spend their time? Gender, ethnicity, and income as predictors of experiences in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 25,* 177–193.

Elliott, R. J., Aggoun, L., & Moore, J. B. (1995). *Hidden markov models: Estimation and control*. New York, NY: Springer.

Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London, UK: Chapman & Hall.

Ferrer, E., & McArdle, J. J. (2004). An experimental analysis of dynamic hypotheses about cognitive abilities and achievement from childhood to early adulthood. *Developmental Psychology, 40,* 935–952.

Ferrer, E., & Widaman, K. F. (2008). Dynamic factor analysis of dyadic affective processes with intergroup differences. In N. A., Card, J. Selig, & T. Little (Eds.), *Modeling dyadic and interdependent data in the developmental and behavioral sciences* (pp. 107–138). New York, NY: Routledge.

Fukuda, K., & Ishihara, K. (1997). Development of human sleep and wakefulness rhythm during the first six months of life: Discontinuous changes at the 7th and 12th week after birth. *Biological Rhythm Research, 28,* 94–103.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *Trends in Cognitive Sciences, 9,* 6–10.

Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika, 24,* 229–252.

Goldberger, A. S., & Duncan, O. D. (Eds.). (1973). *Structural equation models in the social sciences.* New York, NY: Academic Press.

Grimm, K. J. (2006). *A longitudinal dynamic analysis of the impacts of reading on mathematical ability in children and adolescents* (Unpublished doctoral dissertation). Charlottesville, VA: University of Virginia.

Grimm, K. J., & Ram, N. (2011). Growth curve modeling from an SEM perspective. In T. Little, B. Laursen, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 411–431). New York, NY: Guilford Press.

Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in Mplus and OpenMx. *Multivariate Behavioral Research, 45,* 887–909.

Hamaker, E. L., Dolan, C. V., & Molenaar, P. C. M. (2003). ARMA-based SEM when the number of time points *T* exceeds the number of cases *N*: Raw data maximum likelihood. *Structural Equation Modeling, 10,* 352–379.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57,* 357–384.

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

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal, 14,* 202–226. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10705510709336744. doi: 10.1080/10705510709336744

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57,* 253–270.

Hosenfeld, B. (1997). Indicators of discontinuous change in the development of analogical reasoning. *Journal of Experimental Child Psychology, 64,* 367–395.

Hosmer, D. W. (1974). Maximum likelihood estimates of parameters of a mixture of two regression lines. *Communications in Statistics: Theory and Methods, 3,* 995–1006.

Jackson, C. H. (2011). Multi-state models for panel data: The msm package for R. *Journal of Statistical Software, 38,* 1–29. Retrieved from http://www.jstatsoft.org/v38/i08/

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997a). Stemm: A general finite mixture structural equation model. *Journal of Classification, 14,* 23–50. Retrieved from http://dx.doi.org/10.1007/s003579900002. doi: 10.1007/s003579900002

Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997b). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16,* 39–59.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology, 44,* 457–467.

Kim, C.-J., & Nelson, C. R. (1999). State-space models with regime switching: Classical and Gibbs-sampling approaches with applications. Cambridge, MA: MIT Press.
Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development, 12,* 3–120.
Lanza, S. T., & Collins, L. M. (2008). A new SAS procedure for latent transition analysis: Transitions in dating and sexual risk behavior. *Developmental Psychology, 44,* 446–456.
Lanza, S. T., Patrick, M. E., & Maggs, J. L. (2010). Latent transition analysis: Benefits of a latent variable approach to modeling transitions in substance use. *Journal of Drug Issues, 40,* 93–120.
Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston, MA: Houghton Mifflin.
Lubke, G. H. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21–39.
Lubke, G. H., & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14,* 26–47.
Marsh, H. W. (1989). Sex differences in the development of verbal and mathematics constructs: The high school and beyond study. *American Educational Research Journal, 26*, 191–225.
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). Heidelberg, Germany: Springer-Verlag.
McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analysis with incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 139–175). Washington, DC: American Psychological Association.
McLachlan, G., & Peel, D. (1995). *Finite mixture models*. New York, NY: Wiley.
Mercer, J. (1998). *Infant development: A multidisciplinary introduction*. Pacific Grove, CA: Brooks/Cole.
Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. New York, NY: Oxford University Press.
Monroe, W. S., & Englehart, M. D. (1931). *A critical summary of research relating to the teaching of arithmetic* (Vol. 58). Urbana, IL: University of Illinois.
Muscio, R. D. (1962). Factors related to quantitative understanding in the sixth grade. *Arithmetic Teacher, 9,* 258–262.
Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumaker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–33). Mahwah, NJ: Erlbaum.
Muthén, Bengt O., Beyond SEM: General latent variable modeling. *Behaviormetrika, 29,* 81–117.
Muthén, B. O., & Asparouhov, T. (2011, July). *LTA in Mplus: Transition probabilities influenced by covariates* (Mplus Web Notes: No. 13). Los Angeles, CA: Muthén & Muthén.
Muthén, B. O., & Shedden, K. (1992). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics, 55,* 463–469.
Muthén, L. K., & Muthén, B. O. (2001). *Mplus: The comprehensive modeling program for applied researchers. User's guide*. Los Angeles, CA: Author.
Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology, 31,* 327–362.
National Institute of Child Health and Human Development Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Elementary School Journal, 105,* 305–323.
Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569. Retrieved from http://www.scopus.com/record/display.url?view=extended&origin=resultslist&eid=2-s2.0-36849091981
Nylund-Gibson, K., Muthén, B., Nishina, A., Bellmore, A., & Graham, S. (under review). Stability and instability of peer victimization during middle school: Using latent transition analysis with covariates, distal outcomes, and modeling extensions.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York, NY: Basic Books.
Pianta, R. C., Belsky, J., Houts, R., & Morrison, F. (2007). Opportunities to learn in America's elementary classrooms. *Science, 315,* 1795–1796.
R Development Core Team. (2009). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org (ISBN 3-900051-07-0)
SAS Institute Inc. (2008). SAS 9.2 Help and Documentation [Computer software manual]. Cary, NC: Author.
Schmittmann, V. D., Dolan, C. V., van der Maas, H., & Neale, M. C. (2005). Discrete latent Markov models for normally distributed response data. *Multivariate Behavioral Research, 40,* 207–233.
Skaalvik, E. M., & Rankin, R. J. (1990). Math, verbal, and general academic self-concept: The internal/external frame of reference model and gender differences in self-concept structure. *Journal of Educational Psychology, 82,* 546–554.
Snow, C. E., Burns, S. M., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
Tiao, G. C., & Tsay, R. S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting, 13,* 109–131.
Titterington, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, UK: Wiley.
Tong, H., & Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B, 42,* 245–292.
Tueller, S., & Lubke, G. (2010). Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling, 17,* 165–192. doi: 10.1080/10705511003659318
U.S. Department of Education, Office of the Deputy Secretary. (2004). *No child left behind: A toolkit for teachers*. Washington, DC.
U.S. Department of Education, National Center for Education Statistics. (2010). *Early childhood longitudinal study, kindergarten class of 1998–99 (ECLS-K) kindergarten through fifth grade approaches to learning and self-description questionnaire (SDQ) items and public-use data files (nces 2010-070).* Washington, DC: Author.
Van der Maas, H. L. J., Dolan, C. V., Grasman, R., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review, 113,* 842–861.
Van der Maas, H. L. J., & Molenaar, P. C. M. (1992). Stagewise cognitive development: An application of catastrophe theory. *Psychological Review, 99,* 395–417.
Van Dijk, M., & Van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development, 16,* 7–33.
Vermunt, J. K., & Magidson, J. (2005). Structural equation models: Mixtrue models. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1922–1927). Chichester, UK: Wiley.
Visser, I. (2007). *Depmix: An R-package for fitting mixture models on mixed multivariate data with Markov dependencies* (Tech. Rep.). University of Amsterdam. Retrieved from http://cran.r-project.org
Wrigley, C., Saunders, D., & Neuhaus, J. O. (1958). Applications of the quartimax method of rotation to Thurstone's priamry mental abilities study. *Pyschometricka, 23,* 151–168.
Yang, M., & Chow, S.-M. (2010). Using state-space model with regime switching to represent the dynamics of facial electromyography (emg) data. *Psychometrika, 74,* 744–771.
Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika, 62,* 297–330.