PSYCHOMETRIKA—VOL. 78, NO. 4, 740–768
OCTOBER 2013
DOI: 10.1007/S11336-013-9330-8

NONLINEAR REGIME-SWITCHING STATE-SPACE (RSSS) MODELS

SY-MIIN CHOW

THE PENNSYLVANIA STATE UNIVERSITY

GUANGJIAN ZHANG

UNIVERSITY OF NOTRE DAME

Nonlinear dynamic factor analysis models extend standard linear dynamic factor analysis models by allowing time series processes to be nonlinear at the latent level (e.g., involving interaction between two latent processes). In practice, it is often of interest to identify the phases—namely, latent "regimes" or classes—during which a system is characterized by distinctly different dynamics. We propose a new class of models, termed nonlinear regime-switching state-space (RSSS) models, which subsumes regime-switching nonlinear dynamic factor analysis models as a special case. In nonlinear RSSS models, the change processes within regimes, represented using a state-space model, are allowed to be nonlinear. An estimation procedure obtained by combining the extended Kalman filter and the Kim filter is proposed as a way to estimate nonlinear RSSS models. We illustrate the utility of nonlinear RSSS models by fitting a nonlinear dynamic factor analysis model with regime-specific cross-regression parameters to a set of experience sampling affect data. The parallels between nonlinear RSSS models and other well-known discrete change models in the literature are discussed briefly.

Key words: regime-switching, state-space, nonlinear latent variable models, dynamic factor analysis, Kim filter.

1. Nonlinear Regime-Switching State-Space (RSSS) Models

Factor analysis is widely recognized as one of the most important methodological developments in the history of psychometrics. By combining factor analysis and time series analysis, dynamic factor analysis models are one of the better known models of intensive multivariate change processes in the psychometric literature (Browne & Nesselroade, 2005; Engle & Watson, 1981; Geweke & Singleton, 1981; Molenaar, 1985; Nesselroade, McArdle, Aggen, & Meyers, 2002). They have been used to study a broad array of change processes (Chow, Nesselroade, Shifren, & McArdle, 2004; Ferrer & Nesselroade, 2003; Molenaar, 1994a; Sbarra & Ferrer, 2006). Parallel to the increased use of dynamic factor analysis models in substantive applications, various methodological advancements have also been proposed over the last two decades for fitting linear dynamic factor analysis models to continuous as well as categorical data (Molenaar, 1985; Molenaar & Nesselroade, 1998; Browne & Zhang, 2007; Engle & Watson, 1981; Zhang & Nesselroade, 2007; Zhang, Hamaker, & Nesselroade, 2008).

In the present article, we propose a new class of models, termed nonlinear regime-switching state-space (RSSS) models, which subsumes dynamic factor analysis models that show *linear or nonlinear* dynamics at the latent level as a special case. The term "regime-switching" refers to the property that individuals' change mechanisms are contingent on the latent class or "regime" they are in at a particular time point. In addition, individuals are allowed to transition between

Requests for reprints should be sent to Sy-Miin Chow, The Pennsylvania State University, 422 Biobehavioral Health Building, University Park, PA 16801, USA. E-mail: symiin@psu.edu

classes or regimes over time (Hamilton, 1989; Kim & Nelson, 1999). Thus, as in many stagewise theories, change processes are conceptualized as a series of discontinuous progressions through distinct, categorical phases. Many examples of such processes arise in the study of human developmental processes (e.g., Piaget & Inhelder, 1969; Van der Maas & Molenaar, 1992; Fukuda & Ishihara, 1997; Van Dijk & Van Geert, 2007).

Most stagewise theories dictate that the transition between stages unfolds in a unidirectional manner, namely, the transition has to occur from one regime to a later regime in a sequential manner; transition in the reverse direction is considered as rare or not allowed. As distinct from conventional stagewise theories, regime-switching models provide a way to represent change trajectories that unfold continuously within each stage, as well as *how* individuals progress through stages. For instance, Hamilton (1989) proposed a regime-dependent autoregressive model wherein the transition between regimes is modeled as a first-order Markov-switching process. 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), past values of a system (e.g., as in self-exciting threshold autoregressive models; Tiao & Tsay, 1994), and other external covariates of interest (Muthén & Asparouhov, 2011). Thus, such models enrich conventional ways of conceptualizing stagewise processes by offering more ways to represent changes *within* as well as *between* stages. Despite their promises, most of the existing frequentist approaches to fitting regime-switching models assume that the underlying dynamic processes are linear in nature (e.g., Dolan, 2009; Hamilton, 1989; Kim & Nelson, 1999; Muthén & Asparouhov, 2011).

In cases involving nonlinear dynamic factor analysis models, the dynamic processes involved may be nonlinear (Chow, Zu, Shifren, & Zhang, 2011b; Molenaar, 1994b), as well as showing regime-switching properties. Our empirical example describes one such instance in modeling individuals' affective dynamics. Although several approaches have been proposed in the structural equation modeling framework for fitting nonlinear latent variable models (Kenny & Judd, 1984; Marsh, Wen, & Hau, 2004; Schumacker & Marcoulides, 1998), extending these approaches for use with longitudinal, regime-switching processes is not always practical. Due to the way that repeated measurement occasions of the same variable are incorporated as different variables in a structural equation model (SEM), the parameter constraints a researcher has to specify in fitting nonlinear longitudinal SEMs data can be extremely cumbersome (see, e.g., Li, Duncan, & Acock, 2000; Wen, Marsh, & Hau, 2002). When the number of time points exceeds the number of participants in the data set, structural equation modeling–based approaches cannot even be used (Chow, Ho, Hamaker, & Dolan, 2010; Hamaker, Dolan, & Molenaar, 2003). By combining the linearization procedures from the extended Kalman filter (Anderson & Moore, 1979; Molenaar & Newell, 2003) and the Kim filter for estimating *linear* RSSS models (Kim & Nelson, 1999), a new estimation approach, referred to herein as the extended Kim filter, is proposed as an approach for handling parameter and latent variable estimation in our proposed nonlinear RSSS models.

The remainder of the article is organized as follows. We first describe an empirical example that motivated us to develop the nonlinear RSSS estimation technique described in the present article. We then introduce the broader modeling framework that is suited for handling other modeling extensions similar to the model considered in our motivating example. The associated estimation procedures are then outlined. This is followed by a summary of the results from empirical model fitting, as well as a Monte Carlo simulation study. We conclude with some remarks on the strengths and limitations of the proposed technique.

2. Motivating Example

The motivating example in this article was inspired by the Dynamic Model of Activation proposed by Zautra and colleagues (Zautra, Potter, & Reich, 1997; Zautra, Reich, Davis, Potter, & Nicolson, 2000). This model posits that the concurrent association between positive affect (PA) and negative affect (NA) changes over time and context as a function of a time-varying covariate, namely, activation level. In particular, PA and NA are posited to be independent under low activation (e.g., low stress) conditions, but they have been reported to collapse into a unidimensional, bipolar structure under high activation (Zautra et al., 2000). Chow and colleagues (Chow, Tang, Yuan, Song, & Zhu, 2011a; Chow et al., 2011b) demonstrated that, by representing the changes in reciprocal PA-NA linkages as part of a nonlinear dynamic factor analysis model, researchers can further disentangle the directionality of the PA-NA linkage. Specifically, they studied how the lagged influences from PA to NA, as well as those from NA to PA, might change through over-time fluctuations in the *cross-regression parameters*. The resultant model differs from other existing dynamic factor analysis models in the literature (e.g., Browne & Nesselroade, 2005; Nesselroade et al., 2002) in that the dynamic functions that characterize the changes among factors are allowed to be nonlinear.

As in the earlier models considered by Chow and colleagues (Chow et al., 2011a, 2011b), the proposed regime-switching model offers a refinement of Zautra and colleagues' model by providing insights into whether the changes in association between PA and NA are driven more by fluctuations in PA or in NA. The proposed model is distinct from other earlier models, however, in that instances on which individuals show "high-activation" versus "independent" structure of emotions are regarded as two distinct phases of an individual's affective process. That is, we hypothesize that two major regimes characterize the various forms of PA-NA linkage posited in the Dynamic Model of Activation: (1) an "*independent*" regime captures instances on which the lagged influences between PA and NA are zero; (2) a "*high-activation*" regime reflects instances on which the lagged influences between PA and NA intensify when an individual's previous levels of PA and NA were unusually high or low. The resultant illustrative dynamic model is written as

$$\begin{aligned} PA_{it} &= a_P PA_{i,t-1} + b_{PN,S_{it}} NA_{i,t-1} + \zeta_{PA,it}, \\ NA_{it} &= a_N NA_{i,t-1} + b_{NP,S_{it}} PA_{i,t-1} + \zeta_{NA,it}, \end{aligned} \tag{1}$$

$$\begin{aligned} b_{PN,S_{it}} &= \begin{cases} 0 & \text{if } S_{it} = 0, \\ b_{PN0}\left(\frac{\exp(\text{abs}(NA_{i,t-1}))}{1+\exp(\text{abs}(NA_{i,t-1}))}\right) & \text{if } S_{it} = 1, \end{cases} \\ b_{NP,S_{it}} &= \begin{cases} 0 & \text{if } S_{it} = 0, \\ b_{NP0}\left(\frac{\exp(\text{abs}(PA_{i,t-1}))}{1+\exp(\text{abs}(PA_{i,t-1}))}\right) & \text{if } S_{it} = 1, \end{cases} \end{aligned} \tag{2}$$

where abs(.) denotes the absolute function, PA_{it} and NA_{it} correspond to person i's PA and NA factor score at time t, respectively; $b_{PN,S_{it}}$ is the regime-dependent lag-1 NA $\rightarrow$ PA cross-regression weight and $b_{NP,S_{it}}$ is the corresponding regime-dependent lag-1 PA $\rightarrow$ NA cross-regression weight.[1] Within the "independent regime" (i.e., when $S_{it} = 0$), the cross-regression parameters linking PA and NA were constrained to be zero so that yesterday's PA (or NA) has

[1]Note that although the two lag-1 cross-regression parameters, $b_{PN,S_{it}}$ and $b_{NP,S_{it}}$, were allowed to vary over time and could be modeled as latent variables as in Chow et al. (2011b), it was not necessary to do so here because these two parameters did not have their own process noise components. Thus, the model comprises only two latent variables, namely, PA_{it} and NA_{it}. However, as in Chow et al. (2011a), the logistic functions still render the dynamic model nonlinear in PA_{it} and NA_{it}.

no impact on today's NA (or PA). In the "high-activation regime" (i.e., when $S_{it} = 1$), the full deviations in PA $\rightarrow$ NA and NA $\rightarrow$ PA cross-regression parameters from zero are given by b_{NP0} and b_{PN0}, respectively. Such cross-regression effects are fully manifested only when the previous level of PA or NA was extreme—either extremely high or extremely low. The latter feature is reflected by the use of the absolute function, abs(.) in Equation (2).[2] One implication of the specification in Equation (2) is that each individual is allowed to have his/her own cross-regression weights, $b_{PN,S_{it}}$ and $b_{NP,S_{it}}$ that are also allowed to vary *over time*, as governed by the operating regime at a particular time point.

We assume that the shock variables or process noise components in $\boldsymbol{\zeta}_{it} = [\zeta_{PA,it}\zeta_{NA,it}]'$ are distributed in both regimes as

$$\boldsymbol{\zeta}_{it} \sim \text{MVN}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\sigma^2_{\zeta_{PA}} & \\ \sigma_{\zeta_{PA},\zeta_{NA}} & \sigma^2_{\zeta_{NA}}\end{bmatrix}\right), \tag{3}$$

where $\text{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. That is, we assume that the process noise components have the same distribution across regimes. In addition, the same measurement model is assumed across regimes, with

$$\boldsymbol{y}_{it} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{it} + \boldsymbol{\epsilon}_{it}, \quad \text{and} \quad \boldsymbol{\epsilon}_{it} \sim N(\mathbf{0}, \mathbf{R}), \tag{4}$$

where $\boldsymbol{\eta}_{it} = [PA_{it}\ NA_{it}]'$ includes the unobserved PA and NA factor scores for person i at time t, $\boldsymbol{y}_{it}$ is a vector of observed variables used to indicate these latent factors, $\boldsymbol{\Lambda}$ is the factor loading matrix and $\boldsymbol{\epsilon}_{it}$ is the corresponding vector of unique variables. All parameters in the measurement equation in (4) are constrained to be invariant across individuals.

The model depicted in Equations (1–4) are nonlinear in the dynamic functions within regime. As distinct from conventional single-subject time series analysis, we "borrow strengths" from all individuals' data in estimating the person-specific cross-regression weights by constraining seven additional time series parameters to be equal across persons. These parameters include (1) the AR(1) parameters for PA and NA, a_P and a_N, (2) the full NA $\rightarrow$ PA and PA $\rightarrow$ NA cross-regression weights in the high-activation regime, b_{PN0} and b_{NP0}, and (3) the process noise variance and covariance parameters, $\sigma^2_{\zeta_{PA}}$, $\sigma^2_{\zeta_{NA}}$, and $\sigma_{\zeta_{PA},\zeta_{NA}}$. Consistent with conventions in the time series modeling literature, we assume in this motivating example that the data used for model fitting have been demeaned and detrended so there is no intercept term in Equation (1) or (4) and no other systematic trends are present in the data. Alternatively, intercept terms can be added either to the measurement equation in (4) or Equation (1) can be modified to represent deviations in PA and NA from their nonzero (as opposed to zero) equilibrium points.

A transition probability matrix is then used to specify the probability that an individual is in a certain regime conditional on the previous regime. In matrix form, these transition probabilities, which are constrained within the present context to be invariant across people, are written as

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \tag{5}$$

where the jth, kth element of $\mathbf{P}$, denoted as p_{jk} $(j, k = 1, 2, \ldots, M)$, represents the probability of transitioning from regime j at time $t-1$ to regime k at time t for person i. For instance, p_{11} and

[2]It may be worth mentioning that Equations (1–2) differ from the model considered in Chow et al. (2011a) in a number of ways. For instance, Chow et al. (2011a) did not use the absolute function in Equation (2) and allowed the cross-lagged dependencies to materialize in the AR, as opposed to the cross-regression parameters, at extremely *high* levels of PA and NA from the previous day. They also allowed for random effects in some of the time series parameters, all of which are hypothesized to conform to nonparametric distributions modeled within a Bayesian framework.

p_{22} represent the probability of staying within the independent regime and high-activation regime from time $t-1$ to time t, respectively. Depending on a researcher's model specification, the transition between two regimes can be unidirectional or bidirectional in nature. In unidirectional transitions, an individual may be allowed to switch from regime 1 to regime 2, but not the other way round. This is implemented by freeing p_{12} and setting p_{21} to zero. When the transition is bidirectional in nature, an individual is allowed to transition from regime 1 to regime 2 and vice versa, implemented by freeing both p_{12} and p_{21}.

To help shed light on properties of the proposed model, we generated some hypothetical trajectories using the model shown in Equations (1–2). Figure 1 depicts examples of individuals who always stay within one of the two regimes (see Panels A and B), or show a mixture of trajectories from both regimes (see Panel C). All trajectories were specified to start with the same levels of high NA and low PA at $t=1$, and they were all subjected to influences from the same series of random shocks, $\boldsymbol{\zeta}_{it}$.

At the specific parameter values used for this simulation, the effects of the initial random shock at $t=1$ can be seen to decay exponentially over time at different rates in Figure 1, Panels A–C, when only minimal new random shocks are added between $t=2$ and $t<10$. While in the independent regime, PA and NA, with autoregression weights that are less than 1.0 in absolute value and minimal new random shocks prior to $t=10$, approach their respective equilibrium points at zero regardless of the level of the other emotion. In contrast, while in the high-activation regime, the negative cross-regression weights from PA to NA and from NA to PA propel the two emotions to maintain divergent levels as they approach their equilibrium points. That is, when NA is unusually high, PA is unusually low. Compared to the PA and NA trajectories in the independent regime, the return to the equilibrium points unfolds over a longer period while the individual is staying in the high-activation regime. This is because unusually high deviations in one affect lead to high deviations in the other affect, but in the opposite direction (Figure 1, Panel B). As the magnitudes of the process noise variances are increased from $t=10$ and beyond, the new random shocks give rise to ebbs and flows of varying magnitudes that are always manifested in opposing directions. The decay rate is attenuated—that is, the coupling becomes stronger—on days that are preceded by unusually high PA or NA. Figure 1, Panel C, shows the corresponding regulatory trajectories when the individual has a probability of 0.6 and 0.6 of staying within the independent and high activation regime, respectively.

3. Nonlinear Regime-Switching State-Space Models

In this section, we discuss the broader nonlinear regime-switching modeling framework within which the model shown in the motivating example can be structured as a special case. Our general modeling framework can be expressed as

$$\boldsymbol{y}_{it} = \mathbf{d}_{S_{it}} + \boldsymbol{\Lambda}_{S_{it}}\boldsymbol{\eta}_{it} + \mathbf{A}_{S_{it}}\boldsymbol{x}_{it} + \boldsymbol{\epsilon}_{it}, \tag{6}$$

$$\boldsymbol{\eta}_{it} = \mathbf{b}_{S_{it}}(\boldsymbol{\eta}_{i,t-1}, \boldsymbol{x}_{it}) + \boldsymbol{\zeta}_{it}, \tag{7}$$

$$\begin{bmatrix} \boldsymbol{\epsilon}_{it} \\ \boldsymbol{\zeta}_{it} \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{R}_{S_{it}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{S_{it}} \end{bmatrix}\right),$$

where i indexes person and t indexes time. S_{it} is discrete-valued regime indicator that is latent (i.e., unknown) and has to be estimated from the data. The term $\boldsymbol{y}_{it}$ is a $p \times 1$ vector of observed variables at time t, $\boldsymbol{\eta}_{it}$ is a $w \times 1$ vector of unobserved latent variables, $\boldsymbol{x}_{it}$ is a vector of known time-varying covariates which may affect the dynamic and/or measurement functions, $\mathbf{A}_{S_{it}}$ is a matrix of regression weights for the covariates, $\boldsymbol{\Lambda}_{S_{it}}$ is a $p \times w$ factor loading matrix that links

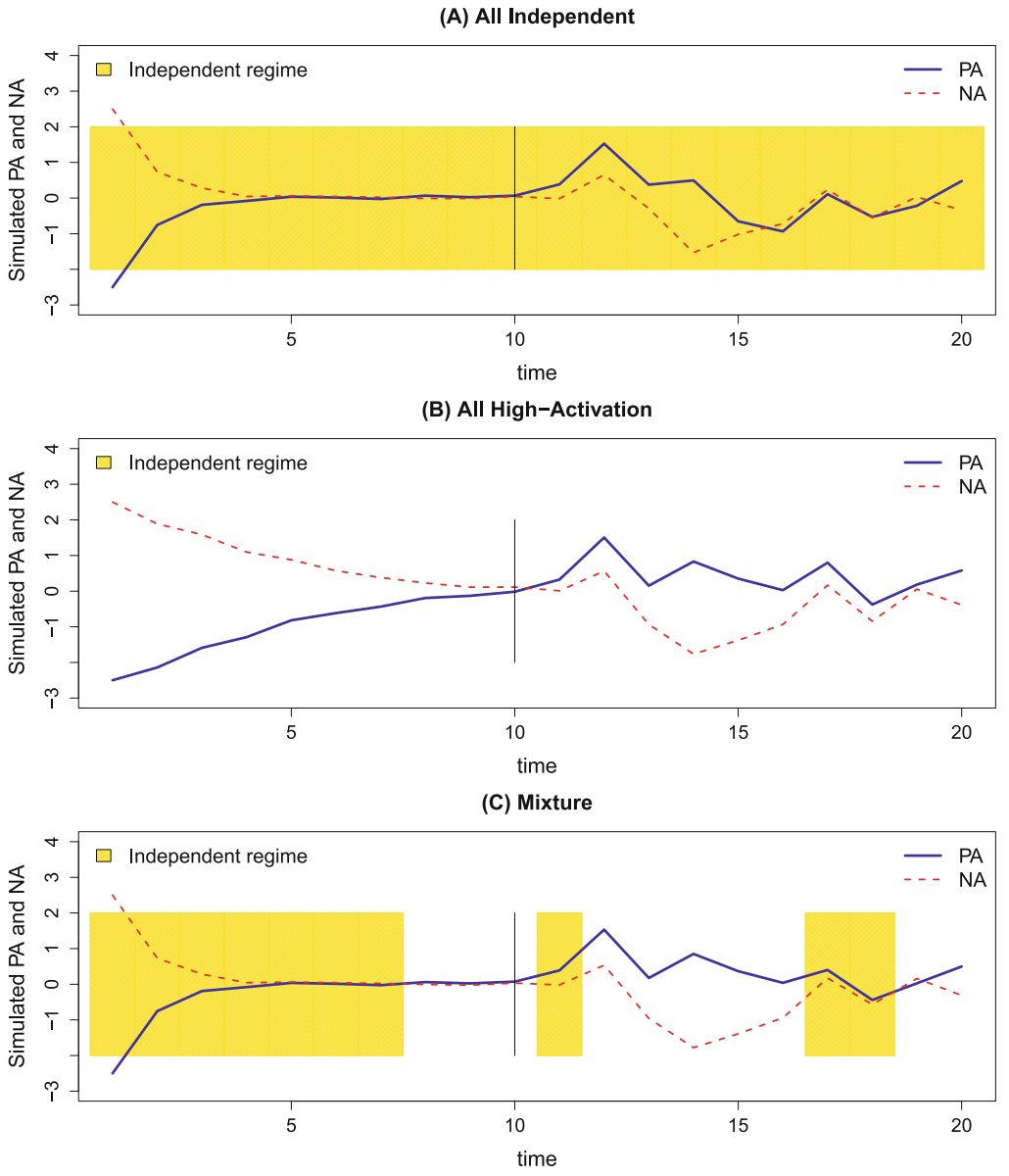

FIGURE 1.

Hypothetical trajectories of PA and NA generated using the proposed regime-switching model, with $a_P = 0.3$, $a_N = 0.3$, $b_{PN0} = -0.6$ and $b_{NP0} = -0.5$. In all scenarios, the equilibrium points of PA and NA are located at zero and the same sequence of random shocks is added to all series. For $t \leq 10$, the process noise variances were set to be very small ($\sigma^2_{\zeta_{PA}} = \sigma^2_{\zeta_{NA}} = 0.001$, $\sigma_{\zeta_{PA},\zeta_{NA}} = 0$); for $t > 10$, the process noise variances were increased to $\sigma^2_{\zeta_{PA}} = \sigma^2_{\zeta_{NA}} = 0.5$. The plots illustrate scenarios where (**A**) all data came from the independent regime; (**B**) all data came from the high-activation regime; and (**C**) the data came from both of the regimes.

the observed variables to the latent variables, and $\mathbf{d}_{S_{it}}$ is $w \times 1$ vector of intercepts. The term $\mathbf{b}_{S_{it}}(.)$ is a $w \times 1$ vector of differentiable (linear or nonlinear) dynamic functions that describe the values of $\boldsymbol{\eta}_{it}$ at time t as related to $\boldsymbol{\eta}_{i,t-1}$ and $\boldsymbol{x}_{it}$; $\boldsymbol{\epsilon}_{it}$ and $\boldsymbol{\zeta}_{it}$ are measurement errors and random shocks (or process noise) assumed to be serially uncorrelated over time and normally

distributed with a mean vector of zeros and regime-dependent covariance matrix, $\mathbf{R}_{S_{it}}$ and $\mathbf{Q}_{S_{it}}$, respectively.

Equations (6) and (7) are the *measurement equation* and *dynamic equation* of the system, respectively. The former describes the relationship between a set of observed variables and a set of latent variables over time. The latter portrays the ways in which the latent variables change over time. The subscript S_{it} associated with $\mathbf{b}_{S_{it}}(.)$, $\mathbf{d}_{S_{it}}$, $\mathbf{\Lambda}_{S_{it}}$, $\mathbf{A}_{S_{it}}$, $\mathbf{R}_{S_{it}}$, and $\mathbf{Q}_{S_{it}}$ indicates that the values of some of the parameters in them may depend on S_{it}, the operating regime for individual i at time t. In practice, not all of these elements are free to vary by regime. Our motivating example illustrated one example of such constrained models. Other than allowing some parameters to differ in values conditional on the (over-person and over-time) changes in S_{it}, the model is used to describe the dynamics of multiple subjects at the group level. Thus, when there is only one regime, these parameters do not differ in value over individuals or time.

When there is no regime dependency, Equations (6) and (7) can be conceived as a nonlinear state-space model with Gaussian distributed measurement and dynamic errors. Alternatively, these equations may be viewed as a nonlinear dynamic factor analysis model, namely, a model that combines a factor analytic model (Equation (6)) and a nonlinear time series model of the latent factors (Equation 7).

To make inferences on S_{it}, it is essential to specify a transition probability matrix. A first-order Markov process is assumed to govern the transition probability patterns, with

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix} \tag{8}$$

where the jth, kth element of $\mathbf{P}$, denoted as p_{jk} $(j, k = 1, 2, \ldots, M)$, represents the probability of transitioning from regime j at time $t-1$ to regime k at time t for person i, or $Pr[S_{it} = k | S_{i,t-1} = j]$, with the constraint that $\sum_{k=1}^{M} p_{jk} = 1$ (Kim & Nelson, 1999). These transition probabilities are regarded as model parameters that are to be estimated with other parameters that appear in $\mathbf{d}_{S_{it}}$, $\mathbf{\Lambda}_{S_{it}}$, $\mathbf{b}_{S_{it}}$, $\mathbf{A}_{S_{it}}$, $\mathbf{R}_{S_{it}}$ and $\mathbf{Q}_{S_{it}}$ in Equations (6) and (7).

4. Estimation Procedures

When the regime-switching model of interest consists only of linear equations, an estimation procedure known as the Kim filter, or the related Kim smoother (Kim & Nelson, 1999), can be used for estimation purposes. When nonlinearities are present either in Equation (6) or (7), one of the simplest approaches of handling such nonlinearities is to linearize the nonlinear equations via Taylor series approximation. The resultant estimation procedures, referred to herein as the *extended Kim filter* and *extended Kim smoother*, are outlined briefly here and described in more detail in the Appendix. The proposed estimation procedure allows each individual to have a different number of total time points, as is the case in our empirical example. However, for ease of presentation, we omit person index from T in our subsequent descriptions. In addition, all latent variable estimates are inherently conditional on the parameter vector, θ, but this dependency is omitted to simplify notations.

The extended Kim filter presented here is essentially an estimation procedure that combines the traditional extended Kalman filter (Anderson & Moore, 1979) and the Hamilton filter (Hamilton, 1989). For descriptions of the Kim filter, which combines the linear Kalman filter and the Hamilton filter, readers are referred to Kim and Nelson (1999). The Kim filter provides a

way to derive estimates of the latent variables in $\boldsymbol{\eta}_{it}$ based on both current and previous regimes and all the manifest observations from $t = 1$ to t, denoted herein as $\mathbf{Y}_{it}$. That is, we obtain $\boldsymbol{\eta}_{i,t|t}^{j,k} \triangleq E[\boldsymbol{\eta}_{it}|S_{i,t-1} = j, S_{it} = k, \mathbf{Y}_{it}]$, as well as $\mathbf{P}_{i,t|t}^{j,k} \triangleq \text{Cov}[\boldsymbol{\eta}_{it}|S_{i,t-1} = j, S_{it} = k, \mathbf{Y}_{it}]$. In contrast, the Hamilton filter offers a way to update the probability of being in the kth regime at time t conditional on manifest observations up to time t (i.e., $Pr[S_{it} = k|\mathbf{Y}_{it}]$).

The extended Kim filter can be implemented in three sequential steps. First, the extended Kalman filter is executed to yield $\boldsymbol{\eta}_{i,t|t}^{j,k}$, and their covariance matrix, $\mathbf{P}_{i,t|t}^{j,k}$. Next, the Hamilton filter is implemented to get the conditional joint regime probability of being in the jth and k regime at, respectively, time $t-1$ and time t, namely, $Pr[S_{i,t-1} = j, S_{it} = k|\mathbf{Y}_{it}]$, as well as the probability of being in the kth regime at time t, namely, $Pr[S_{it} = k|\mathbf{Y}_{it}]$. Third, a "collapsing process" is carried out to compute the estimates, $\boldsymbol{\eta}_{i,t|t}^{k}$ (i.e., $E[\boldsymbol{\eta}_{it}|S_{it} = k, \mathbf{Y}_{it}]$), and the associated covariance matrix, $\mathbf{P}_{i,t|t}^{k}$, by taking weighted averages of the $M \times M$ sets of latent variable estimates $\boldsymbol{\eta}_{i,t|t}^{j,k}$ and their associated covariance matrices, $\mathbf{P}_{i,t|t}^{j,k}$ (with $j = 1, \ldots, M$, $k = 1, \ldots, M$), prior to performing estimation for the next time point. This collapsing process reduces the need to store M^2 new values of $\boldsymbol{\eta}_{i,t|t}^{j,k}$ and $\mathbf{P}_{i,t|t}^{j,k}$ at each time point to just M sets of new marginal estimates, $\boldsymbol{\eta}_{i,t|t}^{k}$ and $\mathbf{P}_{i,t|t}^{k}$. As explained in Kim and Nelson (1999), this collapsing procedure only yields an approximation of $\boldsymbol{\eta}_{i,t|t}^{k}$ and $\mathbf{P}_{i,t|t}^{k}$ due to the truncation of terms that were omitted in the collapsing procedure in previous time points ($t = 1, \ldots, t-2$). The estimation process is performed sequentially for each time point for $t = 1$ to T.

Under normality assumptions of the measurement and process noise components and linearity of the measurement equation in (6), the prediction errors, $\boldsymbol{v}_t^{j,k}$, which capture the discrepancies between the manifest observations and the predictions implied by the model, are multivariate normally distributed. This yields a log-likelihood function, also known as the *prediction error decomposition* function, that can be computed using by-products from the extended Kim filter (i.e., see explanations accompanying Equation (A.8) in the Appendix). This prediction error decomposition function can then be optimized to yield estimates of all the time-invariant parameters in $\boldsymbol{\theta}$, as well as to construct fit indices such as the Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978). However, the resultant estimates are only "approximate" maximum likelihood (ML) estimates in the present context for two reasons. First, as in linear RSSS models, the Kim filter only yields approximate latent variable estimates due to the use of the collapsing procedure to ease computational burden (Kim & Nelson, 1999). Second, in fitting nonlinear RSSS models, additional approximation errors are induced by the truncation errors stemming from the use of only first-order terms in the Taylor series expansion in the extended Kalman filter.

If the entire time series of observations is available for estimation purposes—as in the case in most studies in psychology, one can refine the latent variable estimates for $\boldsymbol{\eta}_{it}$ and the probability of the unobserved regime indicator, S_{it}, based on all the observed information in the sample, yielding the smoothed latent variable estimates, $\boldsymbol{\eta}_{it|T} = E(\boldsymbol{\eta}_{it}|\mathbf{Y}_{iT})$, and the smoothed regime probabilities, $Pr[S_{it} = k|\mathbf{Y}_{iT}]$. These elements can be estimated by means of the extended Kim smoother. Estimates from the extended Kim smoother, $\boldsymbol{\eta}_{i,t|T}$ and $Pr(S_{it} = k|\mathbf{Y}_{iT})$, under some regularity conditions (Bar-Shalom, Li, & Kirubarajan, 2001), are more accurate than those from the extended Kim filter, since the former is based on information from the entire time series rather than from previous information up to the current observations, as in the extended Kim filter. More detailed descriptions of the extended Kim filter, the extended Kim smoother, and other related steps are included in the Appendix.

In sum, our proposed approach utilizes the extended Kalman filter, the extended Kalman smoother, the prediction error decomposition function, and the Hamilton filter with a "collapsing procedure." This results in approximate ML point estimates for all the time-invariant parameters, and smoothed estimates of all the latent variables. Point estimates of all the time-invariant

parameters can be obtained by optimizing the prediction error decomposition function; the corresponding standard errors are then obtained by taking the square root of the diagonal elements of the negative numerical Hessian matrix of the prediction error decomposition function at the point of convergence. As described in the Appendix, information criterion measures such as the Akaike information criterion (AIC; Akaike, 1973) and Bayesian information criterion (BIC; Schwarz, 1978) can also be computed using the prediction error decomposition function.

5. Empirical Data Analysis

5.1. Data Descriptions and Preliminary Screening

To illustrate the utility of the proposed method, we used a subset of the data from the Affective Dynamics and Individual Differences (ADID; Emotions and Dynamic Systems Laboratory, 2010) study. Participants whose ages ranged between 18 and 86 years old enrolled in a laboratory study of emotion regulation, followed by an experience sampling study during which the participants rated their momentary feelings 5 times daily over a month. Only the experience sampling data were used in the present analysis. After removing the data of participants with excessive missingness (>65 % missingness) and data that lacked sufficient response variability, 217 participants were included in the final sample.

The two endogenous latent variables, PA and NA, were measured using items from the Positive Affect and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988) and other items posited in the circumplex model of affect (Larsen & Diener, 1992; Russell, 1980) on a scale of 1 (never) to 4 (very often). We created three item parcels as indicators of each of the two latent factors (PA and NA) via item parceling (Cattell & Barton, 1974; Kishton & Widaman, 1994).[3] All items were assessed 5 times daily at partially randomized intervals that included both daytime assessments as well at least one assessment in the evening. The participants were asked to keep between an hour and a half and four hours between two successive assessments. Because the original data were highly irregularly spaced whereas the proposed methodology is designed to handle equally spaced data, we aggregated the composite scores over every twelve-hour block to yield two measurements per day up to 37 days (a few of the participants continued to provide responses beyond the requested one-month study period).[4]

The total number of time points for each participant ranged from 26 to 74 time points, with an average missing data proportion of 0.18. The proposed RSSS model and a series of alternative models were fitted to data from all participants as a group. Prior to model fitting, we removed the linear time trend in each indicator separately for each individual. Investigation of the autocorrelation and related plots of the residuals indicated that no notable trend was present in the residuals. With the exceptions of a few individuals who showed weekly trends (e.g., statistically significant auto- and partial-autocorrelations at lags 7, 14), the preliminary data screening indicated that there were statistically significant lag-1 partial autocorrelations but no consistent, statistically significant partial autocorrelations at higher lags.

[3]The items included in the three parcels included: (1) for PA parcel 1, elated, affectionate, lively, attentive, active, satisfied and calm; (2) for PA parcel 2, excited, love, enthusiastic, alert, interested, pleased and happy; (3) for PA parcel 3, aroused, inspired, proud, determined, strong and relaxed; (4) for NA parcel 1, angry, sad, distressed, jittery, guilty and afraid; (5) for NA parcel 2, upset, hostile, irritable, tense and ashamed, and (6) for NA parcel 3, depressed, agitated, nervous, anxious and scared.

[4]Although the proposed approach can handle missing values assumed to be missing completely at random or missing at random (Little & Rubin, 2002), the wide-ranging time intervals in the original data would necessitate the insertion of too many "missing values" between some of the observed time points to create a set of equally spaced data. Thus, the data set in its original form is not particularly conducive for the illustration in the present article.

TABLE 1.
Summary of the series of models fitted to the ADID data.

Model label	Descriptions	Pertinent equations
Model 1	One-regime model	$b_{PN,S_{it}} = b_{PN0}\left(\frac{\exp(\text{abs}(NA_{i,t-1}))}{1+\exp(\text{abs}(NA_{i,t-1}))}\right)$ $b_{NP,S_{it}} = b_{NP0}\left(\frac{\exp(\text{abs}(PA_{i,t-1}))}{1+\exp(\text{abs}(PA_{i,t-1}))}\right)$
Model 2	Two-regime nonlinear model	Equations (1–2)
Model 3	Stress-based cross-regression	$b_{PN,S_{it}} = \begin{cases} 0 & \text{if } S_{it}=0 \\ b_{PN0}+b_{PN1}Stress_{it} & \text{if } S_{it}=1 \end{cases}$ $b_{NP,S_{it}} = \begin{cases} 0 & \text{if } S_{it}=0 \\ b_{NP0}+b_{NP1}Stress_{it} & \text{if } S_{it}=1 \end{cases}$
Model 4	Model 3 + regime-dependent autoregression parameters	Equations (1–2) $a_{P,S_{it}} = \begin{cases} a_{P1} & \text{if } S_{it}=0 \\ a_{P2} & \text{if } S_{it}=1 \end{cases}$ $a_{N,S_{it}} = \begin{cases} a_{N1} & \text{if } S_{it}=0 \\ a_{N2} & \text{if } S_{it}=1 \end{cases}$
Model 5	Best-fitting model	Equations (1–2) $a_{P,S_{it}} = \begin{cases} 0 & \text{if } S_{it}=0 \\ a_{P2} & \text{if } S_{it}=1 \end{cases}$ $a_{N,S_{it}} = \begin{cases} 0 & \text{if } S_{it}=0 \\ a_{N2} & \text{if } S_{it}=1 \end{cases}$

Note: AIC for *Models 1–5* = 118410, 118160, 173330, 117790, 117450; BIC for *Models 1–5* = 118530, 118290, 173480, 117940, 117590.

5.2. *Models Considered and Modeling Results*

The model depicted in Equations (1–2) is simply one example of the many models that can be used to describe patterns of change in multivariate time series. We considered a series of alternative models, a summary of which is presented in Table 1. The first model, denoted as *Model 1*, was a one-regime model in which the cross-regression parameters were specified to follow the dynamic functions governing the high activation regime in Equation (2). The second model, *Model 2*, was the two-regime nonlinear process factor analysis model described in the motivating example section (see Equations (1–2)). *Model 3*, the third model we considered, posited that the cross-regression parameters varied as a function of a time-varying covariate, namely, perceived stress as measured using the Perceived Stress Scale (PSS; Cohen, Kamarck, & Mermelstein, 1983). The model is a two-regime model that provided a linear alternative to testing the Dynamic Model of Activation. Specifically, a time-varying covariate, perceived stress, was used to predict individuals' over-time deviations in cross-regression strengths. *Model 4* is the most complex variation considered. In this model, the cross-regression parameters were specified to conform to the same regime-dependent functions as posited in *Model 2*. In addition, the autoregression parameters governing PA and NA were also allowed to be regime-dependent. This model was proposed as an adaptation to *Model 2* based on post-hoc examination of the residual patterns from model fitting. Finally, based on fit indices and evaluations of the autocorrelation patterns in the residuals, *Model 5* was proposed as the best-fitting model, the details of which will be presented later.

We began by fitting *Models 1–3* to the empirical data and compared their AIC and BIC values. Based on the information criterion measures, *Model 2*, namely, the two-regime nonlinear process factor analysis model described in the Motivating Example section, provided the best fit among these three models. To further diagnose possible sources of misfit, we computed the

TABLE 2.
Results from empirical model fitting.

Parameters	Estimates (SE)
λ_{21}	1.20 (0.00)
λ_{31}	1.14 (0.02)
λ_{52}	1.02 (0.00)
λ_{62}	0.95 (0.00)
p_{11}	0.86 (0.01)
p_{22}	0.82 (0.00)
a_{P2}	0.50 (0.02)
a_{N2}	0.81 (0.01)
b_{PN0}	−0.19 (0.02)
b_{NP0}	−0.08 (0.01)
σ_{ϵ_1}	0.28 (0.00)
σ_{ϵ_2}	0.11 (0.01)
σ_{ϵ_3}	0.12 (0.01)
σ_{ϵ_4}	0.13 (0.00)
σ_{ϵ_5}	0.12 (0.00)
σ_{ϵ_6}	0.11 (0.00)
$\sigma_{\zeta_{PA}}$	0.32 (0.01)
$\sigma_{\zeta_{NA}}$	0.19 (0.00)

discrepancies between the lag-0 and lag-1 autocorrelation structures of the composite PA and NA scores, and those obtained using the latent variable scores estimated using the model.[5] Some of the notable discrepancies stemmed from overestimation in the lag-1 positive autocorrelation in NA, especially in the independent regime. One possible way to circumvent this discrepancy is to allow the autoregressive parameters and particularly a_N to also be regime-dependent. Thus, we considered *Model 4*, a two-regime model that extended the model shown in Equations (1–2) by also allowing the autoregression parameters to be regime-dependent. That is, in addition to allowing the cross-regression parameters to assume regime-dependent values as depicted in Equation (2), the autoregression parameters were specified to be regime-dependent. This model showed lower AIC and BIC values than all other models considered, but the autoregression parameters for the independent regime were observed to be close to zero. We thus proceeded to constraining the autoregression parameters in the independent regime to be zero and chose the resultant model, denoted as *Model 5* in Table 1, as the best-fitting model. We focus herein on elaborating results from *Model 5*.

The estimated covariance between the process noises for PA and NA (i.e., $\sigma_{\zeta_{PA},\zeta_{NA}}$) was close to zero. Given that PA and NA were supposed to be two independent dimensions from a theoretical standpoint, we fixed this covariance parameter to be zero. All other parameters were found to be statistically different from zero at the 0.05 level and the corresponding parameter and standard error estimates are summarized in Table 2.

The independent regime was characterized by zero covariance between the process noise components of PA and NA, as well as zero auto- and cross-regression terms. Thus, while in this regime, PA and NA were indeed found to fluctuate as two independent, noise-like processes that showed ebbs and flows as driven by external shocks. The high-activation regime differed from the independent regime in two key ways. First, the moderate to large positive AR(1) parameter

[5]For the composite scores, we aggregated each participant's ratings across item parcels to obtain a composite PA score and a composite NA score for each person and time point. The lagged correlation matrix computed using these composite scores was then compared to the lagged correlation matrix computed using the latent variable scores estimated using Equations (A.9–A.11) in the Appendix.

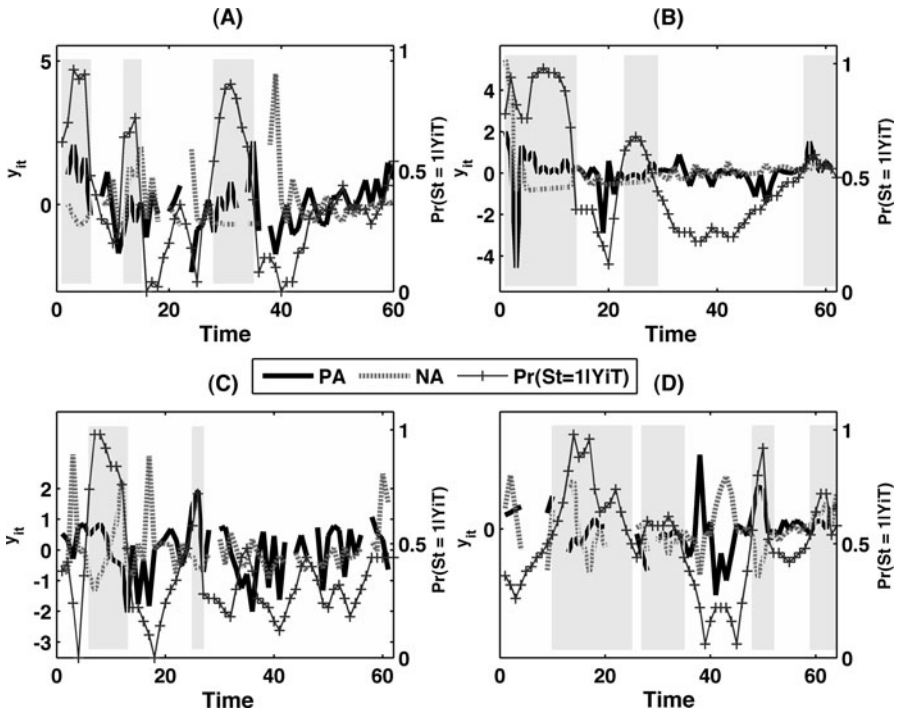

FIGURE 2.
Observed data and estimates from four randomly selected participants. The *shaded regions* represent portions of the data where $P(S_{it} = 1|\mathbf{Y}_{iT}) \geq 0.5$.

estimates in this regime ($a_{P2} = 0.50$ and $a_{N2} = 0.81$) suggested that if an individual showed deviations in PA and NA away from their baseline PA and NA (i.e., zero) in the high-activation regime, such deviations tended to diminish over time relatively slowly.

Second, the negative deviations in cross-regression parameters (i.e., b_{PN0} and b_{NP0}) indicated that PA (NA) from the previous occasion was inversely related to an individual's current NA (PA) in this regime, suggesting that this regime may be interpreted as a high-activation, "reciprocal" phase (Cacioppo & Berntson, 1999). That is, high deviations in PA or NA from its baseline level at $t - 1$ tended to reduce the deviations of the other emotion process from its baseline. In other words, extreme PA (NA) from the previous occasion tended to bring an individual's NA (PA) back to its baseline. Thus, above-baseline NA level at time $t - 1$ (i.e., when $NA_{i,t-1} > 0$) tended to reduce an individual's PA at time t if it was above baseline and elevate the individual's PA if it was below baseline, for instance. A lagged influence of a similar nature also existed in the direction from $PA_{i,t-1}$ to NA_{it}.

The estimated transition probabilities of staying within regime 0 and regime 1 (p_{11} and p_{22}, respectively) suggested that the individuals in the current sample showed a slightly higher probability of staying within the independent regime ($p_{11} = 0.86$) than within the high-activation regime ($p_{22} = 0.82$). Although this difference was small, the higher staying probability of the independent regime suggested that the independent regime was observed at a slightly higher rate than the high-activation regime.

Figure 2 shows the observed data from four randomly selected participants and their corresponding estimates of being in the "high-activation" regime (i.e., $P(S_{it} = 1|\mathbf{Y}_{iT})$). The shaded regions of the plots serve to identify portions of the data where $P(S_{it} = 1|\mathbf{Y}_{iT}) \geq 0.5$, namely, the time points at which a particular individual has greater than or equal to 0.5 probability of being in the "high-activation" regime given the observed data. These shaded regions concurred with

the measurement occasions on which individuals' PA and NA appeared to show divergence in trends and levels. That is, the shaded regions captured the times when one emotion process was high and the other was low. The slightly lower probability of staying within the high-activation regime from time $t-1$ to time t is reflected in the relatively small areas of the shaded regions as compared to the unshaded regions. This was evidenced in the plots of three of the four participants (see Panels A–C), although the participant shown in Panel D did show relatively high occurrences of the high-activation regime throughout the study span.

6. Simulation Study

6.1. Simulation Designs

The purpose of the simulation study was to evaluate the performance of the proposed methodological approach in recovering the true values of the parameters, their associated SEs, the true latent variable scores and the latent regime indicator. Four combinations of sample sizes and time series lengths were considered, namely, with (1) $T = 30$, $n = 100$, (2) $T = 300$, $n = 10$, (3) $T = 60$, $n = 200$ and (4) $T = 200$, $n = 60$. The first condition was characterized by fewer observation points than our empirical example. It provided a moderate-T-small-n comparison condition that might be more reasonable in empirical settings. The second condition had been shown in the past to yield reasonable point and SE estimates in fitting nonlinear dynamic models (Chow et al., 2011b) and it served to provide a large-T-small-n comparison case with the same number of total observation points as the first condition. The third condition was specifically selected to mirror the sample size/time series length of our empirical data whereas the fourth condition provided a large-T-small-n comparison case with the same total number of observation points. Missing data were not the focus of the present study; we included 20 % of missingness in the simulated data following a missing completely at random mechanism to yield a comparable amount of missingness to our empirical data.

Model 2, the model described in the Motivating Example section, was used as our simulation model. The population values of all the time-invariant parameters were chosen to closely approximate those obtained from empirical model fitting when *Model 2* was fitted to the empirical data. We set the factor loading matrix in both regimes to

$$\mathbf{\Lambda}_1 = \mathbf{\Lambda}_2 = \mathbf{\Lambda} = \begin{bmatrix} 1 & 1.20 & 1.20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.10 & 0.95 \end{bmatrix}'.$$

The uniquenesses, $\boldsymbol{\epsilon}_{it}$, were specified to be normally distributed with zero means and the same covariance matrix across regimes, with $\mathbf{R} = \text{diag}[0.28\ 0.10\ 0.12\ 0.13\ 0.12\ 0.11]$. The process noise covariance matrix was set to be invariant across regimes, with $\mathbf{Q} = \text{diag}[0.35\ 0.3]$. With two regimes in total, only one element of each row of the 2×2 transition probability matrix can be freely estimated. We chose to estimate the parameters p_{11} and p_{22} and set the transition probability matrix to

$$\mathbf{P} = \begin{bmatrix} p_{11} = 0.98 & 1 - p_{11} = 0.02 \\ 1 - p_{22} = 0.15 & p_{22} = 0.85 \end{bmatrix} \tag{9}$$

based on the empirical parameter estimates. Other parameters that appear in Equations (1–2) were set to $a_P = 0.2$, $a_N = 0.25$, $b_{PN0} = -0.6$ and $b_{NP0} = -0.8$.

We assumed that the latent variables at the first time point were multivariate normally distributed with zero means and a covariance matrix that was equal to an identity matrix. We discarded the first 50 time points and retained the rest of the simulated data for model fitting purposes. In model fitting, it is necessary to specify the means and covariance matrix associated

TABLE 3.
Summary statistics of parameter estimates for the time-invariant parameters for $T = 30$ and $n = 100$ across 200 Monte Carlo replications.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	SD	$a\widehat{SE}$	RDSE	Coverage of 95 % CIs
λ_{21}	1.20	1.20	0.00	0.00	0.029	0.018	−0.38	0.87
λ_{31}	1.20	1.20	0.00	−0.00	0.028	0.017	−0.40	0.89
λ_{52}	1.10	1.10	0.00	0.00	0.021	0.019	−0.13	0.95
λ_{62}	0.95	0.95	0.00	0.00	0.020	0.016	−0.20	0.95
p_{11}	0.98	0.97	0.01	−0.01	0.040	0.012	−0.70	0.74
p_{22}	0.85	0.83	0.02	−0.02	0.094	0.039	−0.58	0.69
a_P	0.20	0.20	0.00	−0.01	0.025	0.023	−0.07	0.90
a_N	0.25	0.25	0.00	−0.01	0.024	0.024	−0.00	0.93
b_{PN0}	−0.60	−0.62	0.02	0.03	0.151	0.112	−0.26	0.90
b_{NP0}	−0.80	−0.78	0.02	−0.03	0.156	0.120	−0.23	0.87
$\sigma^2_{\epsilon_1}$	0.28	0.27	0.01	−0.05	0.010	0.009	−0.07	0.79
$\sigma^2_{\epsilon_2}$	0.10	0.09	0.01	−0.05	0.008	0.006	−0.21	0.92
$\sigma^2_{\epsilon_3}$	0.12	0.11	0.01	−0.05	0.008	0.007	−0.13	0.92
$\sigma^2_{\epsilon_4}$	0.13	0.12	0.01	−0.05	0.006	0.005	−0.16	0.94
$\sigma^2_{\epsilon_5}$	0.12	0.11	0.01	−0.05	0.006	0.006	−0.00	0.96
$\sigma^2_{\epsilon_6}$	0.11	0.10	0.01	−0.05	0.005	0.005	−0.03	0.94
$\sigma^2_{\zeta_{PA}}$	0.35	0.34	0.01	−0.04	0.017	0.013	−0.23	0.69
$\sigma^2_{\zeta_{NA}}$	0.30	0.29	0.01	−0.04	0.013	0.011	−0.12	0.82

Note: True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_k$, where $\hat{\theta}_k$ = estimate of θ from the kth Monte Carlo replication; RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)^2}$; rBias = relative bias $= \frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)/\theta$; SD = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; RDSE = 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 θ.

with the initial distribution of the latent variables at time 0 (see the Appendix). We set the initial means to be a vector of zeros and the covariance matrix to be an identity matrix.

Two hundred Monte Carlo replications were performed. The root mean squared error (RMSE) and relative bias were used to quantify the performance of the approximate 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, 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 were constructed for each of the $N = 200$ Monte Carlo samples in each condition by adding and subtracting $1.96 * \text{SE}$ estimate in each replication to the parameter estimate from the replication. The coverage performance of a confidence interval was assessed with its empirical coverage rate, namely, the proportion of 95 % CIs covering θ across the Monte Carlo replications.

6.2. *Simulation Results*

Statistical properties of the approximate ML estimator across all conditions are summarized in Tables 3, 4, 5, 6. In general, the point and SE estimates for all sample size configurations were close to the true parameter values and their associated empirical SEs. To facilitate comparisons

TABLE 4.
Summary statistics of parameter estimates for the time-invariant parameters for $T = 300$ and $n = 10$ across 200 Monte Carlo replications.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	SD	Mean $\widehat{SE}$	RDSE	Coverage of 95 % CIs
λ_{21}	1.20	1.20	0.00	0.00	0.029	0.017	−0.42	0.84
λ_{31}	1.20	1.20	0.00	0.00	0.029	0.017	−0.42	0.89
λ_{52}	1.10	1.10	0.00	0.00	0.023	0.018	−0.23	0.94
λ_{62}	0.95	0.95	0.00	0.00	0.021	0.016	−0.26	0.94
p_{11}	0.98	0.98	0.00	−0.00	0.014	0.008	−0.39	0.71
p_{22}	0.85	0.83	0.02	−0.02	0.091	0.038	−0.58	0.68
a_P	0.20	0.20	0.00	−0.00	0.023	0.023	−0.03	0.95
a_N	0.25	0.25	0.00	−0.02	0.023	0.023	0.00	0.94
b_{PN0}	−0.60	−0.59	0.01	−0.02	0.129	0.109	−0.16	0.92
b_{NP0}	−0.80	−0.80	0.00	0.00	0.149	0.121	−0.19	0.89
$\sigma^2_{\epsilon_1}$	0.28	0.27	0.01	−0.05	0.009	0.009	0.04	0.79
$\sigma^2_{\epsilon_2}$	0.10	0.09	0.01	−0.06	0.007	0.006	−0.15	0.94
$\sigma^2_{\epsilon_3}$	0.12	0.11	0.01	−0.05	0.008	0.006	−0.15	0.95
$\sigma^2_{\epsilon_4}$	0.13	0.12	0.01	−0.05	0.006	0.005	−0.14	0.94
$\sigma^2_{\epsilon_5}$	0.12	0.11	0.01	−0.04	0.006	0.006	−0.02	0.96
$\sigma^2_{\epsilon_6}$	0.11	0.10	0.01	−0.05	0.005	0.005	−0.04	0.92
$\sigma^2_{\zeta_{PA}}$	0.35	0.33	0.02	−0.05	0.017	0.012	−0.30	0.61
$\sigma^2_{\zeta_{NA}}$	0.30	0.29	0.01	−0.04	0.012	0.011	−0.13	0.84

Note: True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_k$, where $\hat{\theta}_k$ = estimate of θ from the kth Monte Carlo replication; RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)^2}$; rBias = relative bias $= \frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)/\theta$; SD = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; RDSE = 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 θ.

across sample size configurations, we plotted the RMSEs, biases, and standard deviations (SDs) of the point estimates, biases of the estimated SEs in comparison to the empirical SEs, and coverage rates across the four sample size conditions in Figure 3, Panels A–E, as grouped by the type of parameters. That is, we averaged the outcome measures within each sample size condition as grouped by five types of parameters, including (1) factor loadings parameters in $\mathbf{\Lambda}$, (2) measurement error variances in $\mathbf{R}$,[6] (3) dynamic/time series parameters including a_P, a_N, b_{PN0}, and b_{NP0}, (4) process noise variances in $\mathbf{Q}$ and (5) the transition probability parameters, including p_{11} and p_{22}.

The point estimates associated with the five classes of parameters generally displayed small RMSEs and biases in point estimates across all sample size conditions. As the total observation points increased, improved precision was observed in all five classes of point estimates (see plot of the SDs of the parameter estimates in Figure 3, Panel C), particularly the time series and transition probability parameters. The biases of the process noise and measurement error variance parameters, although relatively small (< 0.015 in absolute value), remained largely constant in magnitude despite the increase in total sample size. These biases may be related to the truncation of higher-order terms in the Taylor series expansion used in the extended Kalman filter.

[6]The parameter $\sigma^2_{\epsilon_1}$ was omitted in computing these average coverage rates because this parameter was characterized by very low coverage rate due to systematic underestimation in the point estimates. To avoid skewing the comparisons across sample size conditions, this parameter was omitted in the computation of the estimates shown in Figure 3, Panel E.

TABLE 5.
Summary statistics of parameter estimates for the time-invariant parameters for $T = 60$ and $n = 200$ across 200 Monte Carlo replications.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	SD	Mean $\widehat{SE}$	RDSE	Coverage of 95 % CIs
λ_{21}	1.20	1.20	0.00	0.00	0.014	0.009	−0.39	0.92
λ_{31}	1.20	1.20	0.00	0.00	0.013	0.009	−0.35	0.93
λ_{52}	1.10	1.10	0.00	−0.00	0.011	0.009	−0.19	0.97
λ_{62}	0.95	0.95	0.00	−0.00	0.010	0.008	−0.17	0.97
p_{11}	0.98	0.98	0.00	−0.00	0.007	0.007	0.04	0.71
p_{22}	0.85	0.84	0.01	−0.01	0.043	0.019	−0.55	0.74
a_P	0.20	0.20	0.00	−0.01	0.012	0.011	−0.01	0.96
a_N	0.25	0.25	0.00	−0.00	0.012	0.011	−0.05	0.97
b_{PN0}	−0.60	−0.62	0.02	0.03	0.061	0.053	−0.12	0.92
b_{NP0}	−0.80	−0.80	0.00	0.00	0.071	0.058	−0.18	0.90
$\sigma^2_{\epsilon_1}$	0.28	0.27	0.01	−0.05	0.005	0.004	−0.08	0.39
$\sigma^2_{\epsilon_2}$	0.10	0.10	0.00	−0.05	0.004	0.003	−0.18	0.96
$\sigma^2_{\epsilon_3}$	0.12	0.11	0.01	−0.05	0.004	0.003	−0.17	0.90
$\sigma^2_{\epsilon_4}$	0.13	0.12	0.01	−0.05	0.003	0.003	−0.11	0.81
$\sigma^2_{\epsilon_5}$	0.12	0.11	0.01	−0.05	0.003	0.003	−0.05	0.95
$\sigma^2_{\epsilon_6}$	0.11	0.10	0.01	−0.05	0.003	0.002	−0.09	0.92
$\sigma^2_{\zeta_{PA}}$	0.35	0.33	0.02	−0.05	0.009	0.006	−0.28	0.81
$\sigma^2_{\zeta_{NA}}$	0.30	0.29	0.01	−0.04	0.006	0.006	−0.14	0.94

Note: True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_k$, where $\hat{\theta}_k$ = estimate of θ from the kth Monte Carlo replication; RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)^2}$; rBias = relative bias $= \frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)/\theta$; SD = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; RDSE = 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 θ.

Biases actually increased slightly in the time series parameters as sample size increased. Inspection of the summary statistics in Tables 3–6 revealed that the slight biases stemmed primarily from the parameters b_{PN0} and b_{NP0}. These parameters captured the full magnitudes of the cross-regression effects during the high-activation regime. Due to the logistic functions used in Equation (2), the full effects of these parameters were only manifested when PA and NA at time $t - 1$ were of extremely high intensity. Such instances were rare even for the (relatively) large sample sizes considered in the present simulation, and this may help explain the biases in the time series parameters in Figure 3, Panel B. Taking into consideration bias and precision information, point estimates of the time series parameters tended to show improvements in RMSEs with increase in the number of time points more so than the number of participants.

Sample size conditions with larger total sample sizes (e.g., $T = 60$, $n = 200$ and $T = 200$, $n = 60$) were observed to exhibit greater efficiency (in the sense of smaller average SE estimates and relatedly, smaller true SEs), as well as smaller biases in the SE estimates. Particularly worth noting was the clear reduction in biases of the SE estimates associated with the transition probability parameters with increase in total observation points. Consistent with the improved precision seen in the point estimates with larger T, smaller biases in the SE estimates were observed among the time series parameters with increase in the number of time points, even when compared against conditions with the same total number of observation points. This improvement was particularly salient when n was also small. This may be because in data of larger T,

TABLE 6.
Summary statistics of parameter estimates for the time-invariant parameters for $T = 200$ and $n = 60$ across 200 Monte Carlo replications.

	True θ	Mean $\hat{\theta}$	RMSE	rBias	SD	$a\widehat{SE}$	RDSE	Coverage of 95 % CIs
λ_{21}	1.20	1.20	0.00	0.00	0.014	0.009	−0.39	0.89
λ_{31}	1.20	1.20	0.00	0.00	0.013	0.009	−0.32	0.94
λ_{52}	1.10	1.10	0.00	0.00	0.012	0.009	−0.28	0.96
λ_{62}	0.95	0.95	0.00	0.00	0.010	0.008	−0.20	0.99
p_{11}	0.98	0.98	0.00	−0.00	0.006	0.004	−0.25	0.69
p_{22}	0.85	0.84	0.01	−0.01	0.040	0.019	−0.53	0.69
a_P	0.20	0.20	0.00	−0.01	0.011	0.011	−0.00	0.97
a_N	0.25	0.25	0.00	−0.02	0.013	0.011	−0.10	0.94
b_{PN0}	−0.60	−0.61	0.01	0.02	0.069	0.056	−0.19	0.90
b_{NP0}	−0.80	−0.80	0.00	−0.00	0.074	0.066	−0.11	0.93
$\sigma^2_{\epsilon_1}$	0.28	0.27	0.01	−0.05	0.005	0.005	−0.02	0.41
$\sigma^2_{\epsilon_2}$	0.10	0.10	0.00	−0.05	0.003	0.003	−0.04	0.97
$\sigma^2_{\epsilon_3}$	0.12	0.11	0.01	−0.05	0.003	0.003	−0.06	0.94
$\sigma^2_{\epsilon_4}$	0.13	0.12	0.01	−0.05	0.003	0.003	−0.07	0.83
$\sigma^2_{\epsilon_5}$	0.12	0.11	0.01	−0.05	0.003	0.003	−0.05	0.94
$\sigma^2_{\epsilon_6}$	0.11	0.10	0.01	−0.05	0.002	0.002	−0.02	0.95
$\sigma^2_{\zeta_{PA}}$	0.35	0.33	0.02	−0.05	0.009	0.006	−0.28	0.74
$\sigma^2_{\zeta_{NA}}$	0.30	0.29	0.01	−0.04	0.007	0.005	−0.21	0.95

Note: True θ = true value of a parameter; Mean $\hat{\theta} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_k$, where $\hat{\theta}_k$ = estimate of θ from the kth Monte Carlo replication; RMSE $= \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)^2}$; rBias = relative bias $= \frac{1}{N}\sum_{k=1}^{N}(\hat{\theta}_k - \theta)/\theta$; SD = standard deviation of estimates across Monte Carlo runs; Mean $\widehat{SE}$ = average standard error estimate across Monte Carlo runs; RDSE = 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 θ.

more lagged information that is free of the influence of initial condition specification is available to convey information concerning dynamics and transition between regimes.

For all conditions, the coverage rates of the 95 % CIs were relatively close to the 0.95 nominal rate for most parameters. Consistent with the simulation results reported for linear RSSS models (Yang & Chow, 2010), larger biases, larger RDSEs and lower coverage rates (namely, compared to the nominal rate of 0.95) were observed for some of the transition probability and process noise variance parameters. Slightly greater biases and lower efficiency were observed when estimating p_{22} as opposed to p_{11} because the former was closer to zero. That is, if the probability of staying within any of the regimes is low (such as in the case of p_{22}, which was equal to 0.85) and either T or n is small, there may be insufficient realizations of data from that particular regime to facilitate estimation. Thus, larger sample sizes (both in terms of n and T) are needed to improve properties of the variance and transition probability parameters.

The negative values of RDSE observed for most parameters in Tables 3–6 suggested that there was a tendency for the SE estimator to underestimate the true variability in most of the parameters. As shown in Figure 3, Panel D, the underestimation was particularly salient for the transition probability parameters and time series parameters, although all biases approached zero as the total observation points increased. The systematic biases in the SE estimates may be related to (1) the truncations of the use of full regime history to estimate the latent variables and all modeling parameters (see the Appendix) and (2) the truncation of higher-order terms

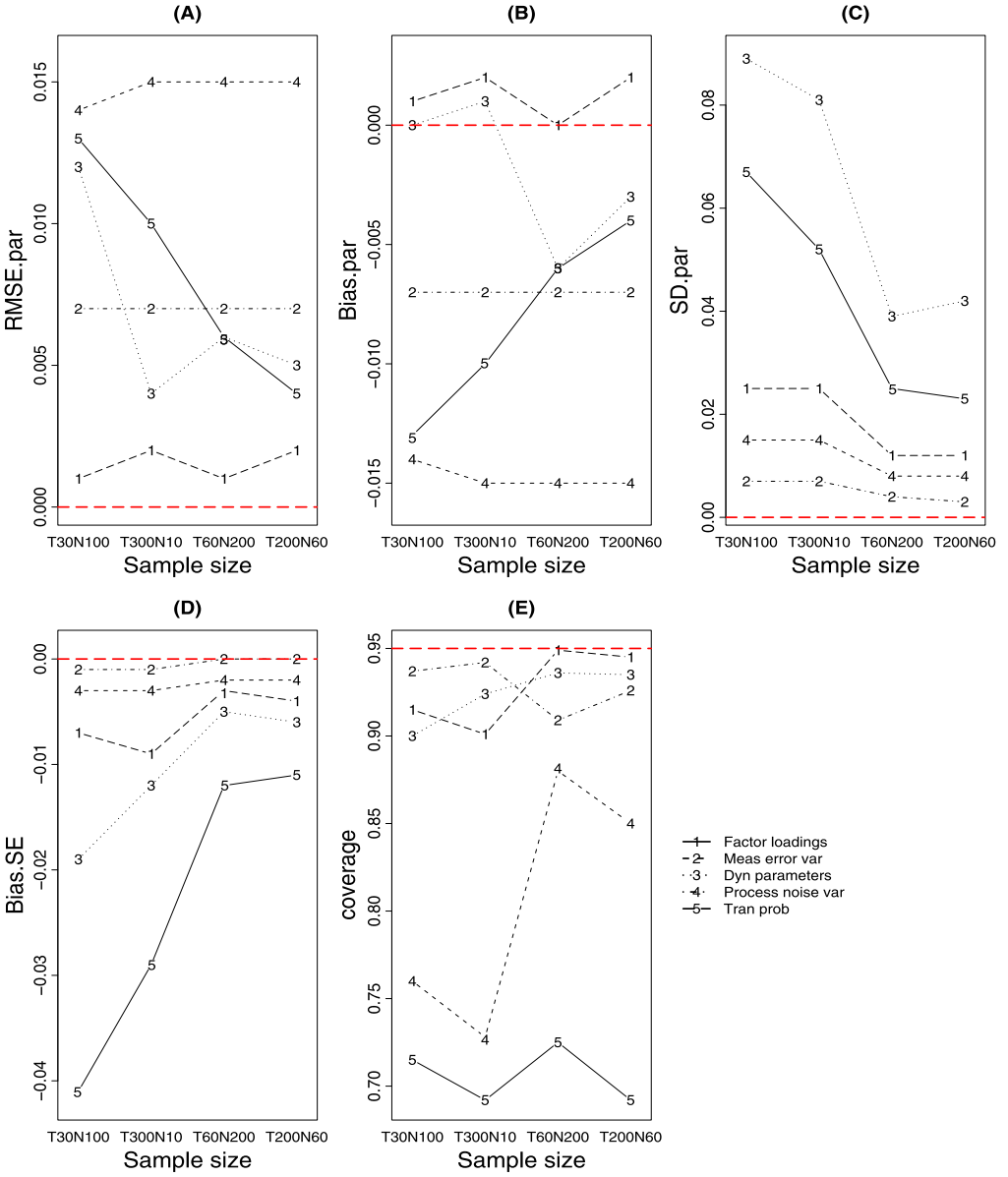

FIGURE 3.
(**A**–**C**) Plots of the average RMSEs, biases and standard deviations of the point estimates, (**D**) biases of the SE estimates, and (**E**) coverage rates. Factor loadings = Factor loading parameters; Meas error var = measurement error variances; Dyn parameters = dynamic/time series parameters including a_P, a_N, b_{PN0} and b_{NP0}; Process noise var = process noise variances and Tran prob = transition probability parameters, including p_{11} and p_{22}.

in the Taylor series expansion in the extended Kalman filter for linearization purposes. Thus, because of these two sources of approximation errors (and hence, mild model misspecification), systematic biases may be expected in the point as well as SE estimates despite improvements with increasing sample size.

Plots of the true values of one of the latent variables, the true regime indicator and their corresponding estimates for two randomly selected cases from each sample size condition are

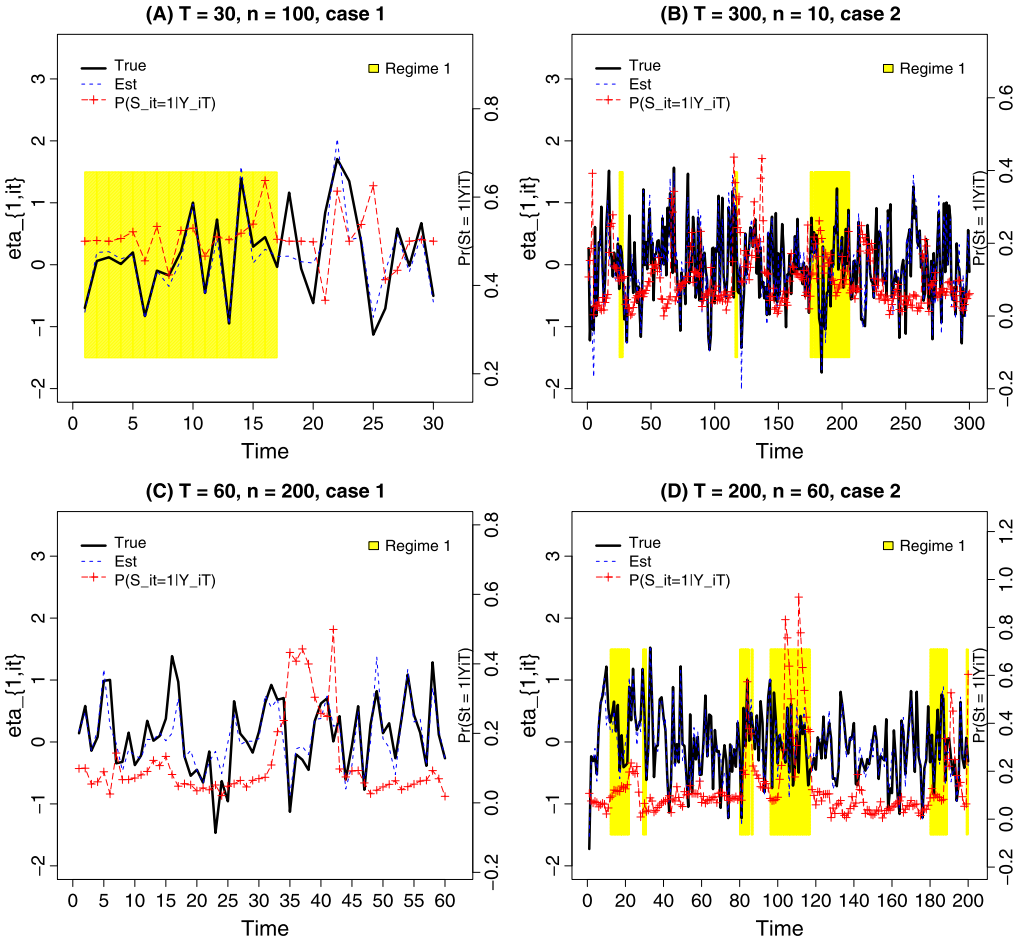

FIGURE 4.
Plots of the true and estimated latent variable scores, $\boldsymbol{\eta}_{i,t|T}$, the true regime and estimated probability of being in regime 1 at time t, $p(S_{it} = 1|\boldsymbol{Y}_{iT})$, for one randomly selected case in each sample size condition. True = true simulated data, Est = $\boldsymbol{\eta}_{i,t|T}$; portions of the data where the true $S_{it} = 1$ are marked as shaded regions.

shown in Figure 4, Panels A–D. The proposed algorithm yielded satisfactory latent variable score estimates (see examples in Figure 4, Panels A–D). We computed RMSEs for the latent variable scores as $\sqrt{\frac{1}{nNT}\sum_{k=1}^{N}\sum_{i=1}^{n}\sum_{t=1}^{T}(\eta_{l,itk} - \eta_{l,itk|Y_{iT}})^2}$ for $l = 1, 2$. Biases in the latent variable estimates were similar across all sample size conditions, with an average RMSE (across all time points, people and Monte Carlo replications) around 0.30 for both latent variables across all sample size conditions. Given the relatively large standard deviations of the latent variables (e.g., as indicated by $\sqrt{Var(\eta_{l,it}|Y_{iT})}$), these RMSEs were within a reasonable range.

Estimates of the probability of being in regime 1, $P(S_{it} = 1|Y_{iT})$, were able to capture some, but not all of the shifts in regimes (see examples in Figure 4, Panels A–D). Greater inaccuracies were evidenced when the time spent staying in a particular regime was brief. To evaluate the performance of the regime indicator estimator, we classified the data for individual i at time t into regime 1 when $P(S_{it} = 1|Y_{iT})$ was greater than or equal to 0.5; the remaining data were assigned to regime 0. Denoting S_{it} as the true regime value, and $\hat{S}_{it}$ as the classified regime value, we computed power (or sensitivity) and Type I error rate (in other words, 1-specificity,

denoted below as α) as

$$power = \frac{\text{number of instances where } S_{it} = 1 \text{ and } \hat{S}_{it} = 1}{\text{number of instances where } S_{it} = 1}, \tag{10}$$

$$\alpha = \frac{\text{number of instances where } S_{it} = 0 \text{ and } \hat{S}_{it} = 1}{\text{number of instances where } S_{it} = 0}. \tag{11}$$

Type 1 error rates were low in both conditions, with α = around 0.01 for all conditions. Power was also low, however. As sample sizes varied in the order depicted in Figure 3 (i.e., from $T = 30$, $n = 100$ to $T = 200$, $n = 60$), power was estimated to be = 0.15, 0.16, 0.14, and 0.14, respectively. This shows that the proposed algorithm does not have enough sensitivity in detecting instances of transition into regime 1 with the specific cut-off values used for classification purposes.

The low power of detecting the correct regime can be understood given the small separation between the two regimes. One possible distance measure for defining the separation between regimes is the multivariate Hosmer's measure of distance between the two regimes (Hosmer, 1974; Dolan & Van der Maas, 1998; Yung, 1997). Here, a state-space equivalent of the measure at a particular time point is given by

$$\max_{h\in\{1,2\}} \left[(\boldsymbol{\mu}_{1,it} - \boldsymbol{\mu}_{2,it})' \boldsymbol{\Sigma}_h^{-1} (\boldsymbol{\mu}_{1,it} - \boldsymbol{\mu}_{2,it})\right]^{1/2}, \tag{12}$$

where $\boldsymbol{\mu}_h = E(\boldsymbol{y}_{it}|S_{it} = h)$ and $\boldsymbol{\Sigma}_h = \text{Cov}(\boldsymbol{y}_{it}, \boldsymbol{y}'_{it}|S_{it} = h)$. When the regimes considered have different $\boldsymbol{\Sigma}_h$, the $\boldsymbol{\Sigma}_h$ that gave rise to larger Hosmer's distance was used. Yung (1997) reported that in cross-sectional mixture structural equation models with one time-point, a Hosmer's distance of 3.8 or above yielded satisfactory estimation results. In the case of our proposed model, the associated Hosmer's distance for each time point was approximately 0.01.[7] Whereas the Hosmer's distance of a single individual does increase with T when accumulated over all time points and led to satisfactory point and SE estimates, the corresponding accuracy associated with accurate regime classification at each time point was clearly less than optimal. The small separation between regimes was due in part to the specification of zero intercepts in the measurement and dynamic equations and relatedly, the use of detrended data. Possible ways to increase regime separation within the context of the proposed nonlinear RSSS models will be outlined in the Discussion section.

Although we did not use the best fitting model from the empirical illustration to construct our simulation model, most of the parameters present in the best-fitting model were also present in *Model 2*. Estimates obtained from fitting *Model 2* were also similar in ranges to those from *Model 5*. Some exceptions and corresponding consequences should be noted, however. First, higher staying probabilities were used in the simulations ($p_{11} = 0.98$ and $p_{22} = 0.85$) than those estimated based on *Model 5*. Based on the present simulation study as well as results reported elsewhere (Yang & Chow, 2010), lower probabilities of staying within regime generally lead to lower accuracy in point estimates and lower efficiency, especially when sample sizes are small. Second, when the autoregression parameters were constrained to be invariant across regimes in

[7]Note that this was only a rough estimate. In the case of linear state-space models that are stationary, closed-form expressions of $E(\boldsymbol{y}_{it}|S_{it})$ and $\text{Cov}(\boldsymbol{y}_{it}, \boldsymbol{y}'_{it}|S_{it})$ can be obtained analytically (see, e.g., p. 121, Harvey, 2001; Du Toit & Browne, 2007). In the case of our general modeling equations, $E(\boldsymbol{y}_{it}|S_{it}) = \mathbf{d}_{S_{it}} + \boldsymbol{\Lambda}_{S_{it}}[\mathbf{b}_{S_{it}}(\boldsymbol{\eta}_{i,t-1}, \boldsymbol{x}_{it})]$ whereas $\text{Cov}(\boldsymbol{y}_{it}, \boldsymbol{y}'_{it}|S_{it}) = \boldsymbol{\Lambda}_{S_{it}} \text{Cov}[\mathbf{b}_{S_{it}}(\boldsymbol{\eta}_{i,t-1}, \boldsymbol{x}_{it})]\boldsymbol{\Lambda}'_{S_{it}}$ and closed-form expressions of these functions cannot be obtained analytically. To yield an approximation, we generated data using the simulation model and a large sample size (i.e., with $T = 1000$ and $n = 1000$). Subsequently, we obtained the empirical means and covariance matrices of $\boldsymbol{y}_{it}|S_{it}$ over all people and time points.

Model 2, the autoregression estimates were positive but closer to zero and the cross-regression parameters in the high-activation regime were larger in absolute magnitude compared to estimates from *Model 5*. These discrepancies reflected how *Model 2* compensated for the between-regime differences in autoregression dynamics by increasing the absolute magnitudes of the negative cross-regression terms. Using the parameter estimates from *Model 5* would have given rise to slightly lower Hosmer's distance than that obtained from *Model 2* (0.005 as compared to 0.01). Thus, if the empirical estimates from *Model 5* were used to construct the simulation study, slight decrements in the performance of the estimation procedure can be expected.

7. Discussion

In the present article, we illustrated the utility of nonlinear RSSS models in representing multivariate processes with distinct dynamics during different portions or phases of the data. Such nonlinear RSSS models include nonlinear dynamic factor analysis models with regime switching as a special case. This class of modeling tools provides a systematic mechanism to probabilistically detect unknown (i.e., latent) regime or phase changes in linear as well as nonlinear dynamic processes. The overall model formulation and associated estimation procedures are flexible enough to accommodate a variety of linear and nonlinear dynamic models.

We illustrated the empirical utility of nonlinear RSSS models using a set of daily affect data. Results from our empirical application suggested that some of the subtle differences in affective dynamics would likely be bypassed if the data were analyzed as if they conformed to only one single regime. Other modeling extensions, are, of course, possible. For instance, a three-regime model in which the cross-regression parameters were depicted to be zero (an independent regime), positive (a coactivated regime) or negative (a reciprocal regime) is another interesting extension. Another modeling extension that has been considered in the context of SEM-based regime-switching models is to use covariates to predict the initial class probabilities and/or the transition probability parameters (e.g., Dolan, Schmittmann, Lubke, & Neale, 2005; Dolan, 2009; Schmittmann, Dolan, van der Maas, & Neale, 2005; Muthén & Asparouhov, 2011; Nylund-Gibson, Muthen, Nishina, Bellmore, & Graham, 2013). This is an interesting extension: In the presence of strong predictors of the transition probability parameters, the covariates may help improve the accuracy of regime classification. One other extension that is interesting but more difficult to implement in the state-space context is to allow the current regime indicator, S_{it}, to not only depend on the regime at a previous time point, namely, $S_{i,t-1}$, but also on other earlier regimes. This is the general framework implemented in SEM-based programs such as Mplus (Muthén & Asparouhov, 2011) and it extends the transition probability model from a first-order Markov process to higher-order Markov processes. The feasibility of adopting higher-order Markov specifications in the state-space framework with intensive repeated measures data is yet to be investigated.

Results from our simulation study showed that the proposed estimation procedures performed well under the sample size configurations considered in the present study. The point estimates generally exhibited good accuracy, with a number of areas in need of improvements. Specifically, slight biases remained in some of the parameters, lower accuracy in SE estimation was observed for some of the transition probability and variance parameters, and the accuracy of regime classification was not satisfactory. The statistical properties of the proposed estimation procedure can be improved by increasing the separation between the two hypothesized regimes. One way of doing so is to increase regime separation as defined through intercept terms. In many time series models, whether linear or nonlinear, intercepts are generally not the modeling focus and data are typically detrended and demeaned prior to model fitting, as was the case in the present study. However, between-regime differences in intercepts can be appropriately utilized

to increase the separation between regimes. For instance, from an affect modeling standpoint, the coactivated regime, in which both PA and NA are hypothesized to be jointly activated, may be characterized by very high intercepts in both PA and NA. The independent regime, in contrast, may be constrained to have generally low intercept levels for PA and NA. By a similar token, between-regime differences in the influences of time-varying covariates can be effectively utilized to increase the separation between regimes.

The proposed RSSS framework and associated estimation procedures offer some unique advantages over other existing approaches in the literature. First, the state-space formulation helps overcome some of the estimation difficulties associated with structural equation modeling-based approaches (e.g., Li et al., 2000; Wen et al., 2002) when intensive repeated measures data are involved (see e.g., Chow et al., 2010). Second, in contrast to standard linear RSSS (Kim & Nelson, 1999) models and linear covariance structure models with regime switching properties (Dolan et al., 2005; Dolan, 2009; Schmittmann et al., 2005), the change processes within each regime are allowed to be linear and/or nonlinear functions of the latent variables in the system as well as other time-varying covariates. Third, all RSSS models, including linear RSSS models (Kim & Nelson, 1999) and the nonlinear extensions considered herein, extend conventional state-space models (Chow et al., 2010; Durbin & Koopman, 2001) by allowing the inclusion of multiple state-space models.

Fourth, the nonlinear RSSS models proposed herein are distinct from another class of models referred to as *nonlinear regime-switching models*, some examples of which include the threshold autoregressive (TAR) models (Tong & Lim, 1980), self-exciting threshold autoregressive (SETAR) models (Tiao & Tsay, 1994) and Markov-switching autoregressive (MS-AR) models (Hamilton, 1989). Even though these regime-switching models are considered nonlinear models because the discrete shifts between regimes render the overall processes nonlinear (i.e., when marginalized or summed over regimes), the change process within each regime is still linear in nature. Fifth, by including an explicit dynamic model within each regime or class, RSSS models also differ from another class of well-known longitudinal models of discrete changes—the hidden Markov models (Elliott, Aggoun, & Moore, 1995), or the related latent transition models (which emphasize categorical indicators; Collins & Wugalter, 1992; Lanza & Collins, 2008). The specification of a continuous model of change within each regime allows the dynamics within regimes to be continuous in nature, even though the shifts between regimes or classes are discrete. In this way, RSSS models are more suited to representing processes wherein, in addition to the progression or shifts through discrete phases, the changes that unfold within regimes are also of interest.

One important difference in estimation procedures between the proposed RSSS modeling framework and linear SEMs with regime switching is worth noting. Within the structural equation modeling framework, longitudinal panel data with a relatively small number of time points are typically used to fit regime-switching models (Dolan et al., 2005; Dolan, 2009; Nylund-Gibson et al., 2013; Schmittmann et al., 2005). Consequently, the computational issues that motivated the "collapsing" procedure implemented in the Kim filter do not arise in this case. Thus, the likelihood expression used by these researchers is exact and does not involve the kind of approximation described in the Appendix. As the number of time points increases, however, the collapsing procedure of the Kim filter allows the estimation process to be computationally feasible, while the proposed extended Kim filter algorithm allows the change processes within regimes to be nonlinear.

Despite the promises of RSSS models, some limitations remain. Model identification may be a key issue, especially when the number of regimes increases and the distinctions between regimes are not pronounced. The increase in computational costs, especially when sample sizes are large, also poses additional challenges. In addition, when multiple regimes exist, multiple local maxima are prone to arise in the likelihood expression, thereby increasing the sensitivity

of the parameter estimates to starting values. As a result, it is recommended to use multiple sets of starting values to check if the corresponding estimation results have converged to the same values.

In our simulation study, the data were assumed to have started for 50 time points prior to data collection. There was, thus, a slight discrepancy between the true and specified initial distributions of the latent variables at the first retained time point. Misspecification in the structure of the initial variable distribution can lead to notably less satisfactory point and especially SE estimates in data of finite lengths. In cases involving linear stationary models, the model-implied means and covariance structures can be used to specify the initial distribution of the latent variables (Du Toit & Browne, 2007; Harvey, 2001). In cases involving nonstationary models, nonlinear adaptations of some of the alternative diffuse filters suggested in the state-space literature (e.g., De Jong, 1991; Koopman, 1997) can be used to replace the extended Kalman filter/smoother in our proposed procedure. In addition, the initial probabilities of the regime indicator at $t = 1$ were specified in the present study using model-implied values computed from the transition probability parameters (see Equation (4.49), p. 71; Kim & Nelson, 1999). Alternatively, these initial probabilities can be modeled explicitly as functions of other covariates (e.g., by using a multinomial logistic regression model as in Muthén & Asparouhov, 2011).

It is important to emphasize that our simulation study was designed to mirror several key features of our empirical data. Thus, our simulation results may be limited in generalizability to other conditions and models of change. In particular, most of the sample size configurations considered in the present study are relatively large in total observation points compared to other standard experience sampling studies. The $T = 30$, $n = 100$ condition is arguably closer in total observation points to most experience sampling studies in the literature (e.g., Chow et al., 2004; Ferrer & Nesselroade, 2003). It is reassuring that this sample size configuration yielded reasonable point and SE estimates. Nevertheless, generalization of the simulation results to other studies has to be done with caution.

We evaluated the performance of the proposed techniques when used with multiple-subject time series data. Frequently encountered in experience sampling studies (Heron & Smyth, 2010), such data are typically characterized by finite time lengths and a small to moderate number of participants. In light of the finite time lengths of such data, we utilized information from all individuals for parameter estimation purposes by constraining all but a subset of time-varying modeling parameters to be invariant across persons. This is in contrast to standard time series approaches that focus on modeling at the individual level. When the assumption of invariance holds across persons, such designs offer one way of pooling information from multiple participants for model estimation purposes. Our general modeling and estimation framework can, however, still be used with single-subject data. In this case, researchers are advised to have an adequate number of time points from each individual before proceeding with model fitting at the individual level. Another alternative is to consider mixed effects variations of the proposed models wherein model fitting is still performed using multiple-subject data, but some interindividual differences in the dynamic parameters are allowed (e.g., Kuppens, Allen, & Sheeber, 2010).

The estimation procedures described in the present article were written in MATLAB. Other statistical programs that can handle matrix operations, such as 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. Standard structural equation modeling programs such as Mplus (Muthén & Muthén, 2001) and Open-Mx (Boker, Neale, Maes, Wilde, Spiegel, & Brick, 2011), the msm (Jackson, 2011) and depmix (Visser, 2007) packages in R, and the PROC LTA procedure in SAS (Lanza & Collins, 2008) can be used to fit hidden Markov models and/or latent transition models. However, structural equation modeling programs are typically not conducive to handling a large number of observations whereas the msm, depmix, and

PROC LTA do not allow users to explicitly define the change processes within regimes/classes (e.g., as a state-space model).

Acknowledgements

Funding for this study was provided by a grant from NSF (BCS-0826844). We would like to thank Manshu Yang for thoughtful comments on earlier drafts of this manuscript.

Appendix

A.1. The Extended Kim Filter and Extended Kim Smoother

We outline the key procedures for implementing the extended Kim filter and extended Kim smoother here. Besides the modifications added to accommodate linearization constraints, the estimation process is identical to that associated with the linear Kim filter and Kim smoother. Interested readers are referred to Kim and Nelson (1999) for further details.

The extended Kim filter algorithm can be decomposed into three parts: the extended Kalman filter (for latent variable estimation), the Hamilton filter (to estimate the latent regime indicator, S_{it}) and a collapsing procedure (to consolidate regime-specific estimates to reduce computational burden). For didactic reasons, we will describe the extended Kalman filter followed by the collapsing process and finally, the Hamilton filter. In actual implementation, however, the Hamilton filter step has to be executed before the collapsing process takes place. MATLAB scripts for implementing these procedures with annotated comments can be downloaded from the first author's website at http://www.personal.psu.edu/quc16/.

A.2. The Extended Kalman Filter (EKF)

The extended Kalman filter (EKF) essentially provides a way to derive longitudinal factor or latent variable scores in real time as a new observation, $\boldsymbol{y}_{it}$, is brought in. Let $\boldsymbol{\eta}_{i,t|t-1}^{j,k} = E(\boldsymbol{\eta}_{it}|S_{it}=k, S_{i,t-1}=j, \mathbf{Y}_{i,t-1})$, $\mathbf{P}_{i,t|t-1}^{j,k} = \mathrm{Cov}(\boldsymbol{\eta}_{it}|S_{it}=k, S_{i,t-1}=j, \mathbf{Y}_{i,t-1})$, $\boldsymbol{v}_{it}^{j,k}$ is the one-step-ahead prediction errors and $\mathbf{F}_{it}^{j,k}$ is the associated covariance matrix; j and k are indices for the previous regime and current regime, respectively. The extended Kalman filter can be expressed as

$$\boldsymbol{\eta}_{i,t|t-1}^{j,k} = \mathbf{b}_k\left(\boldsymbol{\eta}_{i,t-1|t-1}^{j}, \boldsymbol{x}_{it}\right), \tag{A.1}$$

$$\mathbf{P}_{i,t|t-1}^{j,k} = \mathbf{B}_{k,it}\mathbf{P}_{i,t-1|t-1}^{j}\mathbf{B}_{k,it}' + \mathbf{Q}_k, \tag{A.2}$$

$$\boldsymbol{v}_{it}^{j,k} = \boldsymbol{y}_{it} - \left(\mathbf{d}_k + \boldsymbol{\Lambda}_k\boldsymbol{\eta}_{i,t|t-1}^{j,k} + \mathbf{A}_k\boldsymbol{x}_{it}\right), \tag{A.3}$$

$$\mathbf{F}_{it}^{j,k} = \boldsymbol{\Lambda}_k\mathbf{P}_{i,t|t-1}^{j,k}\boldsymbol{\Lambda}_k' + \mathbf{R}_k, \tag{A.4}$$

$$\boldsymbol{\eta}_{i,t|t}^{j,k} = \boldsymbol{\eta}_{i,t|t-1}^{j,k} + \mathbf{K}_{k,it}\boldsymbol{v}_{it}^{j,k}, \tag{A.5}$$

$$\mathbf{P}_{i,t|t}^{j,k} = \mathbf{P}_{i,t|t-1}^{j,k} - \mathbf{K}_{k,it}\boldsymbol{\Lambda}_k\mathbf{P}_{i,t|t-1}^{j,k}, \tag{A.6}$$

where $\mathbf{K}_{k,it} = \mathbf{P}_{i,t|t-1}^{j,k}\boldsymbol{\Lambda}_k'[\mathbf{F}_{it}^{j,k}]^{-1}$ is called the Kalman gain function; $\mathbf{B}_{k,it}$ is the Jacobian matrix that consists of differentiations of the dynamic functions around the latent variable estimates,

$\boldsymbol{\eta}^{j}_{i,t-1|t-1}$, namely, $\mathbf{B}_{k,it} = \frac{\partial \mathbf{b}_k(\boldsymbol{\eta}^{j}_{i,t-1|t-1}, \boldsymbol{x}_{it})}{\partial \boldsymbol{\eta}^{j}_{i,t-1|t-1}}$, with the time-varying covariates, $\boldsymbol{x}_{it}$, fixed at their observed values. The gth row and lth column of $\mathbf{B}_{k,it}$ carries the partial derivative of the gth dynamic function characterizing regime k with respect to the lth latent variable, evaluated at subject i's conditional latent variable estimates from time $t-1$, $\boldsymbol{\eta}^{j}_{i,t-1|t-1}$, from the jth regime. The subject index in $\mathbf{B}_{k,it}$ is used to indicate that the Jacobian matrix has different numerical values because it is evaluated at each person's respective latent variable estimates, not that the dynamic functions are subject-dependent. Because our hypothesized measurement functions are linear, no linearization of the measurement functions is needed.

The EKF summarized in Equations (A.1–A.6) works recursively (i.e., one time point at a time) from time 1 to T and $i = 1, \ldots, n$ until $\boldsymbol{\eta}^{j,k}_{i,t|t}$ and $\mathbf{P}^{j,k}_{i,t|t}$, have been computed for all time points and people. To start the filter, the initial latent variable scores at time $t = 0$, $\boldsymbol{\eta}_0$, are assumed to be distributed as $\boldsymbol{\eta}_0 \sim \text{MVN}(\boldsymbol{\eta}_{0|0}, \mathbf{P}_{0|0})$. Typically, $\boldsymbol{\eta}_0$ is assumed to have a *diffuse* density, that is, $\boldsymbol{\eta}_{0|0}$ is fixed to be a vector of constant values (e.g., a vector of zeros) and the diagonal elements of the covariance matrix $\mathbf{P}_{0|0}$ are set to some arbitrarily large constants.

A.3. The Collapsing Process

At each t, the EKF procedures utilize only the marginal estimates, $\boldsymbol{\eta}^{j}_{i,t-1|t-1}$ and $\mathbf{P}^{j}_{i,t-1|t-1}$, from the previous time point. This is because to ease computational burden, a collapsing procedure is performed on $\boldsymbol{\eta}^{j,k}_{i,t|t}$ and $\mathbf{P}^{j,k}_{i,t|t}$ after each EKF step to yield $\boldsymbol{\eta}^{k}_{i,t|t}$ and $\mathbf{P}^{k}_{i,t|t}$. Given a total of M regimes, if no collapsing is used, the M sets of computations involving $\boldsymbol{\eta}^{j}_{i,t-1|t-1}$ and $\mathbf{P}^{j}_{i,t-1|t-1}$ in Equations (A.1–A.2) would have to be performed using $\boldsymbol{\eta}^{j,k}_{i,t-1|t-1}$ and $\mathbf{P}^{j,k}_{i,t-1|t-1}$ for every possible value of j and k. As a result, the number of possible values of filtered estimates increases directly with time, leading to considerable computational and storage burden if T is large. To circumvent this computational issue, Kim and Nelson (1999) proposed collapsing the $M \times M$ sets of new $\boldsymbol{\eta}^{j,k}_{i,t|t}$ and $\mathbf{P}^{j,k}_{i,t|t}$ at each t as

$$\begin{aligned}
\boldsymbol{\eta}^{k}_{i,t|t} &= \sum_{i=1}^{M} \mathbf{W}_{it} \boldsymbol{\eta}^{j,k}_{i,t|t}, \qquad \mathbf{P}^{k}_{i,t|t} = \sum_{i=1}^{M} \mathbf{W}_{it} \left[\mathbf{P}^{j,k}_{i,t|t} + \left(\boldsymbol{\eta}^{k}_{i,t|t} - \boldsymbol{\eta}^{j,k}_{i,t|t} \right) \left(\boldsymbol{\eta}^{k}_{i,t|t} - \boldsymbol{\eta}^{j,k}_{i,t|t} \right)' \right], \\
\mathbf{W}_{it} &= \frac{Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{it}]}{Pr[S_{it} = k | \mathbf{Y}_{it}]},
\end{aligned} \tag{A.7}$$

where $\mathbf{W}_{it}$ is called the *weighting factor*, the elements of which are computed using the Hamilton filter.

A.4. The Hamilton Filter

The Hamilton filter is also a recursive process and it can be expressed as:

$$\begin{aligned}
&Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{i,t-1}] = Pr[S_{it} = k | S_{i,t-1} = j] \times Pr[S_{i,t-1} = j | \mathbf{Y}_{i,t-1}], \\
&f(\boldsymbol{y}_{it} | \mathbf{Y}_{i,t-1}) = \sum_{k=1}^{M} \sum_{j=1}^{M} f(\boldsymbol{y}_{it} | S_{it} = k, S_{i,t-1} = j, \mathbf{Y}_{i,t-1}) Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{i,t-1}], \\
&Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{it}] = \frac{f(\boldsymbol{y}_{it} | S_{it} = k, S_{i,t-1} = j, \mathbf{Y}_{i,t-1}) Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{i,t-1}]}{f(\boldsymbol{y}_{it} | \mathbf{Y}_{i,t-1})}, \\
&Pr[S_{it} = k | \mathbf{Y}_{it}] = \sum_{j=1}^{M} Pr[S_{i,t-1} = j, S_{it} = k | \mathbf{Y}_{it}],
\end{aligned} \tag{A.8}$$

where $Pr[S_{it} = k|S_{i,t-1} = j]$ are elements of the transition probability matrix shown in Equation (8). $f(\boldsymbol{y}_{it}|S_{it} = k, S_{i,t-1} = j, \mathbf{Y}_{i,t-1})$ is a multivariate normal likelihood function expressed as

$$f(\boldsymbol{y}_{it}|S_{it} = k, S_{i,t-1} = j, \mathbf{Y}_{i,t-1}) = (2\pi)^{-p/2}\left|\mathbf{F}_{it}^{j,k}\right|^{-1/2}\exp\left\{-\frac{1}{2}\left(\boldsymbol{v}_{it}^{j,k}\right)'\left(\mathbf{F}_{it}^{j,k}\right)^{-1}\boldsymbol{v}_{it}^{j,k}\right\}.$$

$f(\boldsymbol{y}_{it}|\mathbf{Y}_{i,t-1})$ in Equation (A.8) is often referred to as the *prediction error decomposition* function (Schweppe, 1965). Taking the log of this value, $\log[f(\boldsymbol{y}_{it}|\mathbf{Y}_{i,t-1})]$, and subsequently summing over $t = 1, \ldots, T$ and then $i = 1, \ldots, n$ yields the overall log-likelihood value, denoted herein as $\log[f(\mathbf{Y}|\boldsymbol{\theta})]$, that can then be maximized using an optimization procedure of choice (e.g., Newton–Raphson) to obtain estimates of $\boldsymbol{\theta}$. Standard errors associated with $\boldsymbol{\theta}$ can be obtained by taking the square root of the diagonal elements of $\mathbf{I}^{-1}$ at the convergence point, where $\mathbf{I}$ is the observed information matrix, obtained by computing the negative numerical Hessian matrix of $\log[f(\mathbf{Y}|\boldsymbol{\theta})]$. Information criterion measures such as the AIC (Akaike, 1973) and BIC (Schwarz, 1978) can be computed using $\log[f(\boldsymbol{y}_{it}|\mathbf{Y}_{i,t-1})]$ as (see p. 80, Harvey, 2001):

$$\begin{aligned}\mathbf{AIC} &= -2\log\left[f(\mathbf{Y}|\boldsymbol{\theta})\right] + 2q\\ \mathbf{BIC} &= -2\log\left[f(\mathbf{Y}|\boldsymbol{\theta})\right] + q\log(nT),\end{aligned}$$

where q is the number of parameters in a model.

Since the prediction error decomposition function in Equation (A.8) is essentially a raw data likelihood function, missing values can be readily accommodated by using only the nonmissing observed elements of $\boldsymbol{y}_{it}$ in computing the prediction errors, $\boldsymbol{v}_{it}^{j,k}$ and their associated covariance matrix. To handle missing data in the EKF, we used the approach suggested by Hamaker and Grasman (2012), that is, to only update the estimates in $\{\boldsymbol{\eta}_{i,t|t}^{j,k}, \mathbf{P}_{i,t|t}^{j,k}, \boldsymbol{\eta}_{i,t|t}^{k}, \mathbf{P}_{i,t|t}^{k}, Pr[S_{i,t-1} = j, S_{it} = k|\mathbf{Y}_{it}], Pr[S_{it} = k|\mathbf{Y}_{it}]\}$ using nonmissing elements from each measurement occasion.

A.5. The Extended Kim Smoother (EKS)

Given estimates from the EKF, the extended Kim smoother (EKS) can be used to obtain more accurate latent variable estimates and regime probabilities based on all observed information from each individual's entire time series. Using $\boldsymbol{\eta}_{i,t|t-1}^{j,k}, \mathbf{P}_{i,t|t-1}^{j,k}, \boldsymbol{\eta}_{i,t|t}^{k}, \mathbf{P}_{i,t|t}^{k}, Pr[S_{it} = k|\mathbf{Y}_{it}]$ and $Pr[S_{it} = k|\mathbf{Y}_{i,t-1}]$, the smoothing procedure can be implemented for $t = T-1, \ldots, 1$ and $i = 1, \ldots, n$ as follows. First, smoothed estimates from regime j to regime k are obtained as

$$\begin{aligned}Pr[S_{i,t+1} = h, S_{it} = k|\mathbf{Y}_{iT}] &= \frac{Pr[S_{i,t+1} = h|\mathbf{Y}_{iT}]Pr[S_{it} = k|\mathbf{Y}_{it}]Pr[S_{i,t+1} = h|S_{it} = k]}{Pr[S_{i,t+1} = h|\mathbf{Y}_{it}]},\\ Pr[S_{it} = k|\mathbf{Y}_{iT}] &= \sum_{h=1}^{M} Pr[S_{i,t+1} = h, S_{it} = k|\mathbf{Y}_{iT}],\\ \boldsymbol{\eta}_{i,t|T}^{k,h} &= \boldsymbol{\eta}_{i,t|t}^{h} + \widetilde{\mathbf{P}}_{t}^{k,h}\left(\boldsymbol{\eta}_{i,t+1|T}^{h} - \boldsymbol{\eta}_{i,t+1|t}^{k,h}\right),\\ \mathbf{P}_{i,t|T}^{k,h} &= \mathbf{P}_{i,t|t}^{h} + \widetilde{\mathbf{P}}_{t}^{k,h}\left(\mathbf{P}_{i,t+1|T}^{h} - \mathbf{P}_{i,t+1|t}^{k,h}\right)\widetilde{\mathbf{P}}_{t}^{k,h},\end{aligned} \tag{A.9}$$

where $\widetilde{\mathbf{P}}_t^{k,h} = \mathbf{P}_{i,t|t}^{k}\mathbf{B}_{h,it}'[\mathbf{P}_{i,t+1|t}^{k,h}]^{-1}$. Similar to the collapsing procedure used in the extended Kim filter, a collapsing process is implemented here as

$$\begin{aligned}\boldsymbol{\eta}_{i,t|T}^{k} &= \sum_{h=1}^{M}\frac{Pr[S_{i,t+1}=h, S_{it}=k|\mathbf{Y}_{iT}]}{Pr[S_{it}=k|\mathbf{Y}_{iT}]}\boldsymbol{\eta}_{i,t|T}^{k,h},\\ \mathbf{P}_{i,t|T}^{k} &= \sum_{h=1}^{M}\frac{Pr[S_{i,t+1}=h, S_{it}=k|\mathbf{Y}_{iT}]}{Pr[S_{it}=k|\mathbf{Y}_{iT}]}\big[\mathbf{P}_{i,t|T}^{k,h} + \big(\boldsymbol{\eta}_{i,t|T}^{k} - \boldsymbol{\eta}_{i,t|T}^{k,h}\big)\big(\boldsymbol{\eta}_{i,t|T}^{k} - \boldsymbol{\eta}_{i,t|T}^{k,h}\big)'\big].\end{aligned} \tag{A.10}$$

Finally, smoothed latent variable estimates and their associated covariance matrix are obtained by summing over the M regimes in effect to yield

$$\begin{aligned}\boldsymbol{\eta}_{i,t|T} &= \sum_{k=1}^{M} Pr[S_{it}=k|\mathbf{Y}_{iT}]\boldsymbol{\eta}_{i,t|T}^{k} \quad \text{and}\\ \mathbf{P}_{i,t|T} &= \sum_{k=1}^{M} Pr[S_{it}=k|\mathbf{Y}_{iT}]\big[\mathbf{P}_{i,t|T}^{k} + \big(\boldsymbol{\eta}_{i,t|T} - \boldsymbol{\eta}_{i,t|T}^{k}\big)\big(\boldsymbol{\eta}_{i,t|T} - \boldsymbol{\eta}_{i,t|T}^{k}\big)'\big].\end{aligned} \tag{A.11}$$

Equations (A.9–A.11) yield three sets of estimates: $\boldsymbol{\eta}_{i,t|T}$, the smoothed latent variable estimates conditional on all observations, the smoothed covariance matrix, $\mathbf{P}_{i,t|T}$, and $Pr[S_{it}=k|\mathbf{Y}_{iT}]$, the smoothed probability for person i to be in regime k at time t.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Anderson, B.D.O., & Moore, J.B. (1979). *Optimal filtering*. Englewood Cliffs: Prentice Hall.

Aptech Systems Inc. (2009). *GAUSS (version 10) [computer software manual]*. Black Diamond: Aptech Systems Inc.

Bar-Shalom, Y., Li, X.R., & Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation: theory algorithms and software*. New York: Wiley.

Boker, S.M., Neale, H., Maes, H., Wilde, M., Spiegel, M., Brick, T., et al. (2011). Openmx: an open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317.

Browne, M.W., & Nesselroade, J.R. (2005). Representing psychological processes with dynamic factor models: some promising uses and extensions of autoregressive moving average time series models. In A. Maydeu-Olivares & J.J. McArdle (Eds.), *Contemporary psychometrics: a festschrift for Roderick P. McDonald* (pp. 415–452). Mahwah: Erlbaum.

Browne, M.W., & Zhang, G. (2007). Developments in the factor analysis of individual time series. In R. Cudeck & R.C. MacCallum (Eds.), *Factor analysis at 100: historical developments and future directions* (pp. 265–291). Mahwah: Erlbaum.

Cacioppo, J.T., & Berntson, G.G. (1999). The affect system: architecture and operating characteristics. *Current Directions in Psychological Science*, *8*, 133–137.

Cattell, R., & Barton, K. (1974). Changes in psychological state measures and time of day. *Psychological Reports*, *35*, 219–222.

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., Nesselroade, J.R., Shifren, K., & McArdle, J. (2004). Dynamic structure of emotions among individuals with Parkinson's disease. *Structural Equation Modeling*, *11*, 560–582.

Chow, S.-M., Tang, N., Yuan, Y., Song, X., & Zhu, H. (2011a). Bayesian estimation of semiparametric dynamic latent variable models using the Dirichlet process prior. *British Journal of Mathematical & Statistical Psychology*, *64*(1), 69–106.

Chow, S.-M., Zu, J., Shifren, K., & Zhang, G. (2011b). Dynamic factor analysis models with time-varying parameters. *Multivariate Behavioral Research*, *46*(2), 303–339.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*(4), 385–396.

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

De Jong, P. (1991). The diffuse Kalman filter. *The Annals of Statistics*, *19*, 1073–1083.

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: Sage.

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*(1), 94–119.
Dolan, C.V., & Van der Maas, H.L.J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*(3), 227–253.
Doornik, J.A. (1998). *Object-oriented matrix programming using Ox 2.0*. London: Timberlake Consultants Press.
Du Toit, S.H.C., & Browne, M.W. (2007). Structural equation modeling of multivariate time series. *Multivariate Behavioral Research*, *42*, 67–101.
Durbin, J., & Koopman, S.J. (2001). *Time series analysis by state space methods*. New York: Oxford University Press.
Elliott, R.J., Aggoun, L., & Moore, J. (1995). *Hidden Markov models: estimation and control*. New York: Springer.
Emotions and Dynamic Systems Laboratory (2010). *The affective dynamics and individual differences (ADID) study: developing non-stationary and network-based methods for modeling the perception and physiology of emotions* Unpublished manual, University of North Carolina at Chapel Hill.
Engle, R.F., & Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, *76*, 774–781.
Ferrer, E., & Nesselroade, J.R. (2003). Modeling affective processes in dyadic relations via dynamic factor analysis. *Emotion*, *3*, 344–360.
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.
Geweke, J.F., & Singleton, K.J. (1981). Maximum likelihood confirmatory factor analysis of economic time series. *International Economic Review*, *22*, 133–137.
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.
Hamaker, E.L., & Grasman, R.P.P.P. (2012). Regime switching state-space models applied to psychological processes: handling missing data and making inferences. *Psychometrika*, *77*(2), 400–422.
Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, *57*, 357–384.
Harvey, A.C. (2001). *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.
Heron, K.E., & Smyth, J.M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behavior treatments. *British Journal of Health Psychology*, *15*, 1–39.
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*(8), 1–29. Available from http://www.jstatsoft.org/v38/i08/.
Kenny, D.A., & Judd, C.M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.
Kim, C.-J., & Nelson, C.R. (1999). *State-space models with regime switching: classical and Gibbs-sampling approaches with applications*. Cambridge: MIT Press.
Kishton, J.M., & Widaman, K.F. (1994). Unidimensional versus domain representative parceling of questionnaire items: an empirical example. *Educational and Psychological Measurement*, *54*, 757–765.
Koopman, S.J. (1997). Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, *92*, 1630–1638.
Kuppens, P., Allen, N.B., & Sheeber, L.B. (2010). Emotional inertia and psychological adjustment. *Psychological Science*, *21*, 984–991.
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*(2), 446–456.
Larsen, R.J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology*, *13*, 25–59.
Li, F., Duncan, T.E., & Acock, A. (2000). Modeling interaction effects in latent growth curve models. *Structural Equation Modeling*, *7*(4), 497–533.
Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data*. New York: Wiley.
Marsh, W.H., Wen, Z.L., & Hau, J.-T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300.
Molenaar, P.C.M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.
Molenaar, P.C.M. (1994a). Dynamic factor analysis of psychophysiological signals. In J.R. Jennings, P.K. Ackles, & M.G.H. Coles (Eds.), *Advances in psychophysiology: a research annual* (Vol. 5, pp. 229–302). London: Jessica Kingsley Publishers.
Molenaar, P.C.M. (1994b). Dynamic latent variable models in developmental psychology. In A. von Eye & C. Clogg (Eds.), *Latent variables analysis: applications for developmental research* (pp. 155–180). Thousand Oaks: Sage.
Molenaar, P.C.M., & Nesselroade, J.R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to block-Toeplitz matrices representing single-subject multivariate time series. *Multivariate Behavioral Research*, *33*, 313–342.
Molenaar, P.C.M., & Newell, K.M. (2003). Direct fit of a theoretical model of phase transition in oscillatory finger motions. *British Journal of Mathematical & Statistical Psychology*, *56*, 199–214.

Muthén, B.O., & Asparouhov, T. (2011). *LTA in Mplus: transition probabilities influenced by covariates* (Mplus Web Notes: No. 13).
Muthén, L.K., & Muthén, B.O. (2001). *Mplus: the comprehensive modeling program for applied researchers: user's guide*. Los Angeles: Muthén & Muthén. 1998–2001.
Nesselroade, J.R., McArdle, J.J., Aggen, S.H., & Meyers, J.M. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D.S. Moskowitz & S.L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: methods and applications* (pp. 235–265). Mahwah: Erlbaum.
Nylund-Gibson, K., Muthen, B., Nishina, A., Bellmore, A., & Graham, S. (2013, 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: Basic Books.
R Development Core Team (2009). *R: a language and environment for statistical computing [computer software manual]*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available from. http://www.R-project.org.
Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(1–sup–6), 1161–1178.
SAS Institute Inc. (2008). *SAS 9.2 help and documentation [computer software manual]*. Cary: SAS Institute Inc.
Sbarra, D.A., & Ferrer, E. (2006). The structure and process of emotional experience following non-marital relationship dissolution: dynamic factor analyses of love, anger, and sadness. *Emotion*, *6*, 224–238.
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*(2), 207–233.
Schumacker, R.E. & Marcoulides, G.A. (Eds.) (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah: Erlbaum.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.
Schweppe, F. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory*, *11*, 61–70.
Tiao, G.C., & Tsay, R.S. (1994). Some advances in non-linear and adaptive modelling in time-series. *Journal of Forecasting*, *13*, 109–131.
Tong, H., & Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B*, *42*, 245–292.
Van der Maas, H.L.J., & Molenaar, P.C.M. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, *99*(3), 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*(1), 7–33.
Visser, I. (2007). *Depmix: an R-package for fitting mixture models on mixed multivariate data with Markov dependencies* (Tech. Rep.). University of Amsterdam. Available from http://cran.r-project.org.
Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scale. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.
Wen, Z., Marsh, H.W., & Hau, K.-T. (2002). Interaction effects in growth modeling: a full model. *Structural Equation Modeling*, *9*(1), 20–39.
Yang, M., & Chow, S.-M. (2010). Using state-space model with regime switching to represent the dynamics of facial electromyography (EMG) data. *Psychometrika*, *74*(4), 744–771. Application and Case Studies.
Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297–330.
Zautra, A.J., Potter, P.T., & Reich, J.W. (1997). The independence of affect is context-dependent: an integrative model of the relationship between positive and negative affect. In M.P. Lawton, K.W. Schaie, & M.P. Lawton (Eds.), *Annual review of gerontology and geriatrics*: *Vol. 17. Focus on adult development* (pp. 75–103). New York: Springer.
Zautra, A.J., Reich, J.W., Davis, M.C., Potter, P.T., & Nicolson, N.A. (2000). The role of stressful events in the relationship between positive and negative affects: evidence from field and experimental studies. *Journal of Personality*, *68*, 927–951.
Zhang, Z., Hamaker, E.L., & Nesselroade, J.R. (2008). Comparisons of four methods for estimating a dynamic factor model. *Structural Equation Modeling*, *15*, 377–402.
Zhang, Z., & Nesselroade, J.R. (2007). Bayesian estimation of categorical dynamic factor models. *Multivariate Behavioral Research*, *42*, 729–756.

Manuscript Received: 24 SEP 2011
Final Version Received: 31 MAY 2012
Published Online Date: 5 MAR 2013