NIH Public Access
Author Manuscript
Br J Math Stat Psychol. Author manuscript; available in PMC 2011 October 23.

Published in final edited form as:
Br J Math Stat Psychol. 2011 February ; 64(Pt 1): 69–106. doi:10.1348/000711010X497262.

Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior

Sy-Miin Chow[1,*], **Niansheng Tang**[2], **Ying Yuan**[1], **Xinyuan Song**[3], and **Hongtu Zhu**[1]
[1]University of North Carolina, Chapel Hill, USA

[2]Yunnan University, Kunming, People's Republic of China

[3]Chinese University of Hong Kong, Ma Liu Shui, People's Republic of China

Abstract

Parameters in time series and other dynamic models often show complex range restrictions and their distributions may deviate substantially from multivariate normal or other standard parametric distributions. We use the truncated Dirichlet process (DP) as a non-parametric prior for such dynamic parameters in a novel nonlinear Bayesian dynamic factor analysis model. This is equivalent to specifying the prior distribution to be a mixture distribution composed of an unknown number of discrete point masses (or clusters). The stick-breaking prior and the blocked Gibbs sampler are used to enable efficient simulation of posterior samples. Using a series of empirical and simulation examples, we illustrate the flexibility of the proposed approach in approximating distributions of very diverse shapes.

1. Introduction

The last two decades have evidenced the emergence of a new class of structural equation models (SEMs), termed latent variable models (LVMs). These models relax the traditional linearity and Gaussian assumptions of the structural equation modelling framework (Jöreskog, 1974) and offer new possibilities for fitting more complex models. Such timely advances include recent developments in fitting dynamic LVMs, defined here as longitudinal models for describing change processes that extend over substantially longer time-spans (e.g., $T > 35$) than those implicated in conventional panel models (typically, $T < 10$).

Some of the seminal work on dynamic LVMs has been spearheaded in part by progress in fitting dynamic factor analysis (DFA) models (Browne & Nesselroade, 2005; McArdle, 1982; Molenaar, 1985; Nesselroade, McArdle, Aggen, & Meyers, 2002; Zhang & Nesselroade, 2007). A DFA model can be viewed as an extension to Cattell, Cattell, and, Rhymer's (1947) P-technique model[1] in which a change model of choice is combined with the standard factor analytic model to account for lagged relationships among factors and manifest variables. The different variants of DFA models proposed in the past two decades have, however, focused exclusively on linear changes (e.g., McArdle, 1982; Molenaar, 1985; Zhang & Browne, 2009; Zhang & Nesselroade, 2007). This is due largely to the many methodological challenges involved in extending DFA models to a nonlinear framework.

© 2010 The British Psychological Society

*Correspondence should be addressed to Dr Sy-Miin Chow, University of North Carolina, CB#3270 Davie Hall, Chapel Hill, NC 27599-3270, USA (symiin@email.unc.edu).

[1]The P-technique model is a common factor model for extracting systematic intra-person patterns from multivariate time series measured on a single individual over time (Jones & Nesselroade, 1990).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

First, the difficulties associated with fitting *cross-sectional* LVMs with nonlinear relationships among factors (as discussed e.g., by Kenny & Judd, 1984) have remained one of most widely investigated research issues in the SEM literature for decades (e.g., Klein & Moosbrugger, 2000; Schumacker, 2002). Generalization to nonlinear dynamic LVMs requires methodological adaptations that can lead to further complication. Second, because different measurement occasions for a single variable are typically treated as different manifest variables in standard SEM practice, the input data covariance matrix for fitting DFA models in SEM software is non-positive definite in cases were the number measurement occasions, T, exceeds the number of participants, N. Even when $T < N$, numerical problems may still arise due to the need to invert a high-dimensional model-implied covariance matrix at each iteration (Hamaker, Dolan, & Molenaar, 2003). Alternative approaches have been proposed to circumvent these issues, including using a block-Toeplitz matrix to replace the data covariance or correlation matrix (Browne & Nesselroade, 2005; Hershberger, Corneal, & Molenaar, 1994; Molenaar, 1985), and using raw data maximum likelihood with special parameter constraints to ensure the positive definiteness of the model-implied covariance matrix (Hamaker *et al.*, 2003). Still, adapting these alternative approaches for use with nonlinear dynamic LVMs is not a straightforward endeavour.

Despite the challenges involved, nonlinear DFA models, as we will illustrate using an empirical example, provide a valuable tool for evaluating substantive questions that are otherwise difficult to test within the linear framework (e.g., Frederickson & Losada, 2005; Gottman, Murray, Swanson, Tyson, & Swanson, 2002). Our aims in the present paper are twofold. First, we seek to present a Bayesian approach to estimating nonlinear dynamic LVMs and propose a novel nonlinear DFA model as a special case. Our second aim is to allow parameters that are of substantive interest in the proposed dynamic model to vary over persons and, more importantly, to conform to non-parametric distributional forms through the use of a non-parametric Dirichlet process (DP) prior. We allow the dynamic but not other modelling parameters (e.g., measurement parameters such as factor loadings) to vary over persons because, in most applications of DFA models, individual differences in the dynamic parameters (i.e., parameters that dictate the nature of the change processes) are often the focus of substantive interest (e.g., Chow, Nesselroade, Shifren, & McArdle, 2004; Ferrer & Nesselroade, 2003). Relaxing the parametric assumptions imposed on these parameters is particularly important from a substantive as well as a practical standpoint. As an example, in fitting DFA models where vector autoregressive (VAR) models or other related variations are used to describe the dynamic relationships among factors, the dynamic parameters (specifically, the auto- and cross-regression parameters) often show complex restrictions of range. That is, moving beyond the stationary ranges may, in certain cases, yield systems that show increasing variance over time2 (Hamilton, 1994; Wei, 1990) – an unlikely scenario in many empirical applications. As a result, the distributions of the dynamic parameters may be non-normal, and even asymmetric.

Prior selection is an important issue in Bayesian data analysis. In many cases, conjugate priors are used for practical reasons to yield posterior distributions of known analytic forms (see Lee, 2007). The multivariate normal distribution, for instance, is one such candidate. Despite its appeal, using multivariate normal distributions as conjugate priors may lead to incorrect posterior inferences in cases where the desired posterior distributions deviate substantially from normality. More recently, there have been a few psychometric applications that utilize non-parametric priors in fitting Bayesian models. Such applications are restricted, however, to cases involving cross-sectional SEMs and item response models

[2]More generally, stationarity refers to the invariance of all statistical properties (e.g., means and covariance functions, in the case of weak stationarity) of a system over time.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(e.g., Duncan & MacEachern, 2008; Lee, Lu, & Song, 2007; Navarro, Griffiths, Steyvers, & Lee, 2006), as well as linear dynamic LVMs (Ansari & Iyengar, 2006). Our proposed approach thus extends previous work that utilizes the DP as a non-parametric prior in Bayesian models to dynamic LVMs with nonlinear change processes at the factor level. Due to the categorical nature of the data used in our empirical example, our approach also generalizes previous approaches of fitting linear DFA models to categorical data (Zhang & Browne, 2009; Zhang & Nesselroade, 2007) to cases involving nonlinear DFA models.

The rest of the present paper is organized as follows. We first summarize features of the semiparametric Bayesian dynamic modelling framework adopted in the paper. We then introduce a motivating empirical example and a novel nonlinear DFA model formulated to test a specific theory of emotions. This is followed by an outline of the broader modelling framework within which our illustrative model can be conceived as a special case. We then present the Markov chain Monte Carlo procedures for fitting the broader model, and results from empirical model fitting and a simulation study. We close with some concluding comments.

2. The Dirichlet process as a non-parametric prior

Non-parametric and semiparametric Bayesian models provide a flexible platform for evaluating the tenability of parametric assumptions in dynamic LVMs. Here, we use the term ‘semiparametric models’ to refer to models with known modelling functions, but unknown distributions for some of the modelling components. In particular, we allow the distributions of the dynamic parameters in our model to conform to non-parametric forms.

When one or more of the distributions in a model of interest are of an unknown form, one common approach is to approximate such distributions non-parametrically by using a finite mixture of parametric distributions, such as a mixture of normal distributions (Lindsay, 1995; McLachlan & Peel, 2000; Sorensen & Alspach, 1971; Titterington, Smith, & Markov, 1985). In some instances, it is of substantive importance to interpret the mixture components as different clusters or classes of individuals with similar characteristics. Unfortunately, when the distributions of interest violate the normality assumption (e.g., in cases involving skewed and heavy-tailed distributions), spurious classes are often detected even when there are no systematic between-class differences (Bauer & Curran, 2003). Non-parametric Bayesian models relax critical dependence on parametric assumptions and, in this way, they help ‘robustify’ parametric models (Antoniak, 1974; Ferguson, 1973). They also serve as a platform for assessing the appropriateness of the parametric assumptions in a model of interest (Dunson, 2008; Karabotsos & Walker, 2009).

Suppose that in an application, an $r \times 1$ vector of modelling components of interest, b_i, for person i, conforms to an unknown distribution such that $b_i \sim \mathscr{P}$, where $\mathscr{P}$ is of an unknown form. In practice, b_i may correspond to a vector of modelling parameters, a vector of latent trait components in an item response model, or a vector of non-Gaussian residuals. One possible non-parametric approach in the Bayesian framework is to specify a DP prior for $\mathscr{P}$ That is, we let $b_i \sim \mathrm{DP}(\alpha P_0)$, where P_0 is a base distribution that serves as a starting-point for constructing the non-parametric distribution. For instance, the (multivariate) normal distribution is a common choice for P_0. The positive quantity α represents the weight a researcher assigns *a priori* to the base distribution, and it reflects the researcher’s certainty of P_0 as the distribution of b_i.

The practical consequences of specifying a DP prior can be better understood in terms of Sethuraman’s (1994) stick-breaking representation. The stick-breaking representation provides an alternative way of conceiving $\mathscr{P} \sim \mathrm{DP}(\alpha P_0)$ as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\mathscr{P}(.)=\sum_{g=1}^{\infty}\pi_g\delta_{Z_g}(.), \quad \text{with } Z_g \overset{\text{i.i.d.}}{\sim} P_0, \tag{1}$$

where $\delta_{Z_g}(.)$ denotes a unit point mass at Z_g, Z_g is the gth matrix consisting of possible values of b_i and π_g is a random probability weight between 0 and 1. Equation (1) thus posits that $\mathscr{P}$consists of a series of point masses – or 'sticks' – of various lengths concentrated at different values of Z_g. For empirical estimation purposes, a truncated DP may be used to approximate equation (1) as (Ishwaran & James, 2001; Ishwaran & Zarepour, 2000; Lee *et al.*, 2007)

$$\mathscr{P}(.) \approx \mathscr{P}_G(.)=\sum_{g=1}^{G}\pi_g\delta_{Z_g}(.), 1 \leq G<\infty, \tag{2}$$

where π_g is obtained as

$$\begin{aligned} \pi_1 &= v_1, \\ \pi_g &= v_g\prod_{l<g}(1-v_l), \quad \text{for } g=2,\ldots,G, \end{aligned} \tag{3}$$

with

$$v_g \overset{\text{i.i.d.}}{\sim} \text{Beta}(1,\alpha), \quad \text{for } g=1,\ldots,G-1, \tag{4}$$

and $\nu_G = 1$ so that $\sum_{g=1}^{G}\pi_g=1$.

To aid interpretation, consider the simple case where $G = 2$. In this case, $\nu_2 = 1$, ν_1 is a random weight between 0 and 1 drawn from the beta distribution; ν_1, the proportion or length of the first stick, is equal to ν_1 and, by equation (3), $\nu_2 = \nu_2(1 - \nu_1) = 1 - \nu_1$. That is, ν_1 is the proportion of a unit probability stick that is broken off and assigned to Z_1, and $1 - \nu_1$ is the remainder of the stick that is assigned to Z_2. In practice, the value of G is either set to a large, predetermined value (e.g., $G \geq 150$) or chosen empirically. For instance, Ishwaran and Zarepour (2000) suggested that the adequacy of the truncation level, G, can be assessed by evaluating moments of the tail probability, $\sum_{g=G}^{\infty}\pi_g$, whose value depends only on G and the hyperparameters that govern the prior distribution of α. A relatively small tail probability is desired, as this indicates that including additional sticks beyond G does not lead to substantial differences in the approximation. Our simulation results showed that a value of G = 300 is more than adequate for the model considered in the present context.

A variety of different distributional forms can be approximated by the discrete probability measure $\mathscr{P}$through different ways of allocating the weights, π_g. The assignment of the weights is, in turn, governed by the random weight, α. Typically, a gamma hyperprior is specified for α and the associated hyperparameters for this gamma prior reflect a researcher's best guess as to how π_g should be distributed, and consequently, how the point masses in equation (2) should resemble the base distribution, P_0. Such prior information is then combined with information from the empirical data to shape the resultant posterior distribution of interest.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To provide a more concrete illustration of the functional role of α, we generated realizations from the DP with different values of α, $G = 300$, $r = 1$ and a univariate $N(0,1)$ base distribution. The associated realizations from $\mathscr{P}_G(.)$ are plotted in Figure 1. The discrete-valued nature of $\mathscr{P}_G(.)$ resulting from the use of the DP prior is portrayed in the plots. In addition, as α increases, samples of ν change from conforming to a roughly uniform distribution to skewed distributions of increasingly restricted ranges dominated by a few weights. Consequently, the distributions of π are characterized by an increasing number of non-zero values of π_g. As a result, progressively more sticks are 'broken off' to form more clusters, and subsequently more unique values of Z_g are observed with non-zero probabilities.

One important fact to note is that as α gets very large, realizations from $\mathscr{P}_G(.)$ show increased resemblance to the standard normal base distribution (see Figure 1d). In fact, in the extreme case where each individual in the sample is assigned to his/her own cluster (not shown here), each individual has one unique set of values for b_i and, consequently, b_i would be distributed as P_0 (Dunson, 2008). In our simulations, we show how the DP can be used as a non-parametric prior for representing a variety of different distributional forms for a set of individual-specific dynamic parameters of interest.

On the technical front, sampling from posterior distributions involving a DP prior can be computationally intensive. One standard approach to enable efficient sampling of b_i within a Markov chain Monte Carlo (MCMC) framework is to represent b_i in terms of a latent variable, L_i, which records each b_i's cluster membership and conveys its values such that $b_i = Z_{L_i}$. In other words, the classification variable L_i is essentially a set of 'pointers' for indicating the values of Z_g associated with person i so that b_i is known when L_i is known. L_i is conditionally independent of Z_g, and it is distributed as

$$L_i|\pi \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{G} \pi_g \delta_g(.). \tag{5}$$

3. Motivating empirical example

Data used in our illustrative example have been previously published elsewhere (see Chow, Ram, Boker, Fujita, & Clore, 2005; Diener, Fujita, & Smith, 1995; Ram *et al.*, 2005). The sample consisted of 179 college students (98 male and 81 female; average age = 20.24 years, *SD* = 1.41) who were asked to provide self-report affect ratings daily for 52 days. After excluding participants with excessive missingness and data anomalies, a total of 174 participants were retained in the final analysis. Four ordinal positive emotion (PE) items and four ordinal negative emotion (NE) items measured on a scale from 1 (= 'none') to 7 (= 'always') were used for model fitting purposes. PE items include joy, contentment, love and affection whereas NE items include unhappiness, anger, depression and anxiety. Individuals' composite PE and NE scores derived from summing these items are plotted in Figure 2.

One linear DFA model that has been used in the past to describe day-to-day changes in PE and NE processes is a process factor analysis model that combines a factor analytic model with a VAR model for describing the relationships among factors (e.g., Chow *et al.*, 2004; Ferrer & Nesselroade, 2003). The dynamics among the latent factors are represented as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\begin{aligned} \mathrm{PE}_{it} &= b_{11}\mathrm{PE}_{i,t-1} + b_{12}\mathrm{NE}_{i,t-1} + \zeta_{1it}, \\ \mathrm{NE}_{it} &= b_{22}\mathrm{NE}_{i,t-1} + b_{21}\mathrm{PE}_{i,t-1} + \zeta_{2it}, \\ (\zeta_{1it}, \zeta_{2it})' &\overset{\text{i.i.d.}}{\sim} N(0, \Psi_\zeta), \Psi_\zeta = \begin{bmatrix} \psi_{\zeta_{11}} & \\ \psi_{\zeta_{12}} & \psi_{\zeta_{22}} \end{bmatrix}. \end{aligned} \tag{6}$$

In this model, PE_{it} and NE_{it} are continuous latent variables representing individual i's underlying PE and NE at time t; ζ_{1it} and ζ_{2it} are process noise terms assumed to follow a multivariate normal distribution; b_{11} and b_{22} are the first-order (or lag-one) autoregressive (AR(1)) parameters, and b_{12} and b_{21} are the lag-one cross-regression parameters. Note that equation (6) only includes auto- and cross-regressive relationships up to the first order because previous inspection of the partial auto- and cross-correlation plots of the PE and NE sum scores from the current data set suggested that the shorter-term dynamics in the data are captured primarily by the lag-one components.3 Higher lag orders and other linear/nonlinear variations of equation (6) can, however, be readily implemented using the broader modelling framework and the associated estimation procedures proposed in the present paper. Thus, higher-order lags can be incorporated as needed in other applications.

Whereas the model shown in equation (6) is a common linear choice for capturing the kinds of relatively rapid fluctuations seen in affect data, we constructed a nonlinear extension to the VAR(1) model in equation (6) to test a theoretically driven model of emotions. Specifically, in their study of affect, Zautra, Reich, Davis, Nicolson, and Potter (2000) proposed a dynamic affect model which postulates that the relative separation between PE and NE changes dynamically as a function of stress. That is, elevated stress is associated with 'shrinkages' in the affective space and the coalescence of positive and negative emotions into a unipolar dimension. Thus, the dynamic affect model suggests that the linkage between PE and NE strengthens with stress but weakens at lower levels of stress. Because we did not have a time-varying indicator of stress, the dynamic affect model will be tested mathematically as

$$\begin{aligned} \mathrm{PE}_{it} &= b_{1,it}\mathrm{PE}_{i,t-1} + \zeta_{1it} = \left[b_{11,i} + b_{12,i}\frac{\exp(\mathrm{NE}_{i,t-1})}{1+\exp(\mathrm{NE}_{i,t-1})} \right] \mathrm{PE}_{i,t-1} + \zeta_{1it}, \\ \mathrm{NE}_{it} &= b_{2,it}\mathrm{NE}_{i,t-1} + \zeta_{2it} = \left[b_{22,i} + b_{21,i}\frac{\exp(\mathrm{PE}_{i,t-1})}{1+\exp(\mathrm{PE}_{i,t-1})} \right] \mathrm{NE}_{i,t-1} + \zeta_{2it}, \\ b_i &\sim \mathrm{DP}(\alpha P_0), b_i = (b_{11,i}, b_{22,i}, b_{12,i}, b_{21,i})', \end{aligned} \tag{7}$$

where $b_{11,i}$ and $b_{22,i}$ are the baseline AR(1) parameters of individual i at extremely low values of $\mathrm{PE}_{i,t-1}$ and $\mathrm{NE}_{i,t-1}$, respectively. If an individual's PE (or NE) was high at time $t-1$, the high PE (NE) affects the subsequent dynamics of the individual's NE (PE) by altering the AR(1) parameter associated with the latter. The parameters $b_{12,i}$ and $b_{21,i}$ are the person-specific deviations in AR(1) parameters for PE_{it} and NE_{it} respectively, when the two factors were at extremely high values at time $t-1$. That is, under high stress, it is natural for an individual to experience a high level of NE. The high NE, in turn, changes the dynamics of PE at the next time point by altering the value of PE's AR(1) parameter by a magnitude of $b_{12,i}$. Whether a reciprocal effect from PE to NE also exists at very high values of PE is reflected in the magnitude of $b_{21,i}$.

In sum, the nonlinear VAR(1) model in equation (7) extends the VAR(1) model into a linear DFA model (see equation (6)) in three ways. First, the AR(1) parameter of each factor is

[3]There was also evidence for weekly cycles in the data. This is beyond the focus of the present paper, but interested readers can refer elsewhere (Chow *et al.*, 2005; Chow, Hamaker, & Allaire, 2009; Ram *et al.*, 2005) for modelling options that do account for cyclic dynamics.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

now a time-varying function of the other factor from a previous time point, so that the dynamic model becomes nonlinear. Second, we allow four of the dynamic parameters, namely, $b_i = (b_{11,i}, b_{22,i}, b_{12,i}, b_{21,i})$ to vary over persons. Third, we use a DP prior for b_i to allow these parameters to conform to non-parametric distributional forms.

4. Dynamic latent variable modelling framework

The nonlinear VAR(1) model in equation (7) can be formulated as a special case of a nonlinear state-space model. The state-space representation provides a flexible framework for representing different dynamic processes and subsumes many dynamic and time series models as special cases. Because latent variables are still the focus of our formulation, we refer to our modelling framework as the dynamic latent variable modelling framework. The resultant modelling framework comprises a dynamic model and a measurement model, which will be described next.

4.1. Dynamic model

The general dynamic model on which the proposed estimation procedures are based is expressed as

$$\eta_{it}=f_t(\eta_{i,t-1}, b_i, \theta_\eta)+\zeta_{it}, \zeta_{it} \overset{\text{i.i.d.}}{\sim} N(0, \Psi_\zeta), b_i|\alpha, P_0 \overset{\text{i.i.d.}}{\sim} \mathrm{DP}(\alpha P_0), \tag{8}$$

where i is the person index and t is the time index; η_{it} is a $w \times 1$ vector of latent variables of interest, ζ_{it} is a vector of process noise components and $f_t(.)$ is a vector of time-varying, differentiable linear or nonlinear functions describing the latent variables at time t in terms of three components (1) their previous history at time $t - 1$, (2) b_i, an $r \times 1$ vector of person-specific parameters, and (3) a vector of parameters that is held invariant over time and persons, denoted as θ_η. In our motivating example, $\eta_{it} = (\mathrm{PE}_{it}, \mathrm{NE}_{it})'$, $b_i = (b_{11,i}, b_{22,i}, b_{12,i}, b_{21,i})'$ and $\theta_\eta = (\Psi_\zeta)$. We specify a DP prior for b_i, and set the base distribution, P_0, to be an r-variate normal distribution with mean vector μ_Z and covariance matrix Ψ_Z.

Beyond our illustrative model, a variety of other dynamic functions can be readily specified as special cases of equation (8). Some examples include unobserved components models with cyclic, seasonal, and irregular components (Durbin & Koopman, 2001; Harvey, 2001), non-parametric spline models (De Jong & Mazzi, 2001), and exact discrete time models (Harvey, 2001). VAR models of higher lags and other related vector autoregressive moving average (VARMA) extensions can also be formulated as special cases of equation (8) by expanding the size of η_{it} to include higher-order lag components.

4.2. Measurement model

Motivated by our empirical data of interest, we now consider a measurement model for ordinal data and elaborate briefly, where appropriate, on how the proposed framework can be extended to include mixed responses. In the conventional state-space framework, the measurement model is used to specify the relationships among a set of latent and manifest variables. In cases involving ordinal manifest data, the same measurement model is defined in terms of a vector of underlying, continuous latent variables as

$$y^*_{it}=\mu+\Lambda\eta_{it}+\varepsilon_{it}, \varepsilon_{it} \overset{\text{i.i.d.}}{\sim} N(0, \Psi_\varepsilon), t=1, 2, \ldots, T; i=1, 2, \ldots n, \tag{9}$$

where y^*_{it} is a $p \times 1$ vector of unobserved continuous response variables underlying a $p \times 1$ vector of manifest ordinal data, y_i, and ε_{it} is a vector of uniquenesses. The vector of time-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

invariant parameters in the measurement model, denoted by θ_ε, includes elements in the vector of intercepts, μ, the $p \times w$ matrix of factor loadings, λ, and Ψ_ε, the $p \times p$ covariance matrix of ε_{it}.

The unobserved continuous latent vector, y^*_{it}, is linked to the manifest ordinal data, y_{it} by

$$y_{it,k}=s \iff \tau_{k,s-1}<y^*_{it,k} \leq \tau_{k,s}, k=1,\ldots,p; s=1,\ldots M, \tag{10}$$

where $\tau_{k,h}$ is a set of threshold values held invariant across persons and time, with

$$-\infty=\tau_{k,0}<\tau_{k,1}<\tau_{k,2}<\cdots<\tau_{k,M-1}<\tau_{k,M}=+\infty. \tag{11}$$

That is, for the kth ordinal variable, $y_{it,k}$, with M categories, there are $M - 1$ threshold parameters. The underlying variable approach summarized in equations (9) – (11) has been shown to be equivalent to the generalized latent trait approach often adopted in the item response theory framework (Bartholomew & Knott, 1999; Jöreskog & Moustaki, 2001).[4] Since only ordinal information is used to identify y^*_{it}, additional constraints need to be imposed to identify the model. For instance, the lowest and highest threshold values of each item can be fixed, as opposed to being freely estimated (Lee & Zhu, 2000). These are the identification constraint adopted in the present study.

5. Bayesian estimation procedures

Throughout, we define $\theta = (\theta_\varepsilon, \theta_\eta)$, $\pi = (\pi_1, \ldots, \pi_G)$, $L = (L_1, \ldots, L_n)$, $H = (\eta_1, \ldots, \eta_n)$ with $\eta_i = (\eta_{i1}, \ldots, \eta_{iT})'$ as the array of latent variables, $\tau = (\tau_{1,2}, \ldots, \tau_{1,M-2}, \ldots, \tau_{p,M-2})$ as a vector of threshold parameters and $b = (b_1, \ldots, b_n)$ as an array of all person-specific parameters. In addition, we denote $Y = (Y_1, Y_2, \ldots, Y_n) \triangleq \{Y_{\text{obs}}, Y_{\text{mis}}\}$, where Y_{obs} is a data array with complete ordinal observations from all persons and time points, Y_{mis} is a data array that includes all missing observations from all persons and time points, with $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})' \triangleq \{y_{i,\text{obs}}, y_{i,\text{mis}}\}$ being a data matrix that includes person i's complete and missing manifest ordinal data up to time T. Their corresponding continuous counterparts are denoted by $Y^*=(Y^*_1,\ldots,Y^*_n) \triangleq \{Y^*_{\text{obs}}, Y^*_{\text{mis}}\}$ with $Y^*_i=(y^*_{i1},\ldots,y^*_{iT})' \triangleq \{y^*_{i,\text{obs}}, y^*_{i,\text{mis}}\}$. Assuming that the data are missing at random with an ignorable missingness mechanism (Little & Rubin, 1987), estimation of all parameters and latent variables will be based only on the observed data set, Y_{obs}.

The general model summarized in equations (7)–(11) can be rewritten as

$$\begin{gathered}(y^*_{it}|\theta_\varepsilon,\eta_{it}) \overset{\text{ind}}{\sim} N_p(\mu+\Lambda\eta_{it},\Psi_\varepsilon), (\eta_{it}|\eta_{i,t-1},b_i=Z_{L_i},\theta_\eta) \overset{\text{ind}}{\sim} N_w[f_t(\eta_{i,t-1},b_i=Z_{L_i},\theta_\eta),\Psi_\zeta],\\ (v_g|\alpha) \overset{\text{i.i.d.}}{\sim} \text{Beta}(1,\alpha), \text{ for } g=1,\ldots,G-1 \text{ and } v_G=1,\\ (\pi_g|v)\sim\text{SB}(v),\ Z_g \overset{\text{i.i.d.}}{\sim} N_r(\mu_Z,\Psi_Z),\ (L_i|\pi) \overset{\text{i.i.d.}}{\sim} \sum_{g=1}^{G}\pi_g\delta_g(.),\end{gathered}$$

where SB(.) denotes the stick-breaking process expressed in equation (3).

[4]The measurement framework adopted in the present paper can be readily generalized to applications involving data measured on mixed scales. In this case, the measurement vector y_{it} is further partitioned into portions with p_1 continuous variables, p_2 binary variables and $p - p_1 - p_2$ ordinal variables. This yields $y_{it} = [y_{it,1}, \ldots, y_{it,p_1}, y_{it,p_1+1}, \ldots, y_{it,p_1+p_2}, y_{it,p_1+p_2+1}, \ldots, y_{it,p}]'$. The appropriate density specification can then be defined for each portion of these data to yield a final measurement model (see Lee & Zhu, 2000).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We specified the prior distributions for our modelling components as

$$\begin{aligned} &\tau \sim c, && (\mu|\mu_0, \Sigma_0) \sim N_p[\mu_0, \Sigma_0], \\ &\left(\Psi_\zeta|w_0, \Psi_{\zeta_0}^{-1}\right) \sim \mathrm{IW}_w\left[w_0, \Psi_{\zeta_0}^{-1}\right], && (\Lambda_k|\psi_{\varepsilon_k}) \sim N_w[\Lambda_{0_k}, \psi_{\varepsilon_k} H_{0\Lambda_k}], \\ &\left(\psi_{\varepsilon_k}^{-1}\right) \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}[\alpha_{0\varepsilon_k}, \beta_{0\varepsilon_k}], && (\mu_Z|\Psi_{\mu_Z}) \sim N_r(\mu_{Z_0}, \Psi_{\mu_Z}), \\ &\left(\Psi_{Z_k}^{-1}|c_1, c_2\right) \overset{\text{i.i.d.}}{\sim} \mathrm{Gamma}(c_1, c_2), && (\alpha|a_1, a_2) \sim \mathrm{Gamma}(a_1, a_2), \end{aligned} \tag{12}$$

where c is a constant used to specify a diffuse prior for the threshold parameters and Λ_k' denotes the kth row of Λ (for $k = 1, \ldots, p$). The components $\mu_0, \Sigma_0, w_0, \Psi_{\zeta_0}^{-1}, \Lambda_{0_k}, H_{0\Lambda_k}, \alpha_{0\varepsilon_k}, \beta_{0\varepsilon_k}, \mu_{Z_0}, \Psi_{\mu_Z}, c_1, c_2, a_1$ and a_2 are all hyperparameters whose values are assumed to be known. Thus, with the exception of the threshold parameters, which were assigned a non-informative prior, standard conjugate priors were specified for all other parametric components in the model. Such priors were thought to provide a reasonable representation of the characteristics of these components, and the associated hyperparameters can be determined in a relatively straightforward manner based on previous applications. Details of our hyperparameter choices are discussed later as we present the empirical results.

The Gibbs sampler will be used to simulate a sequence of random observations from the joint posterior distribution $p(\pi, Z, L, \mu_Z, \Psi_Z, \alpha, H, \theta, \tau, Y^*, Y_{\text{mis}}|Y_{\text{obs}})$, where $Z = (Z_1, \ldots, Z_G)$ is a $G \times r$ matrix containing values of b_i. At the first iteration, samples of Z are drawn from the base distribution, P_0. Then, starting from $\pi^{(0)}, Z^{(0)}, L^{(0)}, \mu_Z^{(0)}, \Psi_Z^{(0)}, \alpha^{(0)}, H^{(0)}, \theta^{(0)}, \tau^{(0)}$ and $Y^{*(0)}$, the Gibbs sampler involves sampling sequentially, for the next iteration, q+1 (until the last iteration, Q), as follows:

- **a.** Generate $\mu_Z^{(q+1)}$ from $p\left(\mu_Z|Z^{(q)}, \Psi_Z^{(q)}\right)$.
- **b.** Generate $\Psi_Z^{(q+1)}$ from $p\left(\Psi_Z|Z^{(q)}, \mu_Z^{(q+1)}\right)$.
- **c.** Generate $\alpha^{(q+1)}$ from $p(\alpha|\pi^{(q)})$.
- **d.** Generate $(\pi^{(q+1)}, Z^{(q+1)})$ from $p\left(\pi, Z|L^{(q)}, \mu_Z^{(q+1)}, \Psi_Z^{(q+1)}, \alpha^{(q+1)}, H^{(q)}, \theta^{(q)}, \tau^{(q)}, Y^{*(q)}, Y_{\text{obs}}\right)$.
- **e.** Generate $L^{(q+1)}$ from $p(L|\pi^{(q+1)}, Z^{(q+1)}, \tau^{(q)}, Y^{*(q)}, \theta^{(q)}, H(q), Y_{\text{obs}})$.
- **f.** Generate $H^{(q+1)}$ from $p(H|\tau^{(q)}, Y^{*(q)}, \theta^{(q)}, Y_{\text{obs}}, b^{(q+1)})$.
- **g.** Generate $\theta^{(q+1)}$ from $p(\theta|\tau^{(q)}, Y^{*(q)}, H^{(q+1)}, Y_{\text{obs}})$.
- **h.** Generate $Y_{\text{mis}}^{*(q+1)}$ from $p\left(Y_{\text{mis}}^*|\theta^{(q+1)}, H^{(q+1)}\right)$.
- **i.** Generate $\left(\tau^{(q+1)}, Y_{\text{obs}}^{*(q+1)}\right)$ from $p(\tau, Y_{\text{obs}}^*|\theta^{(q+1)}, H^{(q+1)}, Y_{\text{obs}})$.

In sum, the Gibbs sampler essentially involves sampling from a series of conditional distributions while each of the modelling components is updated in turn. The conditional distributions needed to implement steps (a)–(i) are summarized in the Appendix. In cases where the associated conditional distributions are of known analytic forms, sampling from these distributions is straightforward (including steps (a)–(c), (e), (g) and (h)). In other cases, Metropolis–Hasting (MH) algorithms are used within the Gibbs sampler (e.g., steps (d), (f) and (i)) to allow sampling from the corresponding conditional distributions. Steps (a)–(e), in particular, are all part of a blocked Gibbs sampling procedure for deriving posterior samples

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

of b_i. Furthermore, no additional step is included to sample the missing ordinal data, Y_{mis}. Because of the assumption of missingness at random, all that is needed are samples of the missing latent continuous data array, Y^*_{mis}, to yield posterior samples of $Y^*=\{Y^*_{\text{obs}}, Y^*_{\text{mis}}\}$, obtained respectively from steps (i) and (h) and subsequently utilized in other sampling steps.

6. Empirical results

The nonlinear DFA model (see equations (7) and (9)–(11)) was fitted to the empirical data. Four ordinal items were used to identify PE_{it}, individual i's latent positive emotion, and four ordinal items were used to identify NE_{it}, individual i's latent negative emotion. We ran three independent Markov chains with different starting values and yielded similar results. We report here the results as aggregated across the three chains.

For identification purposes, we set $\tau_{k,1}$ to $\Phi^{-1}(n_{k,1}/n)$ and $\tau_{k,6}$ to $\Phi^{-1}(n^{-1}\sum_{h=1}^{6} n_{k,h})$, where $\Phi^{-1}(.)$ denotes the inverse standard normal cumulative distribution function and $n_{k,h}$ represents the number of responses endorsing category h on item k across all persons and time points. The hyperparameter values of the prior distributions (see equation (12)) were specified as follows. A diffuse prior was specified for all of the threshold parameters, so c can be set to any arbitrary constant value without affecting the resultant posterior distributions of the threshold parameters. We set μ_0 to an 8×1 vector of ones, Σ_0 to I_8, w_0 to 10, and $\Psi^{-1}_{\zeta_0}$ to

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}.$$

Further, letting k and j denote the row and column, respectively, of the factor loading matrix, Λ, $\Lambda_{0kj} = 0.8$ for each of the freed elements in the factor loading matrix ($j = 1$ for $k = 2$, 3, and 4; $j = 2$ for $k = 6$, 7, and 8) and $H_{0\Lambda kj} = 1$ for each of the freed elements in the factor loading matrix. For the conjugate priors of the measurement error variances, we set $\alpha_{0\varepsilon k}$ to 8 and $\beta_{0\varepsilon k}$ to 10 to yield variance values that were relatively large and diffuse.

To ensure that the approximations obtained from posterior samples of the nonparametric components were not biased by the choice of our hyperparameters, we allowed some of the hyperparameters that governed the base distribution to vary randomly across the three independent Markov chains. Specifically, based on previous results from fitting VAR(1) models to sum scores from the present data set with parametric assumptions, we set $\mu_{Z_{0j}}$ to 0.5 for $j = 1$ and 2 (i.e., corresponding to $b_{11,i}$ and $b_{22,i}$) and to −0.1 for $j = 3$ and 4 (i.e., corresponding to $b_{12,i}$ and $b_{21,i}$). Elements in $\Psi_{\mu Z}$ were sampled randomly from a *Unif* (1, 10) distribution for $b_{11,i}$ and $b_{22,i}$ and a *Unif* (1, 12) distribution for $b_{12,i}$ and $b_{21,i}$ for each independent Markov chain. Furthermore, we set c_1 to 10, and allowed c_2 to be sampled randomly from a *Unif* (3, 7) distribution for elements in Ψ^{-1}_{Z} that corresponded to $b_{11,i}$ and $b_{22,i}$, and from a *Unif* (0.5, 4) distribution for those that corresponded to $b_{12,i}$ and $b_{21,i}$. Note that the different hyperparameter choices for elements in b_i were necessary because we expected $b_{12,i}$ and $b_{21,i}$ to conform to much narrower ranges than the baseline auto regression parameters in stable systems (namely, systems that do not show increasing variance over time). With respect to hyperparameters for the prior distribution of α, we set a_1 to 250 and a_2 to 1 to yield large values of α (and consequently, more unique b_i values) to capture some of the more subtle individual differences in these dynamic parameters.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We computed the estimated potential scale reduction (EPSR; Gelman, 1996) values based on the three independent Markov chains, each initialized with different starting values for the person-invariant parameters, and different hyperparameter values as described above. The EPSR values for all person-invariant parameters became less than 1.2 and the corresponding parameter estimates from different chains stabilized in less than 200 iterations. To allow sufficient burn-in iterations to recover the shapes of the person-specific parameters, we allowed for 18,000 burn-in iterations.

The posterior predictive probability (Gelman, Meng, & Stern, 1996; Lee & Zhu, 2000; Meng, 1994) of the fitted model computed using posterior samples of the continuous data array, Y^*, averaged .67, indicating that the model provided a reasonable fit to the data. Estimates of all person- and time-invariant parameters obtained using 4,000 additional iterations after burn-in are summarized in Table 1. These estimates were averages taken over the three independent chains.

Consistent with previously published results (Ram *et al.*, 2005), all the PE and NE indicators showed relatively little differentiation (i.e., distances between thresholds were small) among the middle categories. High levels of NE were relatively rare, resulting in the threshold values of the NE items clustering around relatively high magnitudes. The baseline levels (as indicated by values of μ) were estimated to be close to zero, with the intercepts of the NE continuous data being significantly greater than zero. At the factor level, PE and NE were found to be weakly negatively correlated, as would be expected based on previous findings concerning the structure of emotions.

We obtained estimates of each person's parameters in b_i, denoted below as $\hat{b}_i$, by averaging posterior samples from the distribution $p(Z_{L_i} | L_i, \mu_Z, \Psi_Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$ as

$$\widehat{b_i} = Q^{-1} \sum_{q=1}^{Q} Z_{L_i,} \tag{13}$$

where $q = 1$ denotes the first iteration after burn-in and Q is the maximum number of iterations. Distributions of these $\hat{b}_i$ estimates across participants (for $i = 1, \ldots, n$) were the focus of our interest.5 These estimates were observed to show very similar distributional forms across the three Markov chains even with different hyperparameter choices. Matrix scatterplots of the $\hat{b}_i$ estimates averaged across the three chains and the corresponding histograms are shown in Figure 3. Based on the plots, distributions of the person-specific autoregressive parameters, $\hat{b}_{11,i}$ and $\hat{b}_{22,i}$, were both highly skewed. Most individuals' estimated autoregressive parameters lay in the moderate to high range (from around 0.5 to close to 1.0), with a small number of individuals showing near-zero autoregressive estimates. That is, the latter subgroup of individuals tended to show very little stability or continuity in their PE and NE from day to day.

The 90% credible interval6 associated with $\hat{b}_{12,i}$ included zero but that for $\hat{b}_{21,i}$ was barely above zero. Thus, on average, the present sample showed unidirectional coupling from PE to NE when PE is at high values, but no coupling in the reverse direction when NE was high. This finding provided partial support for the dynamic affect model postulated by Zautra *et*

[5]Of course, there may be instances where researchers are also interested in obtaining other summary statistics (e.g., standard deviation) from the person-specific distribution, namely, $p(Z_{L_i} | L_i, \mu_Z, \Psi_Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$, to make inferences at the individual level. This is, however, not the focus of the present empirical illustration.

[6]Note that our choice to report the 90%, as opposed to the 95%, credible intervals was largely arbitrary. The results did not differ, however, when the 95% credible intervals were evaluated.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

al. (2000). That is, we found that there was an increased linkage between PE and NE at high values of PE, but the association was driven more by the lagged influence of PE on NE.

Allowing the distributions of the person-specific parameters to deviate from normality also helped reveal novel interrelationships among the auto- and cross-regression parameters. First, although most individuals' $\hat{b}_{12,i}$ estimates clustered around zero, individuals who showed strong continuity in PE (with $b_{11,i}$ estimates close to or above 1.0) tended to also show negative lagged influence from NE to PE. That is, high values of NE served to dampen the fluctuations in PE and bring it back towards its baseline.

A slightly different scenario was observed in the individuals' NE regulation. Almost all individuals showed a positive cross-regression weight from PE at $t-1$ to NE at time t. The positive cross-regression weights suggested that if an individual experienced extremely high PE yesterday, the high PE tended to delay the individual's NE from returning to its baseline. This may, for instance, cause an individual's NE to continue to wander around more extreme (e.g., extremely low) values for a longer period of time. Compared with individuals with high stability in NE, individuals who showed lower stability in NE (i.e., $\hat{b}_{22,i}$ estimates were low) showed greater tendency in this regard (i.e., showing higher positive $\hat{b}_{21,i}$ estimates).

To evaluate whether the non-parametric DP prior yielded any practical differences in the estimation results, we replicated the analysis by using a multivariate normal distribution as a parametric prior for the distribution of b_i. Specifically, we specified the parametric prior as

$$(b_i|\mu_{Z_{\text{Norm}}}, \Psi_{Z_{\text{Norm}}}) \sim N_r(\mu_{Z_{\text{Norm}}}, \Psi_{Z_{\text{Norm}}}),$$

with $\mu_{Z_{\text{Norm}}} \sim N_r([0.8, 0.6, -0.03, 0.26]', \Psi_{Z_{\text{Norm}_0}})$, based on the semiparametric empirical results. To allow for variability in the hyperparameter values across the MCMC chains, $\Psi_{Z_{\text{Norm}_0}}$ was specified to be a diagonal matrix with the first two elements sampled from a *Unif* (0.3, 1) distribution and the last two elements sampled from a *Unif* (0.2, 0.8) distribution for each of the three independent Markov chains. Furthermore, we let $\Psi^{-1}_{Z_{\text{Norm}_k}} \sim \text{Gamma}(d_1, d_2)$ with d_1 set to 2 and d_2 sampled randomly from a *Unif* (0.1, 0.5) distribution for $k = 1$ and 2, and a *Unif* (0.05, 0.3) distribution for $k = 3$ and 4 for each of the independent Markov chains. As in the non-parametric case, these values were selected based on a prior expectation of the ranges of the parameters in b_i in stable systems. The resultant conditional distribution $p(b|\mu_{Z_{\text{Norm}}}, \Psi_{Z_{\text{Norm}}}, \tau, Y^*, \theta, H, Y_{\text{obs}})$ was also non-standard and we used an MH step similar to that for deriving posterior samples of $p(Z|L, \mu_Z, \psi_Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$ to obtain posterior samples from this conditional distribution (see equation (A7) in the Appendix).

The $\hat{b}_i$ estimates obtained from using the multivariate normal prior are plotted in Figure 4, with density plots of the corresponding non-parametric posterior samples overlaid on the histograms in the diagonal panels. It can be seen that, compared with the non-parametric posterior samples, the parametric posterior samples of $\hat{b}_i$ did not capture the full ranges of the $\hat{b}_i$ estimates – in particular, the heavy-tailed nature of $\hat{b}_{11,i}$ and the platykurtic but asymmetric nature of $\hat{b}_{22,i}$. Evaluation of the summary statistics of the posterior samples further revealed that whereas the means of the posterior samples were generally close under both prior specifications, the standard deviations of the estimates were about twice as large in the non-parametric case as in the parametric case. This has some parallels to asymptotic results within the linear frequentist framework, where violation of distributional assumptions has been shown to affect the standard error estimates, but not so much the corresponding point estimators (Ljung & Caines, 1979). Imposing the normality assumption also led to

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

misleading conclusions concerning the interrelationships among the $\hat{b}_i$ estimates. For instance, whereas $\hat{b}_{12,i}$ was found to show a negative relationship with $\hat{b}_{11,i}$ and a quadratic relationship with $\hat{b}_{22,i}$ with the use of the DP prior (see Figure 3, third panels in rows 1 and 2), no such relationships were observed when the multivariate normal prior was used (see Figure 4, third panels in rows 1 and 2).

Only minor differences were observed in the person- and time-invariant parameter estimates when the multivariate normal prior was used (see Table 2). All the estimates and their associated statistics were largely similar, with the exception of the parameters in μ, the intercepts of the continuous data array. Most of the elements in μ were greatly reduced in magnitude and included zero in most of their 90% credible intervals. Such differences likely reflected the discrepancies in the predicted trends of the individuals, due presumably to the greater restrictions of range in the $\hat{b}_i$ estimates in the parametric than in the non-parametric case.

To summarize, by using the DP as a non-parametric prior for parameters in a nonlinear DFA model, we found that the distributions of the two autoregressive parameters, $b_{11,i}$ and $b_{22,i}$, did deviate substantially from normality. Using other parametric (e.g., normal) distributions to approximate these skewed distributions did not reveal the full range of individual differences in these parameters and led to misleading conclusions concerning some of the more complex, and often nonlinear, interrelationships among the $\hat{b}_i$ estimates.

7. Simulation study

To better understand the performance of the proposed procedures under known population conditions, we generated data using the nonlinear DFA model with different distributions for b_i but with approximately the same complete sample size and number of time points as our empirical example, with $n = 170$ and $T = 50$. The true parameters in our simulations were chosen to mirror (though not completely identical to) the estimates obtained from our empirical example. A total of 100 Monte Carlo replications was conducted for each of the simulation conditions described below.

Parameters in the measurement model (see equations (9)–(11)) were identical in all our simulation models and they were set to the values of

$$\begin{aligned}
\Lambda' &= \begin{bmatrix} 1.0 & 0.8 & 0.8 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1.0 & 0.8 & 0.8 & 0.8 \end{bmatrix}, \\
\Psi_\varepsilon &= 0.8 \times I_8, \\
\mu &= (0, 0, \ldots, 0), \\
\tau_k &= (-3.0, -2.0, -1.0, 0.0, 0.5, 2.0), \quad \text{for } k=1, \ldots, 4, \\
&= (-1.0, -0.5, 0.0, 1.0, 1.5, 2.0), \quad \text{for } k=5, \ldots, 8.
\end{aligned} \tag{14}$$

For identification purposes, each factor's loading on the first indicator was set to the true value of 1.0 in model fitting. In all conditions, the true process noise covariance matrix was set to

$$\Psi_\zeta = \begin{bmatrix} 1 & \\ -0.3 & 1 \end{bmatrix}.$$

We tested the effectiveness of using the DP prior to approximate three sets of distributional conditions.

Condition 1. Here, we defined the distributions of b_i as

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\begin{aligned} &b_{11,i} \sim \text{Beta}(3,8), && b_{22,i} \sim \text{Beta}(3,8), \\ &b_{12,i} \sim N(-0.15, 0.001), && b_{21,i} \sim N(-0.15, 0.001), \end{aligned} \tag{15}$$

where $b_{11,i}$, $b_{22,i}$, $b_{12,i}$, and $b_{21,i}$ are as defined in equation (7). This condition was designed to generate positively skewed distributions for the baseline autoregressive parameters, $b_{11,i}$ and $b_{22,i}$.

Condition 2. Here, we specified some of the distributions to be bimodal:

$$\begin{aligned} &b_{11,i} \sim Unif(0.2, 0.4), \text{ for } i=1,\ldots,n/2, && b_{11,i} \sim Unif(0.7, 0.8), \text{ for } i=n/2+1,\ldots,n, \\ &b_{22,i} \sim Unif(0.2, 0.4), \text{ for } i=1,\ldots,n/2, && b_{22,i} \sim Unif(0.7, 0.8), \text{ for } i=n/2+1,\ldots,n, \\ &b_{12,i} \sim N(-0.15, 0.001), && b_{21,i} \sim N(-0.15, 0.001). \end{aligned} \tag{16}$$

Condition 3. Here, we specified the distributions of b_i to be

$$\begin{aligned} &b_{11,i} \sim N(0.6, 0.005), && b_{22,i} \sim N(0.6, 0.005), \\ &b_{12,i} \sim N(-0.15, 0.001) && b_{21,i} \sim N(-0.15, 0.001). \end{aligned} \tag{17}$$

This condition was used to illustrate that even when the normality assumption holds, the DP prior is still general enough to capture characteristics of the multivariate normal distribution as a special case.

The same starting values were used in all conditions. All freed elements in Λ were set to 0.5. Initial Ψ_ε was set to I_8 and μ was set to (0.5, 0.5, …, 0.5). Initial values for τ_k were set to (−3.0, −1.5, −0.5, 0.5, 1.0, 2.0) for k = 1, …, 4 and (−1.0, 0.0, 0.5, 0.8, 1.3, 2.0) for k = 5, …, 8. Starting values for the process noise covariance matrix were set to

$$\Psi_\zeta = \begin{bmatrix} 1.2 & \\ 0.5 & 0.8 \end{bmatrix}$$

In addition, all values in the latent variable vector, η_{it}, were set to 0 for all persons and time points, and initial values of b_i were sampled randomly from $N_r(\mu_{Z_0}, I_r)$.

The same prior distributions and hyperparameters used in the empirical application were adopted in the simulation study with some minor adaptations. Specifically, based on characteristics of the distributions of b_i in each simulation condition (see equations (15)–(17)) and the acceptance rates for the MH step for drawing posterior samples from $p(Z|L, \mu_Z, \Psi_Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$, we specified Ψ_{μ_Z} to be 1.0 and 15.0 and c_2 to be 14.0 and 0.4, respectively, for the first two and last two elements of b_i in the first condition. In the second condition, we changed c_2 to 4.0 and 0.4, respectively, for the first two and last two elements of b_i while keeping other hyperparameters to be the same as those in condition 1. Finally, in condition 3, we changed c_2 to 1.0 and 0.3 and Ψ_{μ_Z} to 1.0 and 20.0, respectively, for the first two and last two elements of b_i. In the first two conditions, the tails of the distributions of the b_i parameters extended relatively far away from the modes of the distributions compared with condition 3. Thus, the hyperparameters were modified accordingly to yield better approximation in the tail areas. In the empirical application, these hyperparameters were drawn randomly from ranges that were broad enough to encompass all the hyperparameter values noted here.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

We organized our simulation results into three sections to summarize results pertaining to factor score (i.e., state) estimation; estimation of all time- and person-invariant parameters; and estimation of all person-specific parameters. A total of 20,000 burn-in iterations was used in each Monte Carlo replication. We used a relatively high number of burn-in iterations due to the greater difficulties involved in recovering the person-specific distributions for $b_{12,i}$ and $b_{21,i}$, whose impact on the system's dynamics was only evident at high values of the latent variables. It is thus harder to recover the shapes of these parameters' distributions given the moderate sample sizes of $n = 170$ and $T = 50$.

7.1. Factor score or state estimation

For illustration purposes, the estimated and true values of factor 1 from one randomly selected case during one particular Monte Carlo run are plotted in Figure 5. The estimates plotted in Figure 5 were the means of the posterior samples drawn from the posterior density of the selected individual, $p(H_i|\tau, Y^*, \theta, Y_{\text{obs}})$, for each time point after the burn-in iterations.

It can be seen that the proposed algorithm was able to recover the true factor scores accurately across all the conditions. Although factor score estimation is not an issue of primary interest in our simulation or empirical examples, it is an important component in studies where the primary interest is to obtain longitudinal factor score estimates or estimates of time-varying parameters that are represented as latent variables (see Young, Pedregal, & Tych, 1999).

7.2. Time- and person-invariant parameters

All time- and person-invariant parameter estimates are summarized in Tables 3–5 for conditions 1, 2, and 3, respectively. Included in the tables are the biases (for parameter l, bias$=\Sigma_{r=1}^{100}\bar{\theta}_{l,r}/100 - \theta_l$ where θ_l is the true value for parameter l and $\bar{\theta}_{l,r}$ is the average of the Gibbs samples of parameter l during the rth Monte Carlo run after burn-in), root mean squared errors (RMSE, given by $\sqrt{\Sigma_{r=1}^{100}(\bar{\theta}_{l,r} - \theta_l)^2/100}$), empirical standard deviations of each parameter across the 100 Monte Carlo runs (denoted by *SD*), standard deviations of the Gibbs samples of each parameter averaged across Monte Carlo runs (denoted by Est *SD*) and the 90% coverage rates (percentage of Monte Carlo runs for which the 90% credible intervals for parameter l contained the corresponding true value, θ_l).

All the parameters were recovered accurately across all three conditions (see biases and RMSEs in Tables 3–5). The average standard deviations (Est *SD*) of the Gibbs samples were also close to the empirical standard deviations of the parameters. The coverage rates computed using the 5th and 95th percentiles of the Gibbs samples were on average close to but slightly below the 90% nominal rate. Average coverage rates were 85.82, 88.26, and 86.86, respectively, for the three conditions.

Condition 1, which was characterized by positively skewed autoregressive parameter distributions, showed comparable biases and RMSEs to other conditions. Slightly greater discrepancies arose in the tail percentiles, thus yielding slightly lower coverage rates for this condition than for the other two conditions. As will be discussed in Section 7.3, this is directly attributable to the biases in estimating the tail areas of the person-specific parameter distributions in condition 1. Such biases in the person-specific parameters also affected the coverage rates of other person-invariant parameters.

The relatively low average coverage rates in the normal condition were contributed largely by the low coverage rates of the parameters in μ. The coverage rates for the parameters in μ were notably lower in condition 3 than in conditions 1 and 2. This may be related to the fact

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

that a small number of the b_i parameters actually lay near the boundary of or within the non-stationary region. By 'non-stationary region', we mean ranges of parameters that would propel a system to show trends or, specifically, continual deviations from its baseline levels, as defined by μ (as in Figure 5c, for example). As a result, the system also shows increasing variance over time. Larger biases are typically observed in such cases irrespective of other features of the simulations. Even though we used normal distributions of relatively restricted ranges for this condition to confine most cases to the stationary region, some of the parameters in b_i still crossed or lay near the boundary of the non-stationary region. These cases led to more extreme μ estimates that affected the coverage rates for μ in condition 3 directly.

7.3. Estimation of person-specific parameters, b_i

Our main focus of interest was to compare the distribution of true b_i (for $i = 1, \ldots, n$) to the distribution of $\hat{b}_i$ obtained from Gibbs sampling. The means and standard deviations of $\hat{b}_i$ (derived using equation (13)) computed across persons are summarized in Table 6. Further details are summarized in Figure 6. Included in the figure are plots of $\hat{b}_{11,i}$ and $\hat{b}_{12,i}$ during one particular Monte Carlo run in comparison to the densities of the true person-specific parameters generated using equations (15)–(17) (see panels (a) and (b), respectively), and quantile–quantile plots of the true $b_{11,i}$ and $b_{12,i}$ values against the $\hat{b}_{11,i}$ and $\hat{b}_{12,i}$ estimates pooled across all Monte Carlo runs.

Generally, the means and standard deviations of the true b_i distributions were recovered very accurately across all conditions (see Table 6). Several additional observations can be noted based on Figure 6. First, the conditional distributions derived from using the DP prior were flexible enough to recover the general shapes of the different distributions of b_i used in all three conditions. Second, the shapes of the distributions were more accurately recovered in condition 3 than in the other two conditions because the associated true densities (i.e., multivariate normal) were of the same form as the base distribution.

Third, in condition 1, discrepancies in the estimation of $b_{11,i}$ arose primarily in the lower tail region. That is, the lower tail extended too far into the negative region, but the densities in the immediately adjacent regions did not rise rapidly enough to capture some of the more subtle changes in the lower quantiles of the positively skewed true distribution. Fourth, in condition 2, the approximation density generally resembled the bimodal density of the true distribution of $b_{11,i}$, but slight discrepancies were observed near the tail areas of the two modes. In particular, the upper tails of the two modes were assigned too much weight whereas the lower tails were too sparsely represented.

Finally, greater biases were observed in the estimates of $\hat{b}_{12,i}$ and $\hat{b}_{21,i}$ than those of $\hat{b}_{11,i}$ and $\hat{b}_{22,i}$ across all conditions (see Table 6 and Figures 6d, 6h, and 6l). This is not surprising, since the nonlinear model posited that the impacts due to $b_{12,i}$ and $b_{12,i}$ would only be fully manifested at extremely high values of the latent variables. Such instances were relatively rare in the simulated data, with only 50 time points and $n = 170$. Thus, even though the distributions constituted by $\hat{b}_{12,i}$ and$\hat{b}_{21,i}$ generally provided reasonable approximations to the shapes of the distributions of $b_{12,i}$ and $b_{21,i}$, the means of the distributions were slightly offset (i.e., biased).

In sum, the proposed estimation procedures were able to recover all the components in the nonlinear DFA model accurately under diverse distributional assumptions for the parameters in b_i. The sample size considered in the present simulation (with $n = 170$ and $T = 50$) yielded reasonable estimates, although larger sample sizes might be needed to improve the accuracy of the $\hat{b}_{12,i}$ and $\hat{b}_{21,i}$ estimates. In addition, although one particular nonlinear DFA model was considered in all the analyses, the proposed procedures are general enough to be used

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

with any dynamic model with differentiable linear and nonlinear functions and normally distributed process noises (i.e., in the form of equation (8)). To illustrate the performance of the proposed estimation procedures within a linear modelling framework, we conducted a supplementary simulation study using a variation of the linear DFA model in equation (6). In particular, we allowed the parameters b_{11}, b_{12}, b_{21}, and b_{22} in equation (6) to vary over persons and used the DP to approximate their corresponding distributions (generated in the same way as in conditions 1, 2, and 3 in the present simulation). Results based on the linear model are comparable to those obtained from the nonlinear model. Further details are not reported here due to space constraints, but they are available as supplementary materials on the first author's website at http://www.unc.edu/~symiin/Sy-Miin%27s%20website/pub.htm

8. Discussion

In the present paper, we used the DP as a non-parametric prior distribution for selected parameters in a nonlinear DFA model. Using the DP as a prior is equivalent to specifying the prior distribution as a mixture distribution composed of an unknown number of discrete point masses. This approach thus provides the flexibility of a non-parametric mixture approach without the need to define the precise number of (or the range of possible numbers of) mixture components required to approximate an unknown distribution. In addition, we also incorporated several MH procedures within a Gibbs-sampling framework to handle some of the non-standard conditional distributions implicated in the proposed nonlinear DFA model.

A series of empirical and simulation examples was used to illustrate the flexibility of the proposed approach in approximating distributions of various shapes (e.g., normal, bimodal, and skewed). Our empirical example revealed that the baseline autoregressive parameters in our proposed DFA model did in fact show substantial deviations from normality. Using a multivariate normal prior did not reveal the full ranges of the associated parameters and their complex interrelationships. Researchers may thus risk bypassing the true nature of the emotional processes being modelled if parametric assumptions are imposed without any evaluation of their tenability.

Our empirical example provided several new insights into Zautra and colleagues' dynamic affect model. Using one particular non-linear model, we validated that the linkage between PE and NE did indeed intensify on the more emotional days. A unidirectional relation was found in the direction from PE to NE for most individuals when their PE was at high extreme values. The lack of coupling in the direction from NE to PE might be attributable to the participants' generally low NE levels. Thus, very few participants actually manifested a similar change in the direction from NE to PE. By allowing the dynamic parameters in the model to assume non-parametric forms, we were able to evaluate such individual differences more thoroughly.

Some comments can be noted concerning the selection of hyperparameters for the base weight, α. In all our examples, we used hyperparameter choices (i.e., α_1 and α_2 in equation (12)) that yielded relatively high values of α to capture clusters or 'sticks' that were relatively far away from the means or modes of the distributions. Higher values of α are typically needed to approximate distributions that are of high dimensions. Generally, given the moderate sample size in the present study, recovering higher-order moments of the distributions of interest can be difficult. We were able to recover the first and second moments reasonably accurately, however.

The present paper is one of the first applications of semiparametric nonlinear dynamic LVMs to studying change in psychology using the DP prior. Many other extensions are, of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

course, possible. For instance, we did not pursue the issue of model comparison in the present paper. When different prior densities are assumed across two models, some of the common model fit indices within the Bayesian framework cannot be utilized directly without some modifications. In particular, computing the Bayes factor via path sampling (Gelman & Meng, 1998) is not a straightforward matter in this case because of the complexity involved in linking the discrepant prior densities from the different models. Other test statistics, such as those developed by Zhu and Zhang (2004) for assessing finite mixture regression models, can potentially be extended to a dynamic LVM framework to provide a more formal assessment of the need to use infinite-order mixture distributions. Another relatively recent model assessment index, termed the *L* measure (Chen, Dey, & Ibrahim, 2004; Ibrahim, Chen, & Sinha, 2001), has also been advocated as an alternative goodness-of-fit index that works well in situations where proper prior distributions cannot be explicitly derived.

With regard to state (or latent variable score) estimation, we only used first-order linearization to derive the proposal distribution in the MH step. The potential utility of using higher-order linearization schemes or other proposal functions (e.g., Geweke & Tanizaki, 2001) in the state density sampling step could be evaluated in future studies. Hybrid algorithms that combine more computationally efficient particle filtering techniques with MCMC algorithms (Doucet, de Freitas, & Gordon, 2001) could also be developed to aid computation speed. In addition, specification of the person-specific parameters could be reformulated within a mixed effects framework to include both fixed and random effects components. Further investigation of the tenability of assuming a missing at random mechanism for the present data set is also warranted.

Using the DP as a non-parametric prior is not without its limitations. The biggest limitation resides perhaps in the discrete nature of the DP, which dictates that different individuals who are assigned to the same ‘cluster’ (where the number of clusters is less than *n*) would have exactly the same parameter values. One way to circumvent this issue is to use the mixture DP (Caron, Davy, Doucet, Duflos, & Vanheeghe, 2008; Escobar & West, 1995) as an alternative choice. In this case, the prior distribution essentially consists of a mixture of different DP priors. In doing so, different individuals may be assigned similar but not identical values on the parameters or constructs of interest. A second limitation is that the accuracy of the DP approximation is still constrained by the choice of the base distribution. In cases where the true distribution of interest deviates too much from the base distribution, the accuracy of the approximation would also deteriorate accordingly.

The nonlinear DFA model proposed in the present paper is but one example of the many dynamic models that can be used to described change processes. A wide array of modelling examples along these lines can be found in the literature on dynamic linear models (West & Harrison, 1997), dynamic generalized linear models (Fahrmeir & Tutz, 1994), regime-switching state-space models (Kim & Nelson, 1999), differential equation models (Molenaar & Newell, 2003; Singer, 2007), and other nonlinear and non-Gaussian dynamic models (Chow, Ferrer, & Hsieh, 2009; Durbin & Koopman, 2001). Whereas formulating dynamic models within a Bayesian framework opens up countless new possibilities for evaluating more complex models, many of the issues inherent to the Bayesian framework also have to be handled with caution. Sensitivity of the modelling results to prior choices and misspecification of other parametric/nonparametric assumptions remains an important issue that deserves more attention from researchers. Parallel to the increase in model complexity are, of course, new challenges for deriving appropriate model fit indices and convergence diagnostics.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To conclude, the DP prior can be used as a flexible non-parametric prior for distributions whose functional forms are unknown. In time series modelling, it is not unusual to encounter parameter distributions that show complex restrictions in range. Very often, the associated distributions are not only non-normal, but also asymmetric. Taking a non-parametric or semiparametric approach allows the assumption of normality to be treated as a testable hypothesis, as opposed to a 'gold standard' by which modellers have to abide. We hope to have illustrated the need to relax some of these parametric assumptions in fitting dynamic LVMs.

Acknowledgments

We would like to thank Frank Fujita for allowing us to use his data for the empirical illustration in this paper. The C ++ scripts used for simulations in the present article are available upon request from the first author. Funding for this study was provided by grants from NSF (BCS-0826844), NIH (UL1-RR025747-01, MH086633, P01CA142538-01 and AG033387), and NSFC (10961026).

References

Ansari A, Iyengar R. Semiparametric Thurstonian models for recurrent choices: A Bayesian analysis. Psychometrika. 2006; 71(4):631–657.

Antoniak CE. Mixtures ofDirichlet processes with applications to Bayesian nonparametric problems. Annals of Statistics. 1974; 2(6):1152–1174.

Bartholomew, DJ.; Knott, M. Latent variable models and factor analysis. 2nd ed.. London: Arnold; 1999.

Bauer D, Curran P. Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. Psychological Methods. 2003; 8:338–363. [PubMed: 14596495]

Browne, MW.; Nesselroade, JR. Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In: Maydeu-Olivares, A.; McArdle, JJ., editors. Contemporary psychometrics: A Festschrift for Roderick P. McDonald. Mahwah, NJ: Erlbaum; 2005. p. 415-452.

Caron F, Davy M, Doucet A, Duflos E, Vanheeghe P. Bayesian inference for linear dynamic models with Dirichlet process mixtures. IEEE Transactions on Signal Processing. 2008; 56(1):71–84.

Cattell RB, Cattell AKS, Rhymer RM. P-technique demonstrated in determining psychophysical source traits in a normal individual. Psychometrika. 1947; 12:267–288. [PubMed: 18921433]

Chen M-H, Dey DK, Ibrahim JG. Bayesian criterion based model assessment for categorical data. Biometrika. 2004; 91:45–63.

Chow, S-M.; Ferrer, E.; Hsieh, F. Statistical methods for modeling human dynamics: An interdisciplinary dialogue. New York: Routledge; 2009.

Chow S-M, Hamaker EJ, Allaire JC. Detecting discrete shifts in dynamics in group-based state-space models. Multivariate Behavioral Research. 2009; 44:465–496.

Chow S-M, Nesselroade JR, Shifren K, McArdle JJ. Dynamic structure of emotions among individuals with Parkinson's disease. Structural Equation Modeling. 2004; 11(4):560–582.

Chow S-M, Ram N, Boker SM, Fujita F, Clore G. Emotion as thermostat: Representing emotion regulation using a damped oscillator model. Emotion. 2005; 5(2):208–225. [PubMed: 15982086]

De Jong P, Mazzi S. Modeling and smoothing unequally spaced sequence data. Statistical Inference for Stochastic Processes. 2001; 4:53–71.

Diener E, Fujita F, Smith H. The personality structure of affect. Journal of Personality and Social Psychology. 1995; 69(1):130–141.

Doucet, A.; de Freitas, N.; Gordon, N. Sequential Monte Carlo methods in practice. New York: Springer; 2001.

Duncan KA, MacEachern SN. Nonparametric Bayesian modelling for item response. Statistical Modeling. 2008; 8(1):41–66.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Dunson, DB. Nonparametric Bayes statistical modeling. New York: Cambridge University Press; 2008. Nonparametric Bayes applications to biostatistics.

Durbin, J.; Koopman, SJ. Time series analysis by state-space methods. New York: Oxford University Press; 2001.

Escobar MD, West M. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association. 1995; 90(430):577–588.

Fahrmeir, L.; Tutz, G. Multivariate statistical modelling based on generalized linear models. Berlin: Springer-Verlag; 1994.

Ferguson TS. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1973; 1(2): 209–230.

Ferrer E, Nesselroade JR. Modeling affective processes in dyadic relations via dynamic factor analysis. Emotion. 2003; 3(4):344–360. [PubMed: 14674828]

Frederickson BL, Losada MF. Positive affect and the complex dynamics of human flourishing. American Psychologist. 2005; 60(7):678–686. [PubMed: 16221001]

Gelman, A. Inference and monitoring convergence. In: Gilks, WR.; Richardson, S.; Spiegelhalter, DJ., editors. Markov chain Monte Carlo in practice. London: Chapman & Hall; 1996. p. 131-144.

Gelman A, Meng XL. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. Statistical Science. 1998; 13:163–185.

Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica. 1996; 6:733–807.

Geweke J, Tanizaki H. Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling. Computational Statistics and Data Analysis. 2001; 37:151–170.

Gottman, JM.; Murray, JD.; Swanson, CC.; Tyson, R.; Swanson, KR., editors. The mathematics of marriage: Dynamic nonlinear models. Cambridge, MA: MIT Press; 2002.

Hamaker EL, Dolan CV, Molenaar PCM. ARMA-based SEM when the number of time points *T* exceeds the number of cases *N*: Raw data maximum likelihood. Structural Equation Modeling. 2003; 10(3):352–379.

Hamilton, JD. Time series analysis. Princeton, NJ: Princeton University Press; 1994.

Harvey, AC. Forecasting, structural time series models and the Kalman filter. Cambridge: Cambridge University Press; 2001.

Hershberger SL, Corneal SE, Molenaar PCM. A dynamic factor analysis of the emotional response patterns underlying stepdaughter/stepfather relationships. Journal of Structural Modeling. 1994; 2:31–52.

Ibrahim JG, Chen M-H, Sinha D. Criterion-based models for Bayesian model assessment. Statistica Sinica. 2001; 11:419–443.

Ishwaran H, James A. Gibbs sampling methods for stick breaking priors. Journal of the American Statistical Association. 2001; 96(453):161–173.

Ishwaran H, Zarepour M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. Biometrika. 2000; 87:371–390.

Jones CJ, Nesselroade JR. Multivariate, replicated, single-subject designs and P-technique factor analysis: A selective review of the literature. Experimental Aging Research. 1990; 16:171–183. [PubMed: 2131263]

Jöreskog, KG. Analyzing psychological data by structural analysis of covariance matrices. In: Krantz, DH.; Atkinson, RC.; Duncan, RL.; Suppes, P., editors. Contemporary developments in mathematical psychology (Vol. 2): Measurement, psychophysics and neural information processing. San Francisco: W. H. Freeman; 1974. p. 1-56.

Jöreskog KG, Moustaki I. Factor analysis of ordinal variables: A comparison of three approaches. Multivariate Behavioral Research. 2001; 36(3):347–387.

Karabotsos G, Walker SG. Coherent psychometric modelling with Bayesian nonparametrics. British Journal of Mathematical and Statistical Psychology. 2009; 62:1–20. [PubMed: 17903345]

Kenny DA, Judd CM. Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin. 1984; 96:201–210.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Kim, C-J.; Nelson, CR. State-space models with regime switching: Classical and Gibbs-sampling approaches with applications. Cambridge MA: MIT Press; 1999.

Klein A, Moosbrugger H. Maximum likelihood estimation of latent interaction effects with the LMS method. Psychometrika. 2000; 65:457–474.

Lee, S-Y. Structural equation modeling: A Bayesian approach. Chichester: Wiley; 2007.

Lee S-Y, Lu B, Song X-Y. Semiparametric Bayesian analysis of structural equation models with fixed covariates. Statistics in Medicine. 2007; 27(13):2341–2360. [PubMed: 18008378]

Lee S-Y, Zhu H-T. Statistical analysis of nonlinear structural equation models with continuous and polytomous data. British Journal of Mathematical and Statistical Psychology. 2000; 53:209–232. [PubMed: 11109705]

Lindley DV, Smith AFM. Bayes estimates for the linear model (with discussion). Journal of the Royal Statistical Society, Series B. 1972; 34:1–42.

Lindsay, BG. Mixturemodels: Theory, geometry and applications. Hayward, CA: Institute of Mathematical Statistics; 1995.

Little, RJA.; Rubin, DB. Statistical analysis with missing data. New York: Wiley; 1987.

Ljung L, Caines PE. Asymptotic normality of prediction error estimators for approximate system models. Stochastics. 1979; 3:26–49.

McArdle, JJ. Structural equation modeling of an individual system: Preliminary results from 'A case study in episodic alcoholism'. Department of Psychology, University of Denver; 1982. Unpublished manuscript

McLachlan, G.; Peel, D. Finite mixture models. New York: Wiley; 2000.

Meng X-L. Posterior predictive *p*-values. Annals of Statistics. 1994; 22:1142–1160.

Molenaar PCM. A dynamic factor model for the analysis of multivariate time series. Psychometrika. 1985; 50(2):181–202.

Molenaar PCM, Newell KM. Direct fit of a theoretical model of phase transition in oscillatory finger motions. British Journal of Mathematical and Statistical Psychology. 2003; 56:199–214. [PubMed: 14633332]

Navarro DJ, Griffiths TL, Steyvers M, Lee MD. Modeling individual differences using Dirichlet processes. Journal of Mathematical Psychology. 2006; 50:101–122.

Nesselroade, JR.; McArdle, JJ.; Aggen, SH.; Meyers, JM. Alternative dynamic factor models for multivariate time-series analyses. In: Moskowitz, DM.; Hershberger, SL., editors. Modeling intraindividual variability with repeated measures data: Advances and techniques. Mahwah, NJ: Erlbaum; 2002. p. 235-265.

Ram N, Chow S-M, Bowles RP, Wang L, Grimm KJ, Fujita F, et al. Recovering cyclicity in pleasant and unpleasant affect using spectral analysis, the Rating Scale model, and planned incompleteness. Psychometrika: Application Reviews and Case Studies. 2005; 70:773–790.

Schumacker RE. Latent variable interaction modeling. Structural Equation Modeling. 2002; 9(1):40–54.

Sethuraman J. A constructive definition of Dirichlet priors. Statistica Sinica. 1994; 4:639–650.

Shi J-Q, Lee S-Y. Using factor analysis to estimate parameters. British Journal of Mathematical and Statistical Psychology. 1998; 51:233–252.

Singer, H. Stochastic differential equation models with sampled data. In: van Montfort, K.; Oud, JHL.; Satorra, A., editors. Longitudinal models in the behavioral and related sciences. Mahwah, NJ: Erlbaum; 2007. p. 73-106.

Sorensen HW, Alspach DL. Recursive Bayesian estimation using Gaussian sums. Automatica. 1971; 7:465–479.

Titterington, DM.; Smith, AFM.; Markov, UE. The statistical analysis of finite mixture distributions. New York: Wiley; 1985.

Wei, WWS. Time series analysis. Redwood City, CA: Addison-Wesley; 1990.

West, M.; Harrison, J. Bayesian forecasting and dynamic models. 2nd ed.. New York: Springer-Verlag; 1997.

Young PC, Pedregal DJ, Tych W. Dynamic harmonic regression. Journal of Forecasting. 1999; 18(6): 369–394.

NIH-PA Author Manuscript

Zautra AJ, Reich JW, Davis MC, Nicolson NA, Potter PT. The role of stressful events in the relationship between positive and negative affects: Evidence from field and experimental studies. Journal of Personality. 2000; 68:927–951. [PubMed: 11001154]

Zhang, G.; Browne, M. Dynamic factor analysis with ordinal variables. In: Chow, S-M.; Ferrer, E.; Hsieh, F., editors. Statistical methods for modeling human dynamics: An interdisciplinary dialogue. New York: Routledge; 2009.

Zhang Z, Nesselroade JR. Bayesian estimation of categorical dynamic factor models. Multivariate Behavioral Research. 2007; 42:729–756.

Zhu HT, Zhang HP. Hypothesis testing in mixture regression models. Journal of Royal Statistical Society, Series B. 2004; 66:3–16.

Appendix

Conditional distributions used in the Gibbs sampling procedures

To estimate the proposed nonlinear LVM, the Gibbs sampler is implemented in which a sequence of sampling steps [steps (a)–(i)] is carried out iteratively. The conditional distributions from which Gibbs samples are obtained are summarized below.

Steps (a)–(e): Conditional distributions related to the non-parametric components

The main idea behind efficient sampling of the non-parametric components is to recast the definition of b_i in terms of the latent variable L_i, $i = 1, \ldots, n$, which records the cluster membership of b_i such that $b_i = Z_{L_i}$. The base distribution in the present context was defined to be an r-variate normal distribution with mean vector μ_Z and covariance matrix Ψ_Z. Conjugate prior distributions were specified for μ_Z, Ψ_Z, and α as in equation (12). To explore the posterior in relation to the non-parametric components, we sample (π, Z, L, μ_Z, ψ_Z, α) by means of the blocked Gibbs sampler to encourage mixing of the Markov chain. That is, Gibbs sampling of the non-parametric components was regrouped into five subsidiary steps (or *blocks*), involving sampling from the conditional distributions $p(\pi, Z|L, \mu_Z, \psi_Z, \alpha, \tau, Y^*, \theta, H, Y_{\text{obs}})$, $p(L|\pi, Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$, $p(\mu_Z|Z, \psi_Z)$, $p(\psi_Z|Z, \mu_Z)$, and $p(\alpha|\pi)$. These five conditional distributions are summarized below.

Block 1. Posterior samples of $[\mu_Z|Z, \Psi_Z]$ can be obtained by sampling from

$$p(\mu_Z|Z, \Psi_Z) \sim N_r(\mu_\mu, \Sigma_\mu), \tag{A1}$$

where $\Sigma_\mu = (G\Psi_Z^{-1} + \Psi_{\mu_Z}^{-1})^{-1}$ and $\mu_\mu = \Sigma_\mu[\Psi_{\mu_Z}^{-1}\mu_{Z_0} + \Psi_Z^{-1}\sum_{g=1}^{G} Z_g]$.

Block 2. For $j = 1, \ldots, r$, each of the diagonal elements of Ψ_Z given Z and μ_Z is distributed as

$$p\left(\psi_{Z_j}^{-1}\middle|Z, \mu_Z\right) \overset{\text{i.i.d.}}{\sim} \text{Gamma}\left(c_1 + \frac{G}{2}, c_2 + \frac{1}{2}\sum_{g=1}^{G}(Z_{g_j} - \mu_{Z_j})^2\right), \tag{A2}$$

where Z_{gj} is the jth element of the values in Z associated with point mass (or cluster) g and μ_{Z_j} is the jth element of μ_Z.

Block 3. Following the derivations detailed elsewhere (Ishwaran & James, 2001; Ishwaran & Zarepour, 2000; Lee *et al.*, 2007), the conditional distribution ($\alpha|\pi$) can be shown to be

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p(\alpha|\pi) \sim \text{Gamma}\left(a_1+G-1, a_2-\sum_{g=1}^{G-1}\log\left(1-v_g^*\right)\right), \tag{A3}$$

where v_g^* is a random weight sampled from the beta distribution and it is sampled within Block 4.

Block 4. As π and α are independent given (Z, τ, Y^*, θ, H, Y_{obs}), the distribution (π, $Z|L$, μ_Z, ψ_Z, α, τ, Y^*, θ, H, Y_{obs}) is proportional to $p(\pi|L, \alpha)p(Z|L, \mu_Z, \psi_Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$. Thus, the conditional distribution can be decomposed into two independent components to be derived separately.

Conditional distribution *p*(π|*L*, α)—It can be shown that the conditional distribution (π|L, α) conforms to a generalized Dirichlet distribution,

$$p(\pi|L,\alpha) \sim g(a_1^*, b_1^*, \ldots, a_{G-1}^*, b_{G-1}^*), \tag{A4}$$

where $a_g^*=1+d_g, b_g^*=\alpha+\sum_{j=g+1}^{G} d_j$, for g = 1, …, G − 1, and d_g is the number of L_is (and thus individuals) whose value equals g. Sampling from the conditional distribution (π|L, α) can be accomplished as follows. First, v_g^* is first drawn from a Beta(a_g^*, b_g^*) distribution. Subsequently, π_g is obtained for g = 1, …, G as

$$\begin{aligned}
&\pi_1=v_1^*,\\
&\pi_G=1-\sum_{g=1}^{G-1}\pi_g,\\
&\pi_g=\prod_{j=1}^{g-1}\left(1-v_j^*\right)v_g^*, \text{ for } g \neq 1 \text{ or } G.
\end{aligned} \tag{A5}$$

Conditional distribution *p*(*Z*|*L*, μ*Z*, ψ*Z*, τ, *Y, θ, *H*, *Y*obs)—**Let $L_1^*, \ldots, L_d^*$ be the d unique L_i values (i.e., unique number of 'clusters'), $Z^L=(Z_{L_1^*}, \ldots, Z_{L_d^*})$, and let $Z^{[L]}$ be components in $Z = (Z_1, \ldots, Z_G)$ other than Z^L. Then

$$p(Z|L,\mu_Z,\psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}})=p(Z^{[L]}|\mu_Z,\psi_Z)p(Z^L|L,\mu_Z,\psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}}),$$

where $p(Z^{[L]}|\ \mu_Z, \psi_Z)$ is simply the r-variate normal distribution, $N_r(\mu_Z, \psi_Z)$ and

$$p(Z^L|L,\mu_Z,\psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}})=\prod_{g=1}^{d} p\left(Z_{L_g^*}\middle|L,\mu_Z,\psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}}\right).$$

It can be shown that the conditional distribution $p(Z_{L_g^*}|L,\mu_Z,\Psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}})$ is non-standard and cannot be derived directly via Gibbs sampling. Specifically,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$p\left(Z_{L_g^*}\middle|L,\mu_Z,\Psi_Z,\tau,Y^*,\theta,H,Y_{\text{obs}}\right)\propto p\left(Z_{L_g^*}\middle|\mu_Z,\Psi_Z\right)\prod_{\{i:L_i=L_g^*\}}\left[p\left(\eta_i|b_i=Z_{L_g^*},\theta_\eta\right)p(y_i^*|\eta_i,\theta_\varepsilon)\right]$$

in which $p(\eta_i|b_i=Z_{L_g^*},\theta_\eta)$ is given by

$$\begin{cases} p(\eta_{i0})\prod_{t=1}^{T}p\left(\eta_{it}|\eta_{it-1},b_i=Z_{L_g^*},\theta_\eta\right), & \text{if } \eta_{i0} \text{ is stochastic,}\\ \prod_{t=1}^{T}p\left(\eta_{it}|\eta_{it-1},b_i=Z_{L_g^*},\theta_\eta\right), & \text{otherwise.}\end{cases} \tag{A6}$$

From equation (A6), it can be noted that multiplication involving the density $p(\eta_{it}|\eta_{i,t-1}, b_i, \theta_\eta)$ results in a conditional density that is non-normal and non-standard due to the nonlinearity of $f_t(.)$ and the fact that $Z_{L_g^*}$ is random, as opposed to fixed within this sampling step. Instead, we adopt an MH step as follows. At the qth iteration with a current value $Z_{L_g^*}^{(q)}$, a new candidate $Z_{L_g^*}$ is generated from the normal distribution $N(Z_{L_g^*}^{(q)},\sigma_b^2\Omega_b)$, where $\Omega_b=\Psi_Z^{-1}+\sum_{\{i:L_i=L_g^*\}}\sum_{t=1}^{T}\Delta_{bit}'\Psi_\zeta^{-1}\Delta_{bit})^{-1}$ and $\Delta_{bit}=\partial\eta_{it}/\partial b_i'|_{b_i=Z_{L_g^*}^{(q)},\eta_{it}=\eta_{i,t-1}}$. The latter is derived by means of the implicit function theorem, namely, $\partial\eta_{it}/\partial b_i'=(\partial f_{i,t+1}/\partial\eta_{it}')^{-1}\partial f_{i,t+1}/\partial b_i'$. The new $Z_{L_g^*}$ is accepted with probability

$$\min\left\{1,\frac{p\left(Z_{L_g^*}\middle|\mu_Z,\psi_Z\right)\prod_{\{i:L_i=L_g^*\}}\prod_{t=1}^{T}p\left(\eta_{it}|\eta_{it-1},b_i=Z_{L_g^*},\theta_\eta\right)}{p\left(Z_{L_g^*}^{(q)}\middle|\mu_Z,\psi_Z\right)\prod_{\{i:L_i=L_g^*\}}\prod_{t=1}^{T}p\left(\eta_{it}|\eta_{it-1},b_i=Z_{L_g^*}^{(q)},\theta_\eta\right)}\right\}. \tag{A7}$$

The variance σ_b^2 can be chosen such that the average acceptance rate is approximately 0.25 or more.

Block 5. The conditional distribution $(L_i|\pi, Z, \tau, Y^*, \theta, H, Y_{\text{obs}})$ is given by

$$(L_i|\pi,Z,\tau,Y^*,\theta,H,Y_{\text{obs}})\overset{\text{i.i.d.}}{\sim}\text{Multinomial}\left(\pi_{ig}^*,g=1,\ldots,G\right), \tag{A8}$$

where π_{ig}^* is proportional to $(\pi_g p(y_i^*|\eta_i,\theta_\varepsilon)p(\eta_i|Z_g,\theta_\eta))$ and $\pi_g(g = 1, \ldots, G)$ are available from step (i.e., block) 4 summarized in equation (A5). Note that because $p(y_i^*|\eta_i,\theta_\varepsilon)$ is fixed conditional on η_i, this component can thus be omitted from the computation.

Step (f): Conditional distribution for latent variable estimates, $p(H|\tau, Y^*, \theta, Y_{\text{obs}})$

The conditional distribution from which posterior samples of the latent variable estimates are obtained can be derived as

$$\begin{aligned} p(H|\tau,Y^*,\theta,Y_{\text{obs}}) &= \prod_{i=1}^{n}p\left(\eta_i|Y_i^*,\theta,b_i\right)\\ &= \prod_{i=1}^{n}\prod_{t=1}^{T}p\left(\eta_{it}|H_{i,t-1},H_{i,t+1}^*,Y_i^*,\theta,b_i\right),\end{aligned}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

where $H_{it} = (\eta_{i1}, \ldots, \eta_{it})$ and $H^*_{i,t+1} = (\eta_{i,t+1}, \ldots, \eta_{iT})$. According to the Gibbs sampler, random draws of η_i from $p(\eta_i|Y^*_i|, \theta, b_i)$ are based on those of η_{it} from $p(\eta_{it}|H_{i,t-1}, H^*_{i,t+1}, Y^*_i, \theta, b_i)$ for each time point. That is, for $t = 1, \ldots, T$:

$$p(\eta_{it}|H_{i,t-1}, H^*_{i,t+1}, Y^*_i, \theta, b_i) \sim \begin{cases} p(y^*_{it}|\eta_{it}, \theta_\varepsilon)p(\eta_{it}|\eta_{i,t-1}, b_i, \theta_\eta)p(\eta_{it+1}|\eta_{it}, b_i, \theta_\eta), & \text{for } t=1, \ldots, T-1, \\ p(y^*_{it}|\eta_{it}, \theta_\varepsilon)p(\eta_{it}|\eta_{i,t-1}, b_i, \theta_\eta), & \text{for } t=T. \end{cases}$$

Note that we could obtain a standard conditional distribution for $t = T$ but not for $t < T$. Specifically, at $t = T$, the conditional distribution $p(\eta_{iT}|H_{i,T-1}, Y^*_i, \theta, b_i)$ is given by $\eta_{iT} \sim N_w(b^*_{iT}, B^*_{iT})$, where $B^*_{iT} = (\Psi^{-1}_\zeta + \Lambda' \Psi^{-1}_\varepsilon \Lambda)^{-1}$ and $b^*_{iT} = B^*_{iT}[\Psi^{-1}_\zeta f_T(\eta_{i,T-1}, b_i) + \Lambda' \Psi^{-1}_\varepsilon (y^*_{iT} - \mu)]$. However, when $t < T$, multiplication involving the density $p(\eta_{i,t+1}|\eta it, b_i, \theta_\eta)$ would result in a conditional density that is non-normal and non-standard. This is due directly to the nonlinearity of $f_t(.)$ and the fact that η_{it} is random, as opposed to fixed, at each t. We adopted the following MH algorithm to sample observations from the posterior density $p(\eta_{it}|H_{i,t-1}, H^*_{i,t+1}, Y^*_i, \theta, b_i)$. At the qth iteration with a current value $\eta^{(q)}_{it}$, a new candidate η_{it} is generated from the normal distribution $N(\eta^{(q)}_{it}, \sigma^2_\eta \Omega_\eta)$, where $\Omega_\eta = (B_{iT^*-1} + \Delta'_{it} \Psi^{-1}_\zeta \Delta_{it})^{-1}$ and $\Delta_{it} = \partial f_{t+1}/\partial \eta'_{it}|_{\eta_{it}=\eta_{i,t-1}}$, and it is accepted with probability

$$\min\left\{1, \frac{p(\eta_{it}|H_{i,t-1}, H^*_{i,t+1}, Y^*_i, \theta, b_i)}{p(\eta^{(q)}_{it}|H_{i,t-1}, H^*_{i,t+1}, Y^*_i, \theta, b_i)}\right\}.$$

The variance σ^2_η can be chosen such that the average acceptance rate is approximately 0.25 or more.

Step (g): Conditional distributions for parameters in θ

Assuming that the parameters in b are independent of those contained in θ, and that parameters in θ_η are conditionally independent of those in θ_ε, the conditional distribution $p(\theta|\tau, Y^*, H, Y_{obs}, b) = p(\theta_\eta|H, b)p(\theta_\varepsilon|\tau, Y^*, H, Y_{obs}, b)$ is derived by computing the latter two densities separately for all the person-invariant parameters in the dynamic and measurement models.

Parameters in the dynamic model—At the dynamic level, the only parametric posterior distribution associated with $p(\theta_\eta|H, b)$ is that of $p(\Psi^*_\zeta|H, b)$. We used a w-dimensional inverse Wishart distribution as the conjugate prior for the process noise covariance matrix, Ψ_ζ, i.e., $p(\Psi_\zeta) \sim \mathrm{IW}_w[w_0, \Psi^{-1}_{\zeta_0}]$ thus yielding

$$p(\Psi_\zeta|H, b) \sim \mathrm{IW}_w\left[nT + w_0, R_\eta + \Psi^{-1}_{\zeta_0}\right],$$

where $R_\eta = \sum_{i=1}^n \sum_{i=1}^T [\eta_{it} - f(\eta_{i,t-1}, b_i)][\eta_{it} - f(\eta_{i,t-1}, b_i)]'$.

Parameters in the measurement model—Following the work of many others (e.g., Lindley & Smith, 1972; Shi & Lee, 1998; Lee & Zhu, 2000), we specified conjugate priors for the distributions of $p(\mu)$, $p(\psi^{-1}_{\varepsilon_k})$ and $p(\Lambda_k|\Psi_{\varepsilon k})$ as in equation (12) for $k = 1, \ldots, p$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To cope with the case of fixed known elements in Λ, let $C = (c_{kj})$ be the index matrix such that $c_{kj} = 0$ if λ_{kj} is known and $c_{kj} = 1$ if λ_{kj} is unknown, and $r_{\varepsilon_k}=\sum_{j=1}^{w} c_{kj}$, where r_{ε_k} denotes the number of freed factor loadings in the kth row of Λ. Further, let H_k ($r_{\varepsilon_k}T \times n$) be a submatrix of H such that its jth row, for which $c_{kj} = 0$, has been deleted, and an $n \times 1$ vector U^*_{kt} such that $U^{*'}_{kt}=(U^*_{k1t}, \ldots, U^*_{knt})$ has elements

$$U^*_{kit}=y^*_{kit} - \mu_k - \sum_{j=1}^{w}\lambda_{kj}\eta_{jit}(1 - c_{kj}), \tag{A9}$$

where y^*_{kit} and μ_k denote the kth element of y^*_{it} and μ, respectively. Then it can be shown that

$$p\left(\psi^{-1}_{\varepsilon_k}|\mu, Y^*H\right) \sim \text{Gamma}\left(\tfrac{nT}{2}+\alpha_{0\varepsilon k}, \beta_{\varepsilon_k}\right), p\left(\Lambda_{\varepsilon_k}|\psi^{-1}_{\varepsilon_k}, \mu, Y^*H\right) \sim N[\upsilon_k, \psi_{\varepsilon_k}\Omega_k],$$
$$p(\mu|\Lambda, \psi_\varepsilon, Y^*, H) \sim N\left[\Omega_\mu\left\{\sum_0^{-1}\mu_0+\Psi^{-1}_\varepsilon\sum_{i=1}^{n}\sum_{t=1}^{T}(y^*_{it} - \Lambda\eta_{it})\right\}, \Omega_\mu\right],$$

where

$$\beta_{\varepsilon k}=\beta_{0\varepsilon k}+\tfrac{1}{2}\left(\sum_{t=1}^{T}U^{*'}_{kt}U^*_{kt} - \upsilon'_k\Omega^{-1}_k\upsilon_k+\Lambda'_{0k}H^{-1}_{0\Lambda_k}\Lambda_{0_k}\right),$$
$$\Omega_k=(H^{-1}_{0\Lambda_k}+\sum_{t=1}^{T}H_{kt}H'_{kt})^{-1}, \upsilon_k=\Omega_k(\sum_{t=1}^{T}H_{kt}U^*_{kt}+H^{-1}_{0\Lambda_k}\Lambda_{0_k}), \text{ and } \Omega_\mu=(\sum_0^{-1}+nT\Psi^{-1}_\varepsilon)^{-1}.$$

Step (h): Conditional distribution $p(Y^*_{mis}|\theta, H)$

Since the y^*_{it} are mutually independent for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, the $y_{it,\text{mis}}$ are also independent of each other for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. In addition, Ψ_ε is assumed to be a diagonal matrix. Thus, $y_{it,\text{mis}}$ is also independent of $y_{it,\text{obs}}$. Because Y^* are missing at random, we have $p\left(y^*_{it,\text{mis}}|\theta_\varepsilon, \eta_{it}\right) \overset{\text{i.i.d.}}{\sim} N\left(\mu_{it,\text{mis}}+\Lambda_{it,\text{mis}}\eta_{it}, \Psi_{\varepsilon it,\text{mis}}\right)$, where $\mu_{it,\text{mis}}$ is a subvector of μ with components corresponding to the missing components in $y_{it,\text{mis}}$, $\Lambda_{it,\text{mis}}$ is a submatrix of Λ with rows corresponding to the missing components in $y_{it,\text{mis}}$, and $\Psi_{\varepsilon it,\text{mis}}$ is a submatrix of Ψ_ε with rows and columns corresponding to the missing components in $y^*_{it,\text{mis}}$.

Step (i): Conditional distribution $p(\tau, Y^*_{obs}|\theta, H, Y_{obs})$

To sample τ and Y^*_{obs}, we first note that

$$p(\tau_k, Y^*_{k,\text{obs}}|\theta, H, Y_{k,\text{obs}})=p(\tau_k|Y_{k,\text{obs}}, \theta, H)p(Y^*_{k,\text{obs}}|\tau_k, Y_{k,\text{obs}}, \theta, H),$$

where

$$p(\tau_k|Y_{k,\text{obs}}, \theta, H) \propto \prod_{i=1}^{n}\prod_{t=1}^{T}\left(\Phi\left(\frac{\tau_{k,y_{it,k+1,\text{obs}}} - \mu_k - \Lambda'_k\eta_{it}}{\psi^{1/2}_{\varepsilon_k}}\right) - \Phi\left(\frac{\tau_{k,y_{it,k,\text{obs}}} - \mu_k - \Lambda'_k\eta_{it}}{\psi^{1/2}_{\varepsilon_k}}\right)\right), \tag{A10}$$

$$p\left(y^*_{it,k,\text{obs}}|\tau_k, Y_{k,\text{obs}}, \theta, H\right)=N(\mu_k+\Lambda'_k\eta_{it}, \psi_{\varepsilon_k})I_{(\tau_{k,y_{it,k,\text{obs}}-1},\tau_{k,y_{it,k,\text{obs}}})}(y^*_{it,k,\text{obs}}). \tag{A11}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

To generate observations from the non-standard and complex joint conditional density of τ_k and $Y^*_{k,\text{obs}}$, the following MH step is embedded within the Gibbs sampler. Specifically, a vector of thresholds $(\tau_{k,2}, \ldots, \tau_{k,M-2})$ is first generated from the truncated normal distribution

$$\tau_{k,s} \sim N\left(\tau^{(q)}_{k,s}, \sigma^2_{\tau_k}\right) I_{\left(\tau_{k,s-1}, \tau^{(q)}_{k,s+1}\right)}(\tau_{k,s}), \text{ for } s=2, \ldots, M-2, \tag{A12}$$

where $\tau^{(q)}_{k,s}$ denotes the value of $\tau_{k,s}$ at the qth iteration of the Gibbs sampler and $\sigma^2_{\tau_k}$ is a preassigned constant. As mentioned earlier, the values of the first ($s = 1$) and last ($s = M - 1$) thresholds are fixed for identification purposes. Each new draw of $\tau_{k,s}$ is then retained with acceptance probability min(1, R_k), where

$$R_k = \prod_{s=2}^{M-2} \frac{\Phi\left[\left(\tau^{(q)}_{k,s+1} - \tau^{(q)}_{k,s}\right)/\sigma_{\tau_k}\right] - \Phi\left[\left(\tau_{k,s-1} - \tau^{(q)}_{k,s}\right)/\sigma_{\tau_k}\right]}{\Phi\left[(\tau_{k,s+1} - \tau_{k,s})/\sigma_{\tau_k}\right] - \Phi\left[(\tau^{(q)}_{k,s-1} - \tau_{k,s})/\sigma_{\tau_k}\right]}$$
$$\times \prod_{i=1}^{n}\prod_{t=1}^{T} \frac{\Phi\left[\psi_{\varepsilon_k}^{-1/2}\left\{\tau_{k,y_{it,k,\text{obs}}} - \mu_k - \Lambda_k'\eta_{it}\right\}\right] - \Phi\left[\psi_{\varepsilon_k}^{-1/2}\{\tau_{k,y_{it,k,\text{obs}}-1} - \mu_k - \Lambda_k'\eta_{it}\}\right]}{\Phi\left[\psi_{\varepsilon_k}^{-1/2}\left\{\tau^{(q)}_{k,y_{it,k,\text{obs}}} - \mu_k - \Lambda_k'\eta_{it}\right\}\right] - \Phi\left[\psi_{\varepsilon_k}^{-1/2}\left\{\tau^{(q)}_{k,y_{it,k,\text{obs}}-1} - \mu_k - \Lambda_k'\eta_{it}\right\}\right]}. \tag{A13}$$

Once the threshold values have been determined, they are then used to generate new draws of $y^*_{it,k,\text{obs}}$ using equation (A11).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

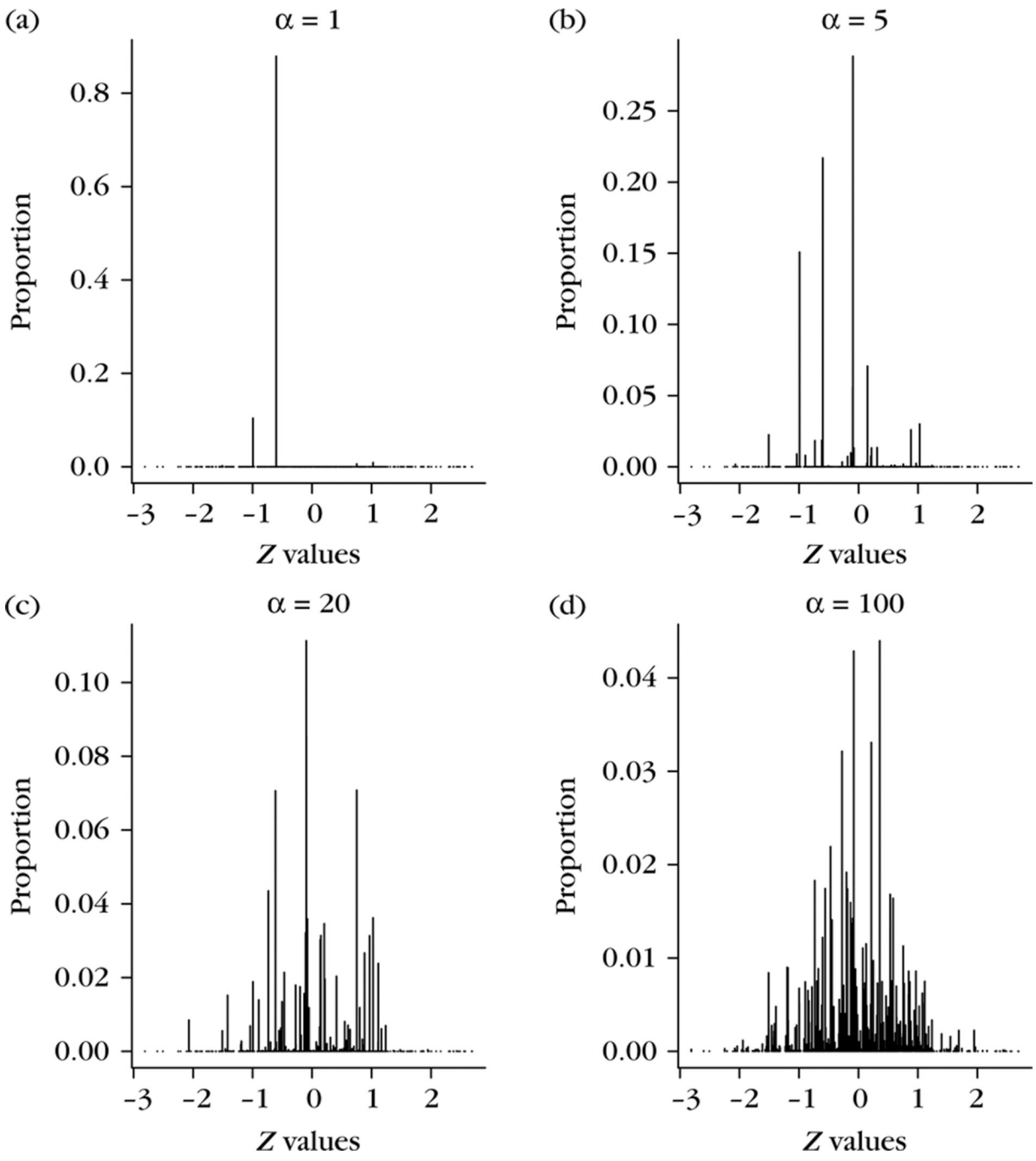

Figure 1.
Plots of realizations from the DP with different values of α and a $N(0,1)$ base distribution.

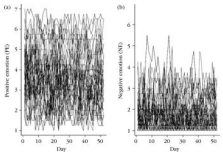

Figure 2.
(a) PE and (b) NE ratings of the participants over 52 days. To avoid cluttering the figure, we have only plotted the trajectories of 50 randomly sampled individuals.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

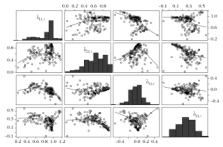

Figure 3.

Matrix scatterplots of the $\hat{b_i}$ estimates using the DP prior, as averaged across three Markov chains. A loess line is imposed on each of the scatterplots. The diagonal plots are histograms of these estimates. For $b_{11,i}$, $M = 0.83$, $SD = 0.20$, 90% CI [0.44, 1.05]; for $b_{22,i}$, $M = 0.64$, $SD = 0.21$, 90% CI [0.26, 0.91]; for $b_{12,i}$, $M = -0.03$, $SD = 0.14$, 90% CI [−0.28, 0.20]; and for $b_{21,i}$, $M = 0.26$, $SD = 0.12$, 90% CI [0.06, 0.46].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

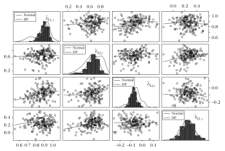

Figure 4.

Matrix scatterplots of the person-specific parameter estimates with a multivariate normal prior, as averaged across three Markov chains. A loess line is imposed on each of the scatterplots. The diagonal plots are histograms of these estimates. For $b_{11,i}$, $M = 0.86$, $SD = 0.10$, 90% CI [0.69, 1.00]; for $b_{22,i}$, $M = 0.63$, $SD = 0.14$, 90% CI [0.36, 0.82]; for $b_{12,i}$, $M = -0.07$, $SD = 0.08$, 90% CI [−0.19, 0.06]; and for $b_{21,i}$, $M = 0.26$, $SD = 0.13$, 90% CI [0.03, 0.47].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

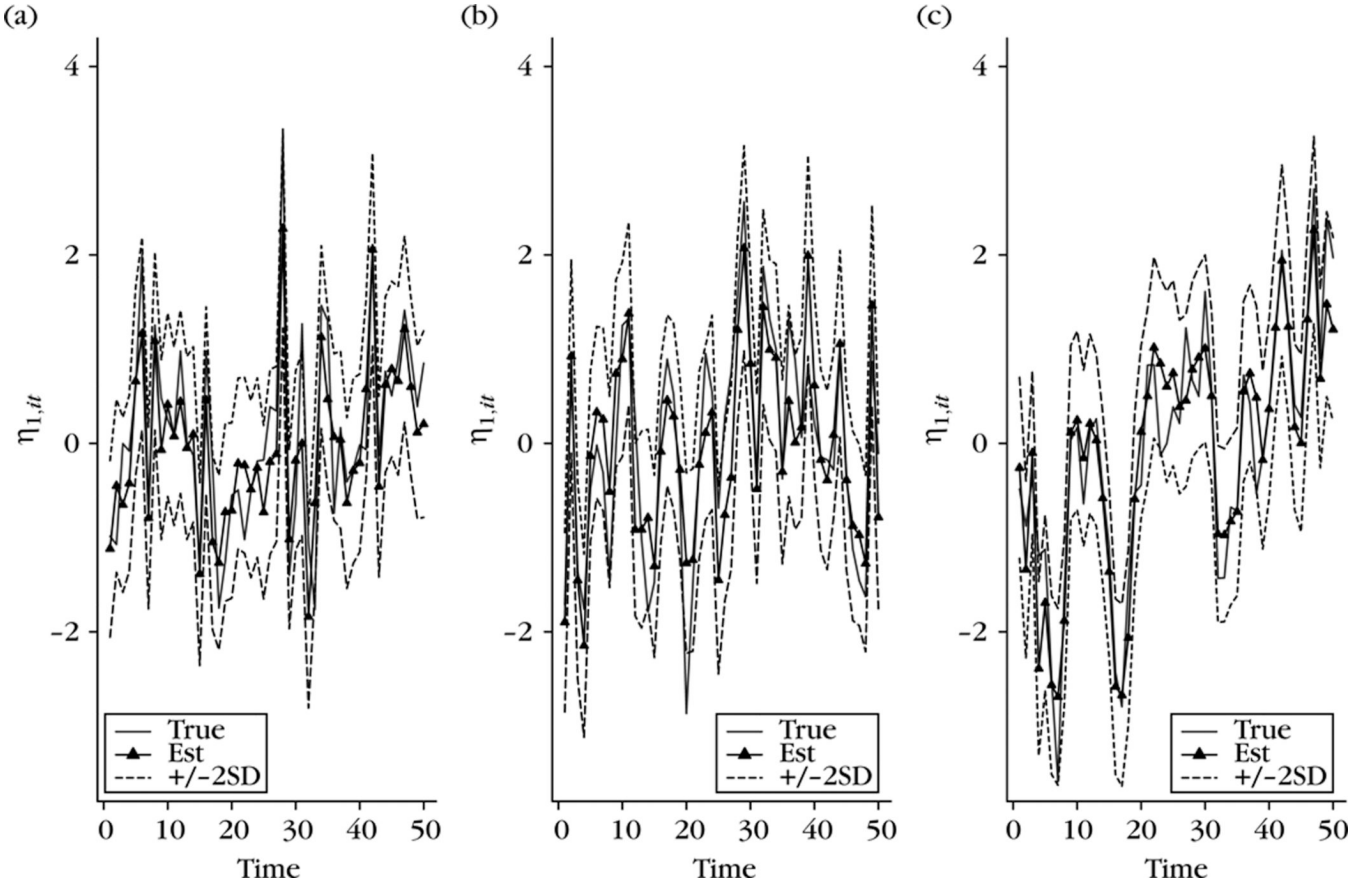

Figure 5.

True factor scores, estimated factor scores and intervals constructed from twice the standard deviation of the posterior samples of the first latent variable from for one randomly selected hypothetical subject in (a) condition 1, (b) condition 2, and (c) condition 3.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

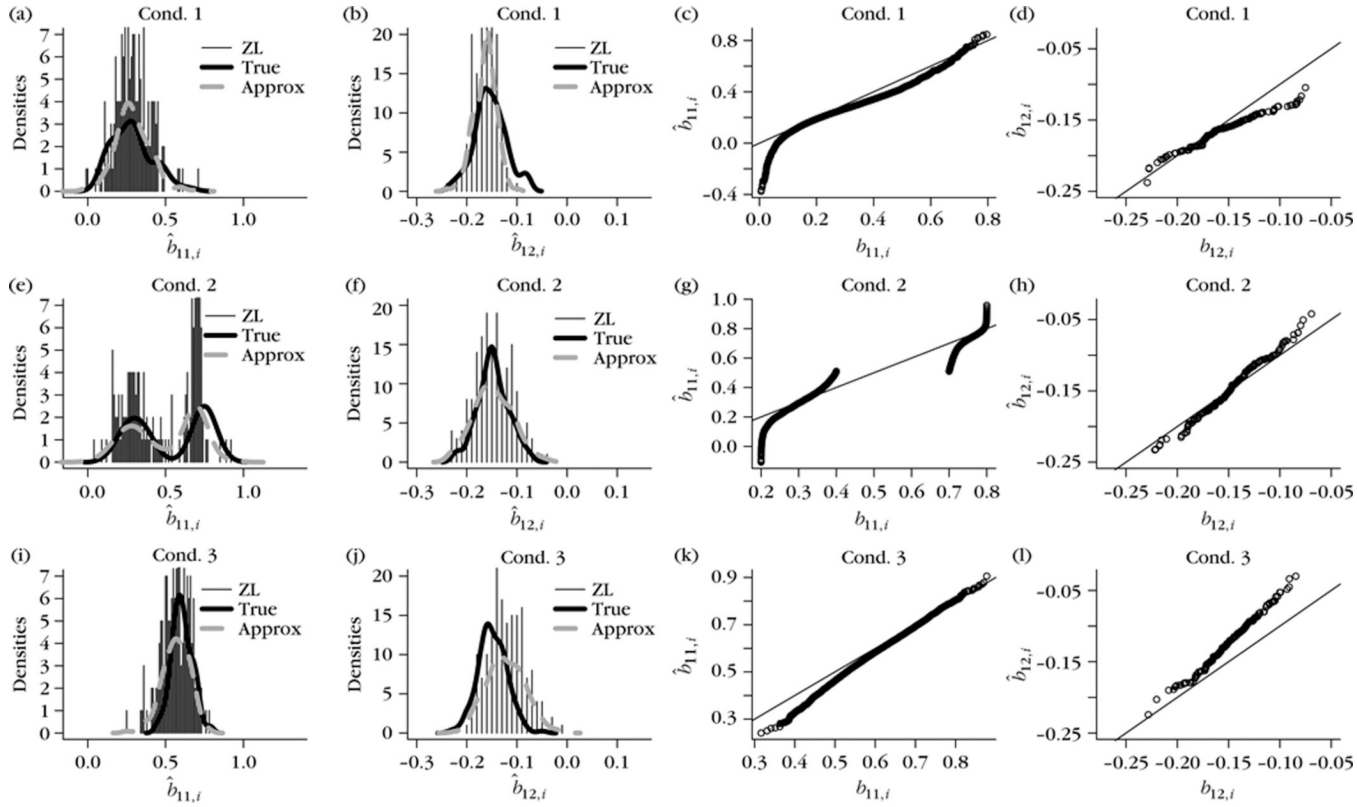

Figure 6.

(a), (e), (i) True density of $b_{11,i}$, the approximation density constituted by $\hat{b_i}$ for conditions 1, 2, and 3, respectively. (b), (f), (j) The corresponding plots associated with $b_{12,i}$. (c), (g), (k) Quantile–quantile plots comparing the true and estimated $b_{11,i}$ pooled across all Monte Carlo runs; the straight line provides the reference for $y = x$. (d), (h), (l) Quantile–quantile plot comparing the true and estimated $b_{12,i}$; the straight line provides the reference for $y = x$.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Bayesian estimates of time- and person-invariant parameters from fitting the nonlinear DFA model to empirical data using the DP prior

Parameters	Mean	*SD*	5th percentile	95th percentile
$[\lambda_{21}, \lambda_{31}, \lambda_{41}]$	[1.08, 1.03, 0.86]	[0.02, 0.02, 0.02]	[1.05, 1.00, 0.84]	[1.11, 1.06, 0.89]
$[\lambda_{62}, \lambda_{72}, \lambda_{82}]$	[1.03, 0.77, 0.69]	[0.04, 0.03, 0.02]	[0.97, 0.72, 0.65]	[1.09, 0.83, 0.73]
ψ_{ε_1}	0.41	0.01	0.39	0.43
ψ_{ε_2}	0.35	0.01	0.34	0.37
ψ_{ε_3}	0.40	0.01	0.38	0.41
ψ_{ε_4}	0.55	0.01	0.53	0.57
ψ_{ε_5}	0.43	0.02	0.39	0.47
ψ_{ε_6}	0.29	0.01	0.26	0.31
ψ_{ε_7}	0.54	0.03	0.48	0.60
ψ_{ε_8}	0.63	0.02	0.60	0.66
μ_1	0.05	0.03	0.01	0.09
μ_2	0.03	0.03	−0.02	0.08
μ_3	0.03	0.03	−0.01	0.08
μ_4	0.01	0.03	−0.03	0.05
μ_5	0.12	0.03	0.08	0.17
μ_6	0.13	0.03	0.09	0.19
μ_7	0.10	0.02	0.06	0.14
μ_8	0.10	0.02	0.05	0.13
$[\tau_{12}, \tau_{13}, \tau_{14}, \tau_{15}]$	[−0.21, 0.39, 0.80, 1.39]	[0.01, 0.02, 0.02, 0.02]	[−0.23, 0.36, 0.77, 1.35]	[−0.19, 0.42, 0.83, 1.42]
$[\tau_{22}, \tau_{23}, \tau_{24}, \tau_{25}]$	[−1.16, −0.47, 0.12, 0.75]	[0.02, 0.02, 0.02, 0.02]	[−1.19, −0.49, 0.09, 0.72]	[−1.13, −0.44, 0.15, 0.78]
$[\tau_{32}, \tau_{33}, \tau_{34}, \tau_{35}]$	[−0.88, −0.31, 0.21, 0.88]	[0.01, 0.01, 0.02, 0.02]	[−0.90, −0.34, 0.18, 0.85]	[−0.86, −0.29, 0.23, 0.91]
$[\tau_{42}, \tau_{43}, \tau_{44}, \tau_{45}]$	[−0.60, 0.20, 0.66, 1.22]	[0.01, 0.02, 0.02, 0.02]	[−0.62, 0.17, 0.62, 1.18]	[−0.58, 0.22, 0.69, 1.25]
$[\tau_{52}, \tau_{53}, \tau_{54}, \tau_{55}]$	[0.91, 1.73, 2.24, 2.71]	[0.03, 0.05, 0.06, 0.06]	[0.86, 1.65, 2.15, 2.61]	[0.95, 1.81, 2.33, 2.80]
$[\tau_{62}, \tau_{63}, \tau_{64}, \tau_{65}]$	[0.68, 1.48, 1.95, 2.45]	[0.02, 0.04, 0.05, 0.05]	[0.64, 1.42, 1.87, 2.35]	[0.72, 1.55, 2.04, 2.54]
$[\tau_{72}, \tau_{73}, \tau_{74}, \tau_{75}]$	[0.80, 1.65, 2.14, 2.66]	[0.03, 0.06, 0.07, 0.08]	[0.74, 1.55, 2.01, 2.52]	[0.85, 1.74, 2.25, 2.79]
$[\tau_{82}, \tau_{83}, \tau_{84}, \tau_{85}]$	[0.20, 0.98, 1.48, 1.95]	[0.02, 0.02, 0.03, 0.03]	[0.18, 0.94, 1.43, 1.90]	[0.23, 1.02, 1.53, 2.00]
ϕ_{11}	0.11	0.00	0.10	0.11
ϕ_{12}	−0.10	0.01	−0.11	−0.09
ϕ_{22}	0.23	0.01	0.21	0.25

Note. The values of the lowest thresholds of the ordinal items, τ_{11}, τ_{21}, τ_{31}, τ_{41}, τ_{51}, τ_{61}, τ_{71}, τ_{81}, were set to −0.71, −1.88, −1.44, −1.34, −0.33, −0.26, −0.13, and −0.67, respectively. The values of the highest thresholds of the ordinal items, τ_{16}, τ_{26}, τ_{36}, τ_{46}, τ_{56}, τ_{66}, τ_{76}, τ_{86}, were set to 2.16, 1.71, 1.72, 1.91, 3.09, 2.95, 3.12, and 2.44, respectively.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Bayesian estimates of time- and person-invariant parameters from fitting the nonlinear DFA model to empirical data using the multivariate normal prior

Parameters	Mean	*SD*	5th percentile	95th percentile
$[\lambda_{21}, \lambda_{31}, \lambda_{41}]$	[1.08, 1.02, 0.86]	[0.02, 0.02, 0.02]	[1.05, 1.00, 0.83]	[1.11, 1.06, 0.89]
$[\lambda_{62}, \lambda_{72}, \lambda_{82}]$	[1.03, 0.77, 0.69]	[0.03, 0.03, 0.02]	[0.99, 0.73, 0.66]	[1.08, 0.81, 0.72]
ψ_{ε_1}	0.41	0.01	0.39	0.43
ψ_{ε_2}	0.35	0.01	0.33	0.36
ψ_{ε_3}	0.40	0.01	0.38	0.42
ψ_{ε_4}	0.55	0.01	0.53	0.57
ψ_{ε_5}	0.43	0.02	0.40	0.46
ψ_{ε_6}	0.29	0.01	0.27	0.31
ψ_{ε_7}	0.54	0.03	0.49	0.59
ψ_{ε_8}	0.63	0.02	0.60	0.66
μ_1	0.04	0.02	0.01	0.07
μ_2	0.01	0.02	−0.03	0.05
μ_3	0.02	0.02	−0.01	0.05
μ_4	0.00	0.02	−0.03	0.04
μ_5	0.00	0.02	−0.03	0.03
μ_6	0.01	0.02	−0.02	0.04
μ_7	0.00	0.02	−0.03	0.03
μ_8	0.00	0.02	−0.02	0.03
$[\tau_{12}, \tau_{13}, \tau_{14}, \tau_{15}]$	[−0.21, 0.39, 0.80, 1.39]	[0.01, 0.02, 0.02, 0.02]	[−0.23, 0.36, 0.77, 1.35]	[−0.19, 0.41, 0.83, 1.42]
$[\tau_{22}, \tau_{23}, \tau_{24}, \tau_{25}]$	[−1.16, −0.47, 0.12, 0.75]	[0.02, 0.02, 0.02, 0.02]	[−1.19, −0.50, 0.09, 0.72]	[−1.13, −0.44, 0.15, 0.78]
$[\tau_{32}, \tau_{33}, \tau_{34}, \tau_{35}]$	[−0.88, −0.31, 0.21, 0.88]	[0.01, 0.02, 0.02, 0.02]	[−0.90, −0.34, 0.18, 0.85]	[−0.86, −0.29, 0.23, 0.91]
$[\tau_{42}, \tau_{43}, \tau_{44}, \tau_{45}]$	[−0.60, 0.19, 0.65, 1.21]	[0.01, 0.02, 0.02, 0.02]	[−0.63, 0.17, 0.62, 1.18]	[−0.58, 0.22, 0.69, 1.25]
$[\tau_{52}, \tau_{53}, \tau_{54}, \tau_{55}]$	[0.91, 1.74, 2.24, 2.72]	[0.02, 0.04, 0.04, 0.05]	[0.88, 1.68, 2.17, 2.63]	[0.95, 1.80, 2.32, 2.79]
$[\tau_{62}, \tau_{63}, \tau_{64}, \tau_{65}]$	[0.69, 1.50, 1.98, 2.48]	[0.02, 0.03, 0.04, 0.05]	[0.65, 1.45, 1.91, 2.39]	[0.72, 1.56, 2.05, 2.55]
$[\tau_{72}, \tau_{73}, \tau_{74}, \tau_{75}]$	[0.79, 1.64, 2.13, 2.66]	[0.03, 0.05, 0.06, 0.06]	[0.75, 1.57, 2.04, 2.56]	[0.84, 1.73, 2.23, 2.77]
$[\tau_{82}, \tau_{83}, \tau_{84}, \tau_{85}]$	[0.20, 0.98, 1.48, 1.95]	[0.02, 0.02, 0.03, 0.03]	[0.18, 0.94, 1.43, 1.90]	[0.23, 1.02, 1.53, 2.00]
ϕ_{11}	0.11	0.00	0.10	0.12
ϕ_{12}	−0.10	0.00	−0.11	−0.09
ϕ_{22}	0.23	0.01	0.22	0.25

Note. The values of the lowest thresholds of the ordinal items, τ_{11}, τ_{21}, τ_{31}, τ_{41}, τ_{51}, τ_{61}, τ_{71}, τ_{81}, were set to −0.71, −1.88, −1.44, −1.34, −0.33, −0.26, −0.13, and −0.67, respectively. The values of the highest thresholds of the ordinal items, τ_{16}, τ_{26}, τ_{36}, τ_{46}, τ_{56}, τ_{66}, τ_{76}, τ_{86}, were set to 2.16, 1.71, 1.72, 1.91, 3.09, 2.95, 3.12, and 2.44, respectively.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Bayesian estimates of the person-invariant parameters with beta-distributed baseline autoregressive parameters (condition 1) in the simulation study

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
[λ_{21}, λ_{31}, λ_{41}]	[0.02, 0.02, 0.02]	[0.02, 0.02, 0.02]	[0.000, −0.006, −0.005]	[0.02, 0.02, 0.02]	[83, 85, 82]
[λ_{62}, λ_{72}, λ_{82}]	[0.02, 0.02, 0.02]	[0.02, 0.02, 0.02]	[−0.002, −0.001, 0.000]	[0.02, 0.02, 0.02]	[87, 89, 92]
ψ_{ε_1}	0.02	0.02	0.003	0.02	87
ψ_{ε_2}	0.02	0.02	0.005	0.02	90
ψ_{ε_3}	0.02	0.02	−0.005	0.02	89
ψ_{ε_4}	0.02	0.02	0.004	0.02	90
ψ_{ε_5}	0.02	0.02	0.003	0.03	91
ψ_{ε_6}	0.02	0.02	0.002	0.02	85
ψ_{ε_7}	0.02	0.02	−0.001	0.02	86
ψ_{ε_8}	0.02	0.02	0.005	0.02	87
μ_1	0.03	0.03	0.000	0.03	80
μ_2	0.03	0.03	−0.002	0.03	88
μ_3	0.03	0.03	0.006	0.03	85
μ_4	0.03	0.03	0.004	0.03	89
μ_5	0.02	0.02	0.006	0.02	90
μ_6	0.02	0.02	0.002	0.02	93
μ_7	0.02	0.02	0.002	0.02	90
μ_8	0.02	0.02	0.004	0.02	89
[τ_{12}, τ_{13}, τ_{14}, τ_{15}]	[0.03, 0.03, 0.02, 0.02]	[0.04, 0.04, 0.03, 0.03]	[−0.003, −0.003, −0.003, −0.001]	[0.04, 0.04, 0.03, 0.03]	[84, 78, 79, 76]
[τ_{22}, τ_{23}, τ_{24}, τ_{25}]	[0.04, 0.03, 0.03, 0.02]	[0.05, 0.04, 0.03, 0.03]	[−0.006, −0.007, −0.004, −0.003]	[0.05, 0.04, 0.03, 0.03]	[83, 83, 83, 81]
[τ_{32}, τ_{33}, τ_{34}, τ_{35}]	[0.04, 0.03, 0.03, 0.02]	[0.04, 0.03, 0.03, 0.02]	[0.009, 0.009, 0.005, 0.003]	[0.04, 0.03, 0.03, 0.02]	[89, 85, 87, 91]
[τ_{42}, τ_{43}, τ_{44}, τ_{45}]	[0.04, 0.03, 0.03, 0.03]	[0.04, 0.04, 0.03, 0.03]	[0.009, 0.004, 0.005, 0.005]	[0.04, 0.04, 0.03, 0.03]	[90, 82, 80, 84]
[τ_{52}, τ_{53}, τ_{54}, τ_{55}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[0.001, 0.001, 0.003, 0.003]	[0.01, 0.02, 0.02, 0.02]	[88, 89, 82, 85]
[τ_{62}, τ_{63}, τ_{64}, τ_{65}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[0.000, 0.000, −0.001, −0.001]	[0.01, 0.02, 0.02, 0.02]	[86, 85, 82, 85]
[τ_{72}, τ_{73}, τ_{74}, τ_{75}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[−0.001, −0.002, −0.001, −0.002]	[0.01, 0.02, 0.02, 0.02]	[89, 84, 88, 86]
[τ_{82}, τ_{83}, τ_{84}, τ_{85}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[−0.001, 0.001, 0.002, 0.002]	[0.01, 0.02, 0.02, 0.02]	[87, 89, 84, 85]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
ϕ_{11}	0.03	0.04	0.007	0.04	79
ϕ_{12}	0.02	0.02	0.002	0.02	89
ϕ_{22}	0.03	0.04	0.001	0.04	88

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Bayesian estimates of the person-invariant parameters with bimodal baseline autoregressive parameter distributions (condition 2) in the simulation study

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
[λ_{21}, λ_{31}, λ_{41}]	[0.02, 0.02, 0.02]	[0.01, 0.02, 0.02]	[0.002, 0.001, 0.002]	[0.01, 0.02, 0.02]	[90, 84, 89]
[λ_{62}, λ_{72}, λ_{82}]	[0.02, 0.02, 0.02]	[0.02, 0.01, 0.01]	[0.000, 0.000, 0.001]	[0.02, 0.01, 0.01]	[87, 91, 93]
ψ_{ε_1}	0.02	0.03	−0.001	0.02	83
ψ_{ε_2}	0.02	0.02	0.003	0.02	88
ψ_{ε_3}	0.02	0.02	−0.000	0.02	90
ψ_{ε_4}	0.02	0.02	0.004	0.02	88
ψ_{ε_5}	0.02	0.02	−0.000	0.03	87
ψ_{ε_6}	0.02	0.02	0.000	0.02	88
ψ_{ε_7}	0.02	0.02	−0.001	0.02	90
ψ_{ε_8}	0.02	0.02	0.001	0.02	87
μ_1	0.03	0.03	0.008	0.03	85
μ_2	0.03	0.03	−0.001	0.03	86
μ_3	0.03	0.03	0.005	0.03	89
μ_4	0.03	0.03	0.002	0.03	83
μ_5	0.02	0.02	0.005	0.02	91
μ_6	0.02	0.02	0.005	0.02	92
μ_7	0.02	0.02	0.003	0.02	91
μ_8	0.02	0.02	0.002	0.02	92
[τ_{12}, τ_{13}, τ_{14}, τ_{15}]	[0.03, 0.03, 0.02, 0.02]	[0.04, 0.03, 0.03, 0.03]	[0.007, 0.004, 0.006, 0.004]	[0.04, 0.03, 0.03, 0.03]	[87, 89, 80, 85]
[τ_{22}, τ_{23}, τ_{24}, τ_{25}]	[0.04, 0.03, 0.03, 0.02]	[0.04, 0.03, 0.03, 0.03]	[−0.006, −0.004, −0.007, −0.004]	[0.04, 0.03, 0.03, 0.03]	[85, 89, 84, 83]
[τ_{32}, τ_{33}, τ_{34}, τ_{35}]	[0.04, 0.03, 0.02, 0.02]	[0.04, 0.03, 0.03, 0.03]	[0.006, 0.002, −0.002, −0.001]	[0.04, 0.03, 0.03, 0.03]	[88, 82, 85, 84]
[τ_{42}, τ_{43}, τ_{44}, τ_{45}]	[0.04, 0.03, 0.02, 0.03]	[0.04, 0.03, 0.02, 0.02]	[−0.003, −0.000, −0.003, −0.001]	[0.04, 0.03, 0.02, 0.02]	[84, 87, 88, 93]
[τ_{52}, τ_{53}, τ_{54}, τ_{55}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.01, 0.02, 0.01]	[−0.001, 0.001, −0.001, −0.001]	[0.01, 0.01, 0.02, 0.02]	[89, 91, 90, 92]
[τ_{62}, τ_{63}, τ_{64}, τ_{65}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[0.003, 0.002, −0.001, 0.001]	[0.01, 0.02, 0.02, 0.02]	[91, 87, 89, 89]
[τ_{72}, τ_{73}, τ_{74}, τ_{75}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.01, 0.02, 0.02]	[−0.003, −0.003, −0.004, −0.002]	[0.01, 0.01, 0.02, 0.02]	[87, 93, 92, 92]
[τ_{82}, τ_{83}, τ_{84}, τ_{85}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[−0.002, −0.003, −0.002, −0.002]	[0.01, 0.02, 0.02, 0.02]	[92, 88, 88, 88]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
ϕ_{11}	0.03	0.03	−0.001	0.04	89
ϕ_{12}	0.02	0.01	0.006	0.02	93
ϕ_{22}	0.03	0.03	−0.002	0.03	94

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 5

Bayesian estimates of the person-invariant parameters with normally distributed baseline autoregressive parameters (condition 3) in the simulation study

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
[λ_{21}, λ_{31}, λ_{41}]	[0.02, 0.02, 0.02]	[0.02, 0.02, 0.02]	[−0.001, 0.001, −0.002]	[0.02, 0.02, 0.02]	[86, 90, 80]
[λ_{62}, λ_{72}, λ_{82}]	[0.02, 0.02, 0.02]	[0.01, 0.02, 0.01]	[−0.002, 0.000, −0.001]	[0.01, 0.02, 0.01]	[91, 87, 89]
ψ_{ε_1}	0.02	0.02	−0.003	0.02	84
ψ_{ε_2}	0.02	0.02	0.000	0.02	87
ψ_{ε_3}	0.02	0.02	0.002	0.02	90
ψ_{ε_4}	0.02	0.02	0.000	0.02	82
ψ_{ε_5}	0.02	0.02	0.009	0.03	87
ψ_{ε_6}	0.02	0.02	0.002	0.02	85
ψ_{ε_7}	0.02	0.02	0.001	0.02	85
ψ_{ε_8}	0.02	0.02	0.001	0.02	86
μ_1	0.03	0.04	0.004	0.04	84
μ_2	0.03	0.03	0.002	0.03	79
μ_3	0.03	0.03	0.001	0.03	84
μ_4	0.03	0.03	−0.001	0.03	84
μ_5	0.03	0.03	0.000	0.02	76
μ_6	0.02	0.02	−0.001	0.02	83
μ_7	0.02	0.03	−0.001	0.03	82
μ_8	0.02	0.02	−0.001	0.02	81
[τ_{12}, τ_{13}, τ_{14}, τ_{15}]	[0.03, 0.03, 0.02, 0.02]	[0.03, 0.03, 0.02, 0.02]	[0.001, 0.002, 0.003, 0.003]	[0.03, 0.03, 0.02, 0.02]	[82, 84, 87, 85]
[τ_{22}, τ_{23}, τ_{24}, τ_{25}]	[0.04, 0.03, 0.03, 0.02]	[0.04, 0.04, 0.03, 0.03]	[−0.003, −0.001, 0.001, 0.001]	[0.04, 0.04, 0.03, 0.03]	[84, 79, 83, 86]
[τ_{32}, τ_{33}, τ_{34}, τ_{35}]	[0.04, 0.03, 0.03, 0.02]	[0.04, 0.03, 0.03, 0.02]	[−0.003, −0.001, −0.001, −0.003]	[0.04, 0.03, 0.03, 0.02]	[83, 86, 86, 86]
[τ_{42}, τ_{43}, τ_{44}, τ_{45}]	[0.04, 0.03, 0.03, 0.02]	[0.04, 0.03, 0.03, 0.03]	[−0.002, −0.001, −0.001, −0.002]	[0.04, 0.03, 0.03, 0.03]	[92, 85, 89, 89]
[τ_{52}, τ_{53}, τ_{54}, τ_{55}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[0.000, 0.001, 0.002, 0.001]	[0.01, 0.02, 0.02, 0.02]	[89, 90, 84, 89]
[τ_{62}, τ_{63}, τ_{64}, τ_{65}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[−0.001, 0.000, −0.001, 0.000]	[0.01, 0.01, 0.02, 0.02]	[94, 92, 85, 89]
[τ_{72}, τ_{73}, τ_{74}, τ_{75}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[0.000, 0.001, 0.000, −0.001]	[0.01, 0.02, 0.02, 0.02]	[89, 92, 94, 89]
[τ_{82}, τ_{83}, τ_{84}, τ_{85}]	[0.01, 0.02, 0.02, 0.02]	[0.01, 0.02, 0.02, 0.02]	[−0.001, 0.000, −0.001, −0.002]	[0.01, 0.02, 0.02, 0.02]	[93, 92, 87, 98]

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Parameters	Est *SD*	*SD*	Bias	RMSE	Coverage (%)
ϕ_{11}	0.03	0.03	0.003	0.03	91
ϕ_{12}	0.02	0.02	0.001	0.02	93
ϕ_{22}	0.03	0.03	0.006	0.03	92

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 6

Means and standard deviations of the b_i estimates

	Condition 1				Condition 2				Condition 3			
Parameters	**Mean**	**Est Mean**	***SD***	**Est *SD***	**Mean**	**Est Mean**	***SD***	**Est *SD***	**Mean**	**Est Mean**	***SD***	**Est *SD***
$b_{11,i}$	0.27	0.24	0.13	0.12	0.53	0.49	0.23	0.23	0.60	0.60	0.07	0.07
$b_{22,i}$	0.27	0.24	0.13	0.11	0.53	0.49	0.23	0.22	0.59	0.59	0.07	0.08
$b_{12,i}$	−0.15	−0.11	0.03	0.03	−0.15	−0.13	0.03	0.04	−0.15	−0.19	0.03	0.04
$b_{21,i}$	−0.15	−0.11	0.03	0.03	−0.15	−0.12	0.03	0.04	−0.15	−0.17	0.03	0.04

Note. Mean, true empirical mean of the distribution; Est mean, mean of the posterior samples; *SD*, true empirical standard deviation of the distribution; Est *SD*, standard deviation of the posterior samples.