

# Chapter 12

## Advances in Dynamic Factor Analysis of Psychological Processes

Peter C.M. Molenaar and Nilam Ram

The currently dominant, myopic approach to statistical analysis in psychology is based on analysis of inter-individual variation. Differences between subjects drawn from a population of subjects provide the information to make inferences about states of affairs at the population level. For instance, the factor structure of a personality test is determined by drawing a random sample of subjects from the population of interest, estimating the item correlation matrix by pooling across the scores of sampled subjects, and generalizing the results of the ensuing factor analysis to the population of subjects.

Pooling across subjects is the hall-mark of analyses of inter-individual variation. Such pooling implies that the individuality of each subject in the population is immaterial to the statistical analysis—subjects in a homogeneous population are considered to be exchangeable like atoms in a homogeneous gas, constituting replications of each other. Accordingly, the population is conceived of as a set of statistical atoms or clones in which the variation between atoms (inter-individual variation) has the same (factor) structure as the variation of each atom in time (intra-individual variation).

This perspective underlying analyses of inter-individual variation would seem to imply that results which hold in a homogeneous population also apply to each of the individual subjects making up this population. That is, the variation between subjects at each point in time has to be qualitatively and quantitatively the same as the variation which characterizes the life trajectory of each individual subject. Therefore results obtained in analyses of inter-individual variation can be, and abundantly are, applied to assess, counsel and treat individual subjects.

It has been shown recently (Molenaar, 2004), however, that in general the inferred states of affairs at the population level, as determined in analyses of inter-individual variation, do not apply at the level of intra-individual variation characterizing the life trajectories of individual subjects making up the population. This is a direct consequence of general mathematical theorems—the so-called classical ergodic

---

P.C.M. Molenaar (✉)  
Developmental Psychology  
Pennsylvania State University  
University Park, PA 16802-1294, Pennsylvania, USA  
e-mail: pxm21@psu.edu

theorems—which have far-reaching implications for the way in which psychological processes have to be analyzed.

These theorems, which are heuristically described below, imply the necessity of using alternative approaches for the analysis of intra-individual variation, based on single-subject and replicated time series analysis. Next an overview is presented of dynamic factor models for the analysis of multivariate time series and the various ways to fit these models to the data. Some persistent misunderstandings in the recent literature on dynamic factor analysis will be addressed and an illustrative empirical application of factor analysis of mood change during pregnancy data is presented. The next topic is innovative—a new dynamic factor model for the analysis of multivariate time series having time-varying statistical characteristics is introduced and applied to simulated data. In the closing section future extensions of dynamic factor analysis are outlined.

## The Classical Ergodic Theorems

The standard approach to statistical analysis in psychology is to draw a random sample of subjects from a presumably homogeneous population of subjects, analyze the structure of inter-individual variation in this sample, and then generalize the results thus obtained to the population. Such analysis of inter-individual variation underlies all known statistical techniques, like analysis of variance, regression analysis, factor analysis, multilevel modeling, mixture modeling, etc. Consequently the standard approach to psychological data analysis aims to describe the state of affairs at the population level, not at the level of individual subjects. Accordingly, the individuality of each of the persons in the sample and population is deemed immaterial: the subjects are considered to be replications devoid of individuality. This is expressed by the assumption that subjects are homogeneous in all respects relevant to the analysis. This essential homogeneity assumption allows for the averaging (pooling) of the scores of the sampled subjects in the determination of statistics (e.g., means, variances, correlations, model parameters) to be generalized to the population.

However, the standard approach based on analysis of inter-individual variation is incorrect if a psychological process under investigation does not obey stringent conditions (Molenaar, 2004). The proof is based on the classical ergodic theorems; theorems of extreme generality which apply to all measurable processes irrespective of their content (cf. Choe, 2005, for a modern proof of the first ergodic theorem of Birkhoff). The conditions concerned are specified at the close of this section.

To appreciate the far-reaching implications of the classical ergodic theorems, it is helpful to first characterize the elementary methodological situation in psychological measurement. Instead of postulating an abstract population of subjects, consider an ensemble of actually existing human subjects whose measurable psychological processes are functions of time and place (the basic Kantian dimensions of phenomenological reality).

To simplify the following discussion, without affecting its generality, the focus will be on time as the basic dimension along which psychological processes are

evolving. The ensuing basic scientific representation of a human subject in psychology therefore is a high-dimensional dynamic system, the output of which consists of a set of time-dependent processes. The system includes important functional subsystems such as the perceptual, emotional, cognitive, and physiological systems, as well as their dynamic interrelationships. The complete set of measurable time-dependent variables characterizing the system's behavior can be represented as the coordinates of a high-dimensional space which will be referred to as the behavior space. The behavior space contains all the scientifically relevant information about a person (cf. De Groot, 1954).

Within the behavior space, inter-individual variation is defined as follows:

- (i) select a fixed subset of variables;
- (ii) select one or more fixed time points as measurement occasions,
- (iii) determine the variation of the scores on the selected variables at the selected time points by pooling across subjects.

Analysis of inter-individual variation thus defined is called R-technique by Cattell (1952).

In contrast, intra-individual variation is defined as follows:

- (i) select a fixed subset of variables;
- (ii) select a fixed subject;
- (iii) determine the variation of the scores of the single subject on the selected variables by pooling across time points.

Analysis of intra-individual variation thus defined is called P-technique by Cattell (1952).

With these preliminary specifications in place, the following heuristic description of the content of the classical ergodic theorems can be given. These theorems detail the conditions that must be met in order to generalize from analyses of inter-individual variation to analyses of intra-individual variation, and vice versa.

The conditions of the ergodic theorems are twofold. First, a process has to be stationary, meaning that the mean function must be constant in time (without trends or cycles) and the sequential dependence must be constant in time (with constant variance and sequential correlations depending only upon the relative distance between time points; cf. Hannan, 1970).

Second, each person in the population must obey the same dynamics. If a dynamic process obeys both conditions, it is called ergodic; if one or both conditions are violated, it is called non-ergodic. For ergodic processes, lawful relationships between inter- and intra-individual variation exist, but for non-ergodic processes these relationships do not exist. Put another way, if the conditions of ergodicity are violated, no a priori relationship exists between results obtained in an analysis of inter-individual variation (R-technique) and results obtained in an analogous analysis of intra-individual variation (P-technique).

The consequences of the classical ergodic theorems affect all psychological statistical methodology (e.g., Molenaar, Huizenga, & Nesselroade, 2003; Borsboom, 2005). Because "development" generally implies that some kind of growth or decline occurs, developmental processes are almost always non-stationary and are,

therefore, non-ergodic. Generally, developmental scientists consider change that occurs in average or mean levels of a process. However, change may also occur in variances or sequential dependencies over time.

## Overview of Dynamic Factor Modeling

Dynamic factor analysis is factor analysis of single-subject multivariate time series. It constitutes a generalization of Cattell's P-technique (Cattell, 1952) in that it takes account of lead-lag patterns in the dynamic relationships between latent factor series and observed series. In contrast, P-technique involves straight-forward application of standard factor analysis to multivariate time series without accommodation of lead-lag sequential dependencies; the reader is referred to Molenaar and Nesselroade (2008) for further discussion of the domain of application of P-technique.

In what follows, bold face lower case letters denote vectors; bold face upper case letters denote matrices; an apostrophe attached to vectors or matrices denotes transposition. Let  $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_p(t)]'$  be a  $p$ -variate time series,  $p \geq 1$ , observed at equidistant time points  $t = 1, 2, \dots, T$ . The mean of  $\mathbf{y}(t)$  at each time point  $t$  is:  $E[\mathbf{y}(t)] = \boldsymbol{\mu}(t)$ . Considered as function of  $t$ ,  $\boldsymbol{\mu}(t)$  denotes the  $p$ -variate mean function (trend) of  $\mathbf{y}(t)$ . If  $\boldsymbol{\mu}(t) = \boldsymbol{\mu}$ , i.e., if the mean function is constant in time, then  $\mathbf{y}(t)$  has a stationary mean function. The sequential covariance of  $\mathbf{y}(t)$  between a given pair of time points  $t_1$  and  $t_2$  is defined as:  $\boldsymbol{\Sigma}(t_1, t_2) = \text{cov}[\mathbf{y}(t_1), \mathbf{y}(t_2)']$ . Considered as function of 2-D time,  $\boldsymbol{\Sigma}(t_1, t_2)$  denotes the  $(p,p)$ -variate covariance function of  $\mathbf{y}(t)$ . If  $\boldsymbol{\Sigma}(t_1, t_2)$  only depends upon the relative time difference  $t_1 - t_2 = u$ , i.e.,  $\boldsymbol{\Sigma}(t_1, t_2) = \boldsymbol{\Sigma}(t_1 - t_2) = \boldsymbol{\Sigma}(u)$ ,  $u = 0, \pm 1, \dots, \pm T - 1$ , then  $\mathbf{y}(t)$  has stationary covariance function depending only on lag  $u$ . If both the mean function and covariance function of  $\mathbf{y}(t)$  are stationary then  $\mathbf{y}(t)$  is called a weakly stationary  $p$ -variate time series.

In the first publication on dynamic factor analysis in psychology the following model for weakly stationary multivariate Gaussian series was considered (Molenaar, 1985):

$$\mathbf{y}(t) = \boldsymbol{\mu} + \boldsymbol{\Lambda}(0)\boldsymbol{\eta}(t) + \boldsymbol{\Lambda}(1)\boldsymbol{\eta}(t-1) + \dots + \boldsymbol{\Lambda}(s)\boldsymbol{\eta}(t-s) + \boldsymbol{\epsilon}(t) \quad (1a)$$

where  $\mathbf{y}(t)$  is an observed  $p$ -variate time series,  $\boldsymbol{\eta}(t)$  is a latent  $q$ -variate factor series and  $\boldsymbol{\epsilon}(t)$  is a  $p$ -variate measurement error series. Because our main interest is not in the constant mean function  $\boldsymbol{\mu}$ , it is conveniently assumed that  $\boldsymbol{\mu} = \mathbf{0}$ . Then all time series in (1a),  $\mathbf{y}(t)$ ,  $\boldsymbol{\eta}(t)$ , and  $\boldsymbol{\epsilon}(t)$ , are zero mean weakly stationary.

The  $\boldsymbol{\Lambda}(u)$ ,  $u = 0, 1, \dots, s$ , are  $(p,q)$ -dimensional matrices of lagged factor loadings, where  $s \geq 0$  is the maximum lag. These lagged factor loadings allow for the possibility that the realization of the latent factor series  $\boldsymbol{\eta}(t)$  at each time  $t$  not only has an instantaneous effect on  $\mathbf{y}(t)$ , but also may have delayed effects at later time points  $t+1, \dots, t+s$ . The linear combination  $\boldsymbol{\Lambda}(0)\boldsymbol{\eta}(t) + \boldsymbol{\Lambda}(1)\boldsymbol{\eta}(t-1) + \dots + \boldsymbol{\Lambda}(s)\boldsymbol{\eta}(t-s)$  is called a convolution.

For later reference the limiting case of (1a) is considered in which  $s=0$ , i.e., the case in which there are no lagged factor loadings:

$$\mathbf{y}(t) = \mathbf{\Lambda}(0)\boldsymbol{\eta}(t) + \boldsymbol{\varepsilon}(t) \quad (1b)$$

Equation (1b) is a special instance of (1a). It has been assigned several labels, including state-space model (e.g., Molenaar 1985) and process factor model (e.g., Browne & Nesselroade, 2005). In what follows (1b) will be referred to as a state-space model. As will be explained shortly, (1b) has a special property not shared by (1a).

The dynamic factor model (1a) was inspired by Brillinger's (1975) principal component analysis of multivariate weakly stationary time series. It differs from Brillinger's approach in a number of respects, perhaps the most important of which is that in (1a) the convolution of lagged factor loadings and latent factor series is finite and only depends at each time point  $t$  upon earlier realizations  $\boldsymbol{\eta}(t), \dots, \boldsymbol{\eta}(t-s)$  of the latent factor series, whereas in Brillinger's model this convolution is infinite and also depends upon future realizations  $\boldsymbol{\eta}(t+1), \boldsymbol{\eta}(t+2), \dots$ . Another important difference relates to the statistical characteristics of the measurement error series in each of the two models.

To complete the definition of the dynamic factor model under consideration, the covariance functions of the time series occurring in (1a)–(1b) have to be specified. Let  $\text{diag-}\mathbf{A}$  denote a square diagonal matrix  $\mathbf{A}$  (all off-diagonal elements being zero). Then the covariance functions associated with the right-hand side of (1a) are:

$$\begin{aligned} \text{cov}[\boldsymbol{\varepsilon}(t), \boldsymbol{\varepsilon}(t-u)'] &= \text{diag-}\boldsymbol{\Theta}(u); u = 0, \pm 1, \dots \\ \text{cov}[\boldsymbol{\eta}(t), \boldsymbol{\eta}(t-u)'] &= \boldsymbol{\Psi}(u) = 0, \pm 1, \dots \end{aligned} \quad (1c)$$

The first equation of (1c) defines the covariance function of the measurement error process. The univariate measurement error process  $\varepsilon_k(t)$  associated with the  $k$ th observed univariate series  $y_k(t)$ ,  $k \in \{1, \dots, p\}$ , is allowed to have nonzero sequential covariance:  $\text{cov}[\varepsilon_k(t), \varepsilon_k(t-u)] \neq 0$  for  $\forall u$ . However, measurement error processes  $\varepsilon_k(t)$  and  $\varepsilon_m(t)$  associated with different observed univariate series  $y_k(t)$  and  $y_m(t)$ ,  $k \neq m \in \{1, \dots, p\}$ , are assumed to be uncorrelated at all lags  $u$ :  $\text{cov}[\varepsilon_k(t), \varepsilon_m(t-u)] = 0$  for  $\forall u$ . The second equation in (1c) defines the covariance function of the latent factor series.

**Some intricacies associated with (1)** It was proven in Molenaar (1985) that under certain conditions the covariance function of the latent factor series defined in (1c) is not identifiable. That means that under certain conditions the variances and sequential covariances in  $\boldsymbol{\Psi}(u)$ ,  $u = 0, \pm 1, \dots$ , cannot be estimated, but have to be fixed a priori. The conditions concerned are twofold. Firstly, the maximum lag  $s$  of the matrices of factor loadings  $\mathbf{\Lambda}(u)$ ,  $u = 0, 1, \dots, s$ , has to be larger than zero:  $s > 0$ . Secondly, all factor loadings in  $\mathbf{\Lambda}(u)$ ,  $u = 0, 1, \dots, s$ , should be free parameters. That is, the dynamic factor model should be exploratory, having no a priori pattern of fixed factor loadings.

If both these conditions obtain then the covariance function of the latent factor series has to be fixed a priori.  $\boldsymbol{\Psi}(u)$ ,  $u=0, \pm 1, \dots$ , then can be fixed at any possible covariance function without affecting the goodness of fit of the model.

In Molenaar (1985) the simplest covariance function for  $\boldsymbol{\eta}(t)$  was chosen:  $\text{cov}[\boldsymbol{\eta}(t), \boldsymbol{\eta}(t-u)] = \delta(u)\mathbf{I}_q$ , where  $\delta(u)$  is the Kronecker delta ( $\delta(u) = 1$  if  $u = 0$ ;  $\delta(u) = 0$  if  $u \neq 0$ ) and  $\mathbf{I}_q$  is the  $(q,q)$ -dimensional identity matrix. This particular choice implies that the latent factor series lacks instantaneous as well as sequential dependencies. Accordingly,  $\boldsymbol{\eta}(t)$  can be conceived of as a sequence of random shocks, often referred to in the engineering literature as a white noise sequence. But quite other choices are possible (cf. Molenaar & Nesselrode, 2001).

If the dynamic factor model is confirmatory, i.e., if an a priori pattern of fixed factor loadings has been specified in  $\boldsymbol{\Lambda}(u)$ ,  $u=0,1,\dots,s$ , then the covariance function of the latent factor series is identifiable. Also if  $s=0$ , i.e., if the state space model (1b) applies, then the covariance function of the latent factor series is identifiable. In these cases  $\boldsymbol{\Psi}(u)$ ,  $u = 0, \pm 1, \dots$ , can be freely estimated or, alternatively, a parametric time series model for  $\boldsymbol{\eta}(t)$  (and hence for its covariance function) can be considered. For instance, the latent factor series  $\boldsymbol{\eta}(t)$  can be represented by a vector autoregression:  $\boldsymbol{\eta}(t) = \mathbf{B}\boldsymbol{\eta}(t-1) + \boldsymbol{\zeta}(t)$ , where  $\mathbf{B}$  is a  $(q,q)$ -dimensional matrix of regression coefficients and  $\boldsymbol{\zeta}(t)$  is a  $q$ -variate white noise sequence.

Given that for the state-space model (1b) the covariance function (or a parametric time series model) of the latent factor series always is identifiable, it would seem rational to restrict attention to this type of model. At least for exploratory dynamic factor analyses, this would preclude the need to have to arbitrarily fix  $\boldsymbol{\Psi}(u)$ ,  $u=0,\pm 1,\dots$ , which is necessary in such applications of (1a). However, it can be shown that for certain types of psychological time series the state-space model is too restrictive. In particular when the effect of  $\boldsymbol{\eta}(t)$  on  $\mathbf{y}(t)$  is delayed, the state space model will be inappropriate. For these time series one needs the general dynamic factor model (1a) in which  $s > 0$ . Such delays occur, for instance, in multi-lead EEG registrations of electrocortical brain fields caused by a finite set of underlying neural sources. The activity of the neural sources is represented by the latent factor series  $\boldsymbol{\eta}(t)$  and the EEG registrations by the manifest series  $\mathbf{y}(t)$ . The effects of each neural source travel with finite speed along long-range axons to their target regions. Because target regions are located at different distances from a given neural source and the effects of a volley of action potentials of a neural source on the ongoing activity of target areas take time to dissipate, the relationships between  $\boldsymbol{\eta}(t)$  and  $\mathbf{y}(t)$  will show complex patterns of delays.

But the pattern of delayed relationships between the latent factor series and the observed series does not have to be complex in order to invalidate the state space model (1b). This will be illustrated with data simulated with the following simple dynamic 1-factor model. Let  $\mathbf{y}(t)$  be a 5-variate time series and  $\boldsymbol{\eta}(t)$  a univariate latent factor series. Let  $\boldsymbol{\Lambda}(0) = [1.0, 0.9, 0.8, 0.7, 0.6]'$  and  $\boldsymbol{\Lambda}(1) = [0, -0.5, 0, 0, 0]'$ . Hence  $s=1$ . Notice that all elements of  $\boldsymbol{\Lambda}(1)$  are zero, except the lagged loading on  $y_2(t)$ :  $\lambda_2(1) = -0.5$ . Hence there only is a simple delayed effect of  $\boldsymbol{\eta}(t)$  on  $y_2(t)$ . Let all measurement errors be white noise series having variance equal to  $\theta_k(0) = 0.5$ ,  $k = 1, \dots, 5$ . The autoregressive model for the univariate latent factor series is:  $\boldsymbol{\eta}(t) = 0.7\boldsymbol{\eta}(t-1) + \boldsymbol{\zeta}(t)$ , where the variance of the white process noise is  $\text{var}[\boldsymbol{\zeta}(t)] = 1.0$ .

A 5-variate time series of length  $T=400$  has been generated. Although the simulation model almost is a state-space model (save for one delayed factor loading  $\lambda_2(1) = -0.5$ ), the state-space model (1b) with univariate state process  $\boldsymbol{\eta}(t)$  does

not fit the data. Using a window width of 5 (cf. Molenaar, 1985), chi-square=751.95, degrees of freedom=313, prob.=0.0; non-normed fit index=0.93; comparative fit index=0.93. The non-normed fit index and the comparative fit index should have values larger than 0.95 for acceptable model fits. In contrast, the general dynamic 1-factor model (1a) with  $s=5$  yields an acceptable fit: chi-square=235.78, degrees of freedom=295, prob.=1.0; non-normed fit index=1.0; comparative fit index=1.0.

**Alternative ways to fit (1)** Analysis of stationary multivariate Gaussian time series based on the dynamic factor model (1) can proceed in various ways:

- 1) Based on the sequential covariance function, arranged in a so-called block-Toeplitz matrix (see Molenaar, 1985, for a detailed description). This can be carried out by means of commercially available structural equation modeling software. Browne and Zhang (2005, 2007) provide an alternative method for fitting a dynamic factor analysis model to a sequential autocorrelation function. Their approach, implemented in the DyFA computer program, does not use a block-Toeplitz matrix in the estimation process. It fits the model directly to the autocorrelation function without duplicating the constituent autocorrelation matrices to form a block-Toeplitz matrix. The DyFa program can be downloaded from: <http://faculty.psy.ohio-state.edu/browne/software.php>
- 2) Based on Expectation-Maximization of the raw data likelihood associated with state-space models (1b), where the latent factor series is estimated by means of the recursive Kalman filter (Expectation Step) and the parameters are estimated by means of multivariate regression (Maximization Step). See Hamaker, Dolan, and Molenaar (2005); Hamaker, Nesselrode, and Molenaar (2007) for applications; the software used in these applications can be obtained from: <http://users.fmg.uva.nl/cdolan/>
- 3) In the frequency domain, after discrete Fourier transformation. This yields a set of frequency-dependent complex-valued factor models that can be subjected to standard ML factor analysis (Molenaar, 1987). The software for dynamic factor analysis in the frequency domain can be obtained from the first author. Special nonlinearly constrained variants of this approach have been developed for the purpose of neural source estimation in brain imaging (Grasman, Huizenga, Waldorp, Böcker, & Molenaar, 2005). The software concerned can be obtained from: <http://users.fmg.uva.nl/rgrasman/>
- 4) Rewriting the model as a state-space model, with extended state containing not only the latent factor series but also the unknown parameters. This results in a nonlinear state-space model for which the extended Kalman filter is used to estimate the extended state, including the parameters. This will be discussed further in a later section.

### ***Application in Replicated Time Series Design***

Dynamic factor analysis constitutes a generalization of Cattell's P-technique (Cattell, Cattell, & Rhymer, 1947). P-technique involves application of the standard factor



model to the zero-lag covariance matrix  $\Sigma(0)$  of an observed multivariate time series (cf. Jones & Nesselrode, 1990), whereas dynamic factor modeling involves analysis of the observed series' complete sequential covariance function  $\Sigma(u)$ ,  $u = 0, \pm 1, \dots$ . Excellent recent discussions and illustrations of dynamic factor analysis of psychological time series can be found in Browne and Nesselrode (2005); Browne and Zhang (2007); Ferrer and Nesselrode (2003); Ferrer (2006); Hamaker, Dolan, and Molenaar, (2005); Kim, Zhu, Chang, Bentler, and Ernst (2007); Mumma (2004); Nesselrode, McArdle, Aggen, and Meyers (2002); Sbarra and Ferrer (2006); Shiftren, Hooker, Wood, and Nesselrode (1997); and Wood and Brown (1994). Dynamic factor modeling is increasingly prominent in econometrics, see for instance Forni, Hallin, Lippi, and Reichlin (2005); Stock and Watson (2005).

In this section an innovative application of dynamic factor analysis will be presented using multivariate time series data of multiple subjects and performing new types of statistical tests to uncover nomothetic relationships underlying idiographic observations. It concerns a multi-subject dynamic factor analysis of mood change during pregnancy data is presented, using the block-Toeplitz approach introduced in Molenaar (1985). A complete description of the data is given in Lebo and Nesselrode (1978). Here only part of the data will be analyzed (we thank Dr. Michael Lebo for making the data available). Only the data of three of the five subjects will be analyzed. Moreover, only five items will be selected: Enthusiastic, Energetic, Active, Peppy, and Lively. Keeping the same numbering of subjects as in Lebo and Nesselrode (1978), the length of this 5-variate observed time series (daily measurements) is  $T=112$  for Subject 1,  $T=110$  for Subject 3, and  $T=103$  for Subject 5.

State space models (1b) with 2-variate latent factor series are fitted to the data of the three subjects, using the block-Toeplitz correlation matrix approach with window width of 2 (cf. Molenaar, 1985). A window width of 2 implies that the covariance function of the latent factor series can be estimated up to lag  $\pm 1$ :  $\text{cov}[\eta(t), \eta(t-u)'] = \Psi(u)$ ,  $u = 0, \pm 1$ . No equality constraints across subjects are imposed on the (5,2)-D matrices of factor loadings. The subject-specific patterns of free and fixed factor loadings in  $\Lambda(0)$  were determined in preliminary confirmatory oblique P-technique analyses. The measurement errors are assumed to lack sequential dependencies. No equality constraints across subjects are imposed on the zero lag covariance matrices  $\text{cov}[\epsilon(t), \epsilon(t)'] = \text{diag-}\Theta(0)$  of the measurement errors. In contrast, it is assumed that the sequential correlation function of the bivariate latent factor series,  $\text{cor}[\eta(t), \eta(t-u)'] = \Psi(u)$ ,  $u = 0, \pm 1$ , is invariant across the three subjects.

Notice that the state-space models specified above are partly subject-specific and partly invariant across subjects. No equality constraints across subjects have been imposed on the factor loadings and measurement error variances, hence  $\Lambda(0)$  and  $\text{diag-}\Theta(0)$  are subject-specific. But the sequential correlation function of the latent factor series  $\Psi(u)$ ,  $u = 0, \pm 1$ , is constrained to be invariant across the three subjects. This pattern of subject-specific factor loadings and measurement error variances in combination with invariant sequential correlation function of the latent factor series implements the new definition of measurement equivalence proposed in Nesselrode, Gerstorf, Hardy, and Ram (2007). Whereas traditional definitions of measurement equivalence require that at least the matrices of factor loadings are



invariant across subjects, it is argued by Nesselroade et al. (2007) that factor loadings should be allowed to be subject-specific while correlations among factor series should be invariant across subjects.

The overall fit of the state-space models to the replicated 5-variate time series is excellent: chi-square = 114.40, degrees of freedom = 129, prob. = 0.82; non-normed fit index = 1.0; comparative fit index = 1.0. Hence the assessments of the three subjects can be considered measurement equivalent in the sense of Nesselroade et al. (2007). For Subject 1  $\eta_1(t)$  has significant positive factor loadings on Enthusiastic, Energetic, Active, and Lively, whereas  $\eta_1(t)$  has zero loading on Peppy. For this subject  $\eta_2(t)$  has unit loading on Peppy (no measurement error) and significant loading of 0.30 on Energetic. For Subject 2  $\eta_1(t)$  has significant positive factor loadings on Enthusiastic, Energetic, Active and Lively, whereas  $\eta_1(t)$  has zero loading on Peppy. For this subject  $\eta_2(t)$  has unit loading on Peppy (no measurement error) and significant loadings of 0.26 on Enthusiastic and 0.24 on Lively. Finally, for Subject 3  $\eta_1(t)$  has significant positive factor loadings on Energetic, Active, Peppy and Lively, whereas  $\eta_1(t)$  has zero loading on Enthusiastic. For this subject  $\eta_2(t)$  has unit loading on Enthusiastic (no measurement error) and a significant loading of 0.17 on Lively.

Given that  $\eta_1(t)$  has significant positive factor loadings on four of the five items for each subject and following Lebo and Nesselroade (1978), this factor series can be interpreted as Energy. Interpretation based on the pattern of factor loadings associated with  $\eta_2(t)$  also is unambiguous for Subjects 1 and 3—this factor series has unit loading on Peppy in combination with zero measurement error variance. Hence  $\eta_2(t)$  for Subjects 1 and 3 could be interpreted as Peppy. But for Subject 5  $\eta_2(t)$  has unit loading on Enthusiastic in combination with zero measurement error variance. According to this pattern of factor loadings,  $\eta_2(t)$  for Subject 5 should be interpreted as Enthusiastic. It therefore is problematic to assign an interpretation to  $\eta_2(t)$  which is invariant across the three subjects, at least if this interpretation is based on the patterns of subject-specific factor loadings.

An alternative way of assigning interpretations to  $\eta_1(t)$  and  $\eta_2(t)$  follows from a suggestion which is more in line with the basic tenet of the new definition of measurement equivalence of Nesselroade et al. (2007). Instead of considering patterns of factors loadings, interpretations of  $\eta_1(t)$  and  $\eta_2(t)$  are based on the sequential correlation function of the latent factor series. Nesselroade (2007, p. 258) also suggests this alternative way of interpreting factor series, in that "... invariance might be sought at the process level by focusing on patterns of auto- and cross-regression of latent factors, for example, in individual level dynamic factor models". The estimated sequential correlation function  $\Psi(u)$ ,  $u = 0, \pm 1$  of the latent factor series  $\boldsymbol{\eta}(t)$ , assumed to be invariant across the three subjects, has significant zero lag correlation:  $\text{est.-cor}[\eta_1(t), \eta_2(t)] = 0.69$ . Also the autocorrelation of  $\eta_1(t)$  at lag  $\pm 1$  is significant:  $\text{est.-cor}[\eta_1(t), \eta_2(t \pm 1)] = 0.19$ . The remaining elements of  $\text{est.-}\Psi(\pm 1)$  are not significant. Consequently,  $\eta_1(t)$  has a certain degree of stability (i.e., significant autocorrelation) and can be interpreted as Stable Energy, whereas  $\eta_2(t)$  resembles a white noise sequence and can be interpreted as Unstable Energy.

In conclusion of this section, it has been shown that using the multi-group option in commercially available structural equation modeling software, in combination with the block-Toeplitz matrix approach introduced in Molenaar (1985), it is possible to test for equivalences across different subjects using data obtained in a replicated time series design. In this approach different subjects constitute different “groups”. In this way it is possible to detect nomothetic relationships based on idiographic data, or stated otherwise, to detect inter-individual relationships based on intra-individual variation. This approach has been illustrated by implementing the new definition of measurement equivalence proposed by Nesselroade et al. (2007). But it can be applied in many other situations, including testing for measurement equivalence according to the traditional definitions.

## Nonstationary Dynamic Factor Analysis

To reiterate, the two criteria for ergodicity are stationarity in time and homogeneity across subjects. In the previous application we focused on testing for homogeneity of the sequential correlations function across three subjects. Now we turn to a consideration of stationarity; the other criterion for ergodicity. A new approach is presented to test for stationarity and model nonstationary processes. The new approach is based on a state-space model with arbitrarily time-varying (nonstationary) parameters. The model concerned is:

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{\Lambda}[\boldsymbol{\theta}(t)]\boldsymbol{\eta}(t) + \mathbf{v}(t) \\ \boldsymbol{\eta}(t+1) &= \mathbf{B}[\boldsymbol{\theta}(t)]\boldsymbol{\eta}(t) + \boldsymbol{\zeta}(t+1) \\ \boldsymbol{\theta}(t+1) &= \boldsymbol{\theta}(t) + \boldsymbol{\xi}(t+1) \end{aligned} \quad (2a)$$

In (2a)  $\mathbf{y}(t)$  denotes the observed  $p$ -variate time series;  $\boldsymbol{\eta}(t)$  the  $q$ -variate latent factor series (state process). The first equation of (2a) shows that factor loadings in  $\mathbf{\Lambda}[\boldsymbol{\theta}(t)]$  depend upon a time-varying parameter-vector  $\boldsymbol{\theta}(t)$ . The second equation describes the time evolution of the state process  $\boldsymbol{\eta}(t)$ ; the autoregressive weights in  $\mathbf{B}[\boldsymbol{\theta}(t)]$  depend upon  $\boldsymbol{\theta}(t)$  and therefore can be arbitrarily time-varying. The third equation in (2a) describes the time-dependent variation of the unknown parameters. The  $r$ -variate parameter vector process  $\boldsymbol{\theta}(t)$  obeys a random walk with Gaussian white noise innovations  $\boldsymbol{\xi}(t)$ . The covariance functions associated with (2a) are given in (2b):

$$\begin{aligned} \text{cov}[\mathbf{v}(t), \mathbf{v}(t-u)'] &= \delta(u)\text{diag-}\boldsymbol{\Xi} \\ \text{cov}[\boldsymbol{\zeta}(t), \boldsymbol{\zeta}(t-u)'] &= \delta(u)\text{diag-}\boldsymbol{\Psi} \\ \text{cov}[\boldsymbol{\xi}(t), \boldsymbol{\xi}(t-u)'] &= \delta(u)\text{diag-}\boldsymbol{\Phi} \end{aligned} \quad (2b)$$

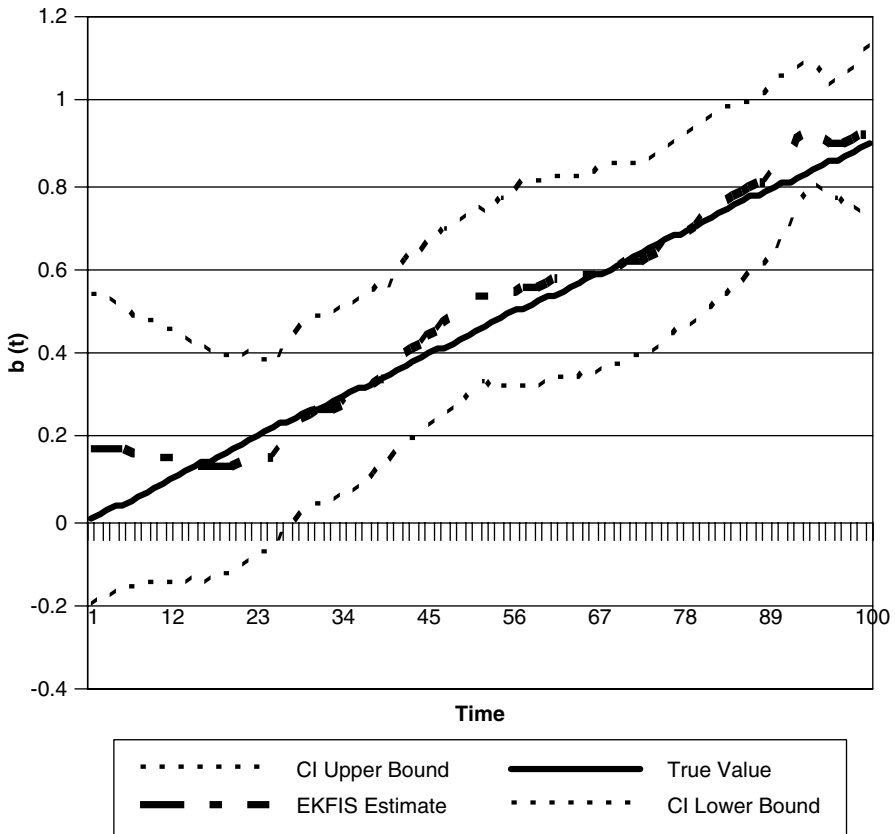
To fit the state-space model with time-varying parameters to an observed multivariate time series, use is made of a combination of the second and fourth estimation techniques mentioned at the end of section “Some intricacies associated with (1)”. That is, a combination of the EM algorithm and the extended Kalman filter/

smoother. First, an extended state process is defined. The extended state process consists of the original latent factor series and the time-varying parameter process:  $\mathbf{x}(t)' = [\boldsymbol{\eta}(t)', \boldsymbol{\theta}(t)']$ . Then, using the extended state process  $\mathbf{x}(t)$ , (2a) is rewritten as the following nonlinear state-space model:

$$\begin{aligned} y(t) &= \mathbf{h}[\mathbf{x}(t), t] + \mathbf{v}(t) \\ \mathbf{x}(t + 1) &= \mathbf{f}[\mathbf{x}(t), t] + \mathbf{w}(t) \end{aligned} \tag{3}$$

The vector-valued nonlinear functions  $\mathbf{h}[\mathbf{x}(t), t]$  and  $\mathbf{f}[\mathbf{x}(t), t]$  consist of products of the entries of  $\mathbf{x}(t)$ . The  $(q+r)$ -dimensional innovations process  $\mathbf{w}(t)$  is defined as the composition of the innovation processes  $\boldsymbol{\zeta}(t)$  and  $\boldsymbol{\xi}(t) : \mathbf{w}(t)' = [\boldsymbol{\zeta}(t)', \boldsymbol{\xi}(t)']$ .

In Fig. 12.1 an illustrative result is shown of an application of this new technique to a simulated time series. A 4-variate time series has been generated by means of the state-space model with time-varying parameters. The model has a univariate latent state process ( $q=1$ ). The autoregressive coefficient  $\mathbf{B}[\boldsymbol{\theta}(t)] = b(t)$  in the true process model for the latent state [second equation in (2a)] increases linearly from 0.0 to 0.9 over the observation interval comprising  $T=100$  time points. Depicted is the esti-



**Fig. 12.1** EKFIS estimate of time-varying coefficient  $b(t)$  in the autoregressive model for the latent factor scores

mate of this autoregressive coefficient  $b(t)$  obtained by means of (3). It is clear that the estimated trajectory closely tracks the true time-varying path of this parameter.

## Discussion and Conclusion

The future of dynamic factor analysis is challenging because of the necessity to focus on the structure on intra-individual variation in the study of nonergodic psychological processes. This necessity follows directly from the classical ergodic theorems. In case subjects are heterogeneous with respect to a particular psychological process, that is, in case person-specific dynamics describe the intra-individual variation of this process, one can only obtain valid information about such non-ergodic process by means of dedicated time series analysis. In a similar vein, in case a psychological process is nonstationary, like developmental and learning processes, but also many clinical and biomedical processes, one also can only obtain valid information about such a nonergodic process by means of dedicated time series analysis.

The implications of the classical ergodic theorems have a very broad impact, not only in psychometrics and psychology but also in the biological and medical sciences. They imply that growth processes as well as disease processes have to be analyzed in organism-specific and patient-specific ways, focusing on the intra-individual variation concerned. We presently are involved in a series of pilot studies using stochastic person-specific control of day-today intra-individual variation in disease processes such as asthma, diabetes, and daily stress. In these projects dynamic factor analysis of non-stationary time series is used to track momentary changes in a disease process as function of medication dose, environmental, emotional and contingent stressors. Then, using predictive control methods based on the fitted time-varying state-space patient models, optimal medication dose is determined at each point in time in a patient-specific way. It is known that the effects of medication, in particular for diabetes and asthma, are patient-specific. Dynamic factor analysis of nonstationary multivariate time series is excellently equipped to accommodate substantial patient-specific reactions to medication and counteract the occurrence of contingent disturbances occurring under normal daily life circumstances.

## References

- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
- Brillinger, D. R. (1975). *Time series: Data analysis and theory*. New York: Holt, Rinehart & Winston.
- Browne, M. W., & Nesselrode, J. R. (2005). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Psychometrics: A festschrift to Roderick P. McDonald* (pp. 415–452). Mahwah, NJ: Erlbaum.
- Browne, M. W., & Zhang, G. (2005). User's Guide: DyFa: Dynamic factor analysis of lagged correlation matrices, version 2.03. [www document and computer program], from <http://quanrm2.psy.ohio-state.edu/browne/>.

- Browne, M. W., & Zhang, G. (2007). Developments in the factor analysis of individual time series. In R. C. MacCallum & R. Cudeck (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahwah, NJ: Erlbaum.
- Cattell, R. B. (1952). The three basic factor-analytic designs—Their interrelations and derivatives. *Psychological Bulletin*, *49*, 499–520.
- Cattell, R. B., Cattell, A. K. S., & Rhymer, R. D. (1947). P-technique demonstrated in determining psychophysiological source traits in a normal individual. *Psychometrika*, *12*, 267–288.
- Choe, G. H. (2005). *Computational ergodic theory*. Berlin: Springer-Verlag.
- De Groot, A. D. (1954). Scientific personality diagnosis. *Acta Psychologica*, *10*, 220–241.
- Ferrer, E. (2006). Application of dynamic factor analysis to affective processes in dyads. In A. Ong & M. van Dulmen (Eds.), *Handbook of methods in positive psychology* (pp. 41–58). Oxford: Oxford University Press.
- Ferrer, E., & Nesselroade, J. R. (2003). Modeling affective processes in dyadic relations via dynamic factor analysis. *Emotion*, *3*, 344–360.
- Forni, M., Hallin, M., Lippi, F., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*, 830–840.
- Grasman, R. P. P. P., Huijzena, H. M., Waldorp, L. J., Böcker, K. B. E., & Molenaar, P. C. M. (2005). Stochastic maximum likelihood mean and cross-spectrum modeling in neuron-magnetic source estimation. *Digital Signal Processing*, *15*, 56–72.
- Hamaker, E. J., Dolan, C. V., & Molenaar, P. C. M. (2005). Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis. *Multivariate Behavioral Research*, *40*, 207–233.
- Hamaker, E. J., Nesselroade, J. R., & Molenaar, P. C. M. (2007). The integrated state-space model. *Journal of Research in Personality*, *41*, 295–315.
- Hannan, E. J. (1970). *Multiple time series*. New York: Wiley.
- Jones, C. J., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject, repeated measures design and P-technique factor analysis: A review of intra-individual change studies. *Experimental Aging Research*, *4*, 171–183.
- Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, *28*, 85–93.
- Lebo, A. M., & Nesselroade, J. R. (1978). Intraindividual differences dimensions of mood change during pregnancy identified in five P-technique factor analyses. *Journal of Research in Personality*, *12*, 205–224.
- Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.
- Molenaar, P. C. M. (1987). Dynamic factor analysis in the frequency domain: Causal modeling of multivariate psychophysiological time series. *Multivariate Behavioral Research*, *22*, 329–353.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218.
- Molenaar, P. C. M., Huijzena, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of interindividual and intraindividual variability: A theoretical and empirical vindication of developmental systems theory. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development: Dialogues with lifespan psychology* (pp. 339–360). Dordrecht: Kluwer Academic Publishers.
- Molenaar, P. C. M., & Nesselroade, J. R. (2001). Rotation in the dynamic factor modeling of multivariate stationary time series. *Psychometrika*, *66*, 99–107.
- Molenaar, P. C. M., & Nesselroade, J. R. (2008). The recoverability of P-technique factor analysis (to appear in *Multivariate Behavioral Research*).
- Mumma, G. H. (2004). Validation of idiosyncratic cognitive schema in cognitive case formulations: An intraindividual idiographic approach. *Psychological Assessment*, *16*, 211–230.
- Nesselroade, J. R. (2007). Factoring at the individual level: Some matters for the second century of factor analysis. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 249–264). Mahwah, NJ: Erlbaum.

- Nesselroade, J. R., Gerstorf, D., Hardy, S. A., & Ram, N. (2007). Idiographic filters for psychological constructs. *Measurement, 5*, 217–235.
- Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Dynamic factor analysis models for representing process in multivariate time series. In D. S. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Methods and applications* (pp. 235–265). Mahwah, NJ: Erlbaum.
- Sbarra, D. A., & Ferrer, E. (2006). The structure and process of emotional experience following nonmarital relationship dissolution: Dynamic factor analysis of love, anger, and sadness. *Emotion, 6*, 224–238.
- Shifren, K., Hooker, K., Wood, P., & Nesselroade, J. R. (1997). Structure and variation of mood in individuals with Parkinson's disease: A dynamic factor analysis. *Psychology and Aging, 12*, 328–339.
- Stock, J. H., & Watson, M. W. (2005). *Implications of dynamic factor models for VAR analysis*. Mimeo: Princeton University.
- Wood, P., & Brwon, D. (1994). The study of intra-individual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin, 116*, 166–186.