Dynamic Factor Analysis: Modeling Person-Specific Process

Ram, Annette Bosse, and Peter C. M. Molenberghs

Abstract

Modern data collection technologies are providing large data sets, with many repeated observations of many individuals on many variables—and new opportunities for application of analytical techniques that consider individuals as unique, complex, multivariate, dynamic entities. In this chapter we review the conceptual and technical background for dynamic factor analysis and provide a primer for application to multivariate time-series data. Step-by-step procedures are illustrated using daily diary data obtained from three women over 102+ days. Specifically, we provide background on and demonstrate (1) formulation of DFA research questions; (2) study design and data collection; (3) variable selection and data pre-processing procedures; (4) the fitting and evaluation of person-specific DFA models; and (5) examination of between-person differences/similarities. We conclude by noting to some extensions that might be elaborated and used to articulate additional complexities of within-person process.

Keywords: longitudinal, P-technique, dynamic systems, idiographic, ecological momentary assessment

A number of "dynamic" longitudinal models are being adapted and used to more clearly understand biological, psychological, and behavioral processes. Generally, the aim of these models is to articulate and test hypotheses about how an established series of events or actions transform an entity from one state to another. Specifically, the objective is to model how an individual's state at one point in time is influenced by his or her past states and/or influences his or her future states.

Recent advances in mobile computing technology have opened new possibilities to obtain biobehavioral data, model it in real-time, and remotely deploy interventions at population scale. The electronic devices many of us now carry with us as we go about our daily lives provide a wide array of opportunities to collect more and more data from more and more participants and to deliver time- and context-specific guidance to them. Such data streams have tremendous implications for how biobehavioral phenomena can be approached, both in principle and in practice. As new study designs bring moment-to-moment data "online" it shall be possible to track, model, and guide the progression of behavioral transformations—in real-time and in real-time, as individuals go about their daily lives.

In this chapter, we provide an overview of one approach for modeling within-person processes—dynamic factor analysis (Molenberghs, 1985)—that holds substantial utility for application to these emerging data streams.

The speed and capacity of modern computers brings with it new possibilities for computation. As has been shown with the advent of Internet
search engine data mining, we can now estimate the parameters of thousands of models to a given set of data in less than 1 second. This means that it is now feasible to implement person-specific approaches to data analysis. Rather than presupposing, "top-down," that all individuals fall into a single population described by an "average" process, we can instead take a "bottom-up" approach and model individual biobehavioral functioning, one person at a time (Carroll, 1966; Lamiell, 1981; Molenar, 2004; Nesselroade, 2007; Ram & Geerson, 2009; Steen, 1911; Valois, 1985).

This provides an opportunity to test basic assumptions of homogeneity and equivalence of within- and between-person structures that are known to be problematic (i.e., ergodicity of dynamic processes; Molenar, 2004), eliminates the need to interpret sample-level findings as though they apply to within-person processes (i.e., commit an ecological fallacy; Etes, 1956; Robinson, 1950) and sets the stage for personalized intervention at population scale.

In the following sections, we review the conceptual and technical background for person-specific dynamic models and provide a primer for application of dynamic factor analysis to multivariate time series data. Step-by-step procedures are illustrated using "classic" daily diary data obtained from a small number of women over 100- plus days (The Lebo Data: Lebo & Nesselroade, 1978). Finally, we point to some extensions of the dynamic factor model that may be used to anticipate even more complex aspects of within-person change.

Background

Factor analysis is a method for investigating the structure of a set of variables. The basic principle is to represent the covariation among many observed variables in terms of linear relations among a smaller number of abstract or latent variables. The underlying idea is that if two or more characteristics covary in a systematic manner, they may reflect a shared underlying construct. In practice, the patterns of covariation reflect the latent dimensions that lie beneath the observed variables (Cattell, 1965; Tennen & Ritzes, 1989). P-technique. P-technique factor analysis is the application of factor analysis to P-data—a matrix of variable (x multi-variable (x single person) matrix of scores (Carroll, Carroll, & Rhymer, 1947). Applied to this multivariate time series data, the P-technique factor model provides a parsimonious description of inter-individual variation and covariance. As such, the P-technique model provides a framework for examining inter-individual differences in the latent dimensions that lie beneath repeated measures for an individual. The modeling approach has been used in numerous areas to describe individual-level differences among factors, personality, psychophysiology, and other domains (see Jones & Nesselroade, 1999; Luborsky & Mira, 1972; and Russell, Jones, & Miller, 2007, for reviews). For example, portraying our forthcoming illustrations, consider the data of six individuals who obtain from one person over 100-plus days that are plotted on the upper panel of Figure 21.1. The objective of a P-technique analysis is to parsimoniously describe the relationships among this woman's daily self-reports of cheerful, happy, contented, vigorous, tired, and well-being.

The objective of the analysis is to test a hypothesis that the day-to-day variation in the observed variables can be adequately described by a manifest factor of two latent, unobserved factors: positive well-being and fatigue. As will be described in detail below, the common factor model is applied to time-series data to obtain a more parsimonious description of the processes that contribute to an individual's day-to-day experiences. This latent process representation is plotted in the lower panel of Figure 21.1.

Time series. P-data, like those shown in the upper panel of Figure 21.1, consist of observations obtained from the same entity on multiple occasions. Given that organisms maintain some sort of continuity over time, repeated measurements obtained from the same person are likely to be related. Thus, time series data likely violate a key assumption required by many statistical analyses—specifically, that observations are independent and identically distributed. Early critiques of P-technique factor analysis pointed out that the model ignored the time dependency in the data (e.g., Anderson, 1963; Carrell, 1967; Tennen & Ritzes, 1989). In the decades that followed, time series analysis emerged as a way to explicitly model and accommodate dependence in such data (e.g., Box & Jenkins, 1976; Jenkins & Watts, 1968). A plethora of techniques are now available for dealing with and making use of the time ordering, sequences, and dependencies inherent in time series data (Shumway & Stoffer, 2006). Of particular importance for our purposes here was the advent of autoregression (and moving average) models, wherein relations among successive occasions are modeled as autoregressive and cross-regression.

Dynamic factor analysis. Molenar (1985) introduced dynamic factor analysis (DFA) as a combination of P-technique factor analysis and time series analysis. The objective was to both deal with the independence violations and provide a framework for modeling the dynamic nature of ongoing processes. In brief, the underlying notion of the DFA model is that the (multivariate) state of the individual at any given time is a function of both concurrent influences and past states. Following our example, an individual's present level of fatigue may, in part, be influenced by what happened yesterday. That fatigue may linger or carryover from one day to the next. Similarly, regulatory processes may promote the maintenance of well-being from one day to the next. Events that influence well-being may contribute not only to current levels but also carry forward for some limited amount of time. The DFA framework provides an opportunity to explicitly model such processes.

Often articulated as state-space models, many fields make use of DFA-type frameworks (Dunbar & Koopman, 2001). In fact, much of the machinery that takes us from place to place (e.g., planes, trains, automobiles) depends on such frameworks to model, forecast, and help guide movements in real-time. Subtractive applications in psychology include modeling of affective and psychophysiological changes (Chow et al., 2004; Ferrer & Nesselroade, 2003; Gates et al., 2010; Wood & Brown, 1994), where ongoing processes (e.g., adaptation, regulation, homeostasis) can be extracted from time series data collected on relatively short time-scales.

Person-specific approach. Person-specific approaches seek to individualize the dynamics of the adaptive, regulatory, and other processes that proceed at the individual level (Nesselroade, 2007; Ram & Geerson, 2009). The objective is to extract a viable representation of an ongoing process from the covariation that manifests in multiple observations of a person across time (Nesselroade, 1991). The focus is on describing, explaining, predicting, and potentially modifying individual behavior, not sample- or population-level behavior. Knowledge about how variables relate across individuals at a single time-point (between-person covariation) cannot be used to make inferences about any individuals' actual behavior (Estes, 1956; Robinson, 1950; see also recent discussions in Steen & Bauer, 2010, and associated comments).

Recently, Molenar (2004) underscored this point using mathematical proofs. Outlining the relevance of ergodicity theorems, he demonstrated that
within- and between-person structures are equivalent (i.e., ergodic) only under very strict (and likely rare) conditions—namely, (1) stationarity of variables' autoregressive and (2) equivalence of the relations among variables for all individuals. Following this logic, it is simply not possible to theorize about between-person differences (the hallmark of most of our inquiries) and meet the methodological requirement that the patterns of variability across time are identical for all individuals. In sum, when using between-person analytic techniques to examine psychological phenomena, we are very likely forced to commit egregious ecological fallacies that are, by definition, at odds with the very phenomenon (i.e., processes) we want to examine. As such, it seems imperative that researchers make use of person-specific analysis frameworks in the formulation and testing of psychological theory (Hamaker, Dolan, & Molenaar, 2005).

Interindividual differences in the individual-level processes can be studied in a subsequent step. After providing some additional mathematical background, we shall, in the sections that follow, work through a step-by-step illustration of how dynamic factor models can be implemented with sufficiently long data streams.

Technical Background

P-technique factor analysis is procedurally similar to the familiar between-person (R-technique) factor analysis to which most researchers are exposed as part of their graduate research methods training. What differs are the data to which the models are applied. In the usual R-technique factor analysis, the common factor model is applied to multivariate observations obtained from multiple subjects at a single measurement occasion (a person × variables matrix of scores). In contrast, in P-technique factor analysis, the common factor model is applied to multivariate single subject time series data (an occasions × variables matrix of scores). The model can be written as

\[ y(t) = \Lambda \eta(t) + \epsilon(t), \]

where \( y(t) \) is a \( p \)-variate time series of observations indexed by time \( t \) (e.g., \( \{1, 2, \ldots, T\} \)), \( \Lambda \) is a \( p \times k \) loading matrix, \( \eta(t) \) is a \( k \)-variate time series of latent factor scores, and \( \epsilon(t) \) is a \( p \)-variate residual (specific error + measurement error) time series. An example model is depicted graphically in Figure 21.2a. The path model depicts how a six-variate \( y(t) \) time series (squares labeled 1 to 6) is "driven" by two common factor score series (circles labeled \( \eta \) and \( \eta_n \)).

In P-technique factor analysis, the common factor model, \( y(t) = \Lambda \eta(t) + \epsilon(t) \), is used to model data obtained from one individual over many occasions, \( t = 1 \) to \( T \), under the assumption that the observations are independent. Depicted graphically in Figure 21.2a, there are no sequential dependencies (arrows) between the variables (latent and manifest) at occasion \( t-1 \) and those at \( t \). The labels for the two occasions could be swapped, \( t \) and \( t-1 \), without effect on the model fit or model parameters. Given organismic continuity, this is an unlikely circumstance. Rarely would we find that repeated measures obtained from the same organism are truly independent observations in the sense that there is no relation between the states on different occasions (see Fiske & Fiske, 1955, and Ram & Geerse, 2009, for discussions of net intra-individual variability).

Dynamic factor analysis (Molenaar, 1985) relaxes the assumption that all observations are independent observations of an individual's states. The occasion-to-occasion dependencies of a time series with equally spaced observations are modeled explicitly (addressing some of the early critiques of P-technique; e.g., Anderson, 1965) and allowing for carryover, spillover, or systemic memory from one occasion to the next. A few configurations of the model have been presented and used (see, e.g., Neselroade, McAdie, Agresti, & Meyers, 2002). In a simplifying, the dynamic factor model can be written as

\[ y(t) = \Lambda \eta(t) + \epsilon(t) \]

where \( \Lambda \) is a \( p \times k \) matrix and \( \eta(t) \) is a \( k \)-variate latent state series that is now modeled as a function of \( k \) lagged 

\[ \eta(t-1) + \eta(t-2) \]

where the \( q \)-variate latent state series \( \eta(t) \) is now modeled as a function of \( k \) lagged states, \( \eta(t-1) + \eta(t-2) \), that are weighted by \( \Lambda \) to \( \eta \). Present time "disturbances" are then introduced as a \( k \)-variate set of latent "innovations," \( \xi(t) \), and residual (measurement + specific error) time series. An example model is depicted graphically in Figure 21.2b. The path model depicts how a six-variate \( y(t) \) time series (squares labeled 1 to 6) is "driven" by two common factor score series (circles labeled \( \eta \) and \( \eta_n \)).

The measurement error terms can be incorporated at the measurement level through between-occasion correlations, not shown in Figure 21.2b. Substituting Equation 3 into Equation 2 and with a little rearranging, we obtain

\[ y(t) = \Lambda \eta(t) + \epsilon(t) \]

where the time dependencies are now modeled through a set of lagged factor loadings, \( \Lambda_k, k = 0, 1, 2, \ldots \). We note that there are a number of nuanced differences between DFA with different configurations and the nature of the processes captured or implied by each model (see, e.g., Brown & Neselroade, 2005; Browne & Zhang, 2007; Neselroade et al., 2002). In particular, Equation 4 is a special case of, and can be rotated to, Equation 5 (Molenaar & Neselroade, 2001). Selection of one configuration over another should, therefore, be informed by substantive considerations—how the specific parameters of the model can be mapped to the particular process of interest. In our forthcoming example discussion, we use the (state-space) representation given in Equations 4 (with additional practical, but unnecessary, constraint that the measurement errors are uncorrelated over time).

Investigations of individual-level processes require consideration of the time-structured variability (Ram & Geerse, 2009). The theories presume transactions or activities that connect an individual's prior state to his or her present and future states—behavioral transformations that are contiguous. The data in which those processes manifest are, by definition, not independent and require explicit rendering of the sequential dependencies. Dynamic factor analysis offers a robust framework for modeling process-oriented theory in time series data (Molenaar, 2010). In recent years, the methodological literature surrounding DFA and availability of software programs for model estimation has expanded (see Webpages for C. Dolan, L. L. Lo, G. Zhang, and Z. Zhang, among others). The literature illustrates use of maximum likelihood, ordinary least squares, Kalman filter, and Bayesian approaches (Zhang, Hamaker, & Neselroade, 2008), use of the model for explorations of reliability of change (Lane & Shroot, 2010), and implementations as structural equation models (Chow et al., 2010). In sum, the availability of software tools and computational power now afford the possibility to consider person-specific DFA with relative ease and speed.
Five Steps for Conducting Dynamic Factor Analysis

The following sections describe a five-step heuristic for conducting DFA. We illustrate implementation using data obtained from a small sample of pregnant women who provided ratings of their daily mood on 100-plus consecutive days (Lebo & Nesselroade, 1978; see also Brosi & Ram, 2012; Molenar & Nesselroade, 2001; Nesselroade & Molenar, 1999; Nesselroade et al., 2007; Nesselroade, McAdile, Aggen, & Meyers, 2002; Zhang & Browne, 2010). We purposely note that there is some irony in our use of this "classic" data set, given that we believe the applications of DFA will expand as modern technologies provide for more and more intensive data collection. However, there is an interesting contrast between the explosion of experience sampling/ behavioral momentary assessment studies and the still low availability of relatively lengthy psychologically oriented time series data suitable for DFA—particularly data with equal spacing between assessments.

Step 1: Empirical Illustration

As we all know from personal experience and observations of others (as well as a plethora of empirical research), individuals’ affective states fluctuate from day to day in response to endogenous and exogenous events. There is also evidence of regulatory processes that mediate these responses. Our specific interest here was in articulating two processes: a stability maintenance process that promotes persistence of an individual’s positive well-being from one day to the next and a buffering process wherein increases in well-being contribute to decreases in feelings of fatigue the following day. To operationalize these questions in available data, we made use of a measurement model wherein daily day-to-day changes in well-being and fatigue would manifest in daily reports of affective feeling states and a dynamic model capturing dependencies between consecutive day’s well-being and fatigue. Specifically, we sought to confirm that for each individual in the study, there was systematic day-to-day spillover and buffering. Further, we were interested in between-person differences in these stability maintenance and buffering processes. Considering that the endogenous and exogenous variables driving the structure and dynamics of affective experiences differ from one individual to the next (e.g., people live in different contexts), we expected substantial between-person differences in how these processes would manifest in different women. We explored whether individuals in the study would be characterized by similar or different processes.

Step 2: Study Design and Data Collection

The data requirements for dynamic factor analysis are that each individual must be measured on multiple variables repeatedly on many occasions. The resulting time series (P-data) must be (1) of considerable length, (2) collected on a time-scale that matches the phenomena of interest, and (3) sampled at equally spaced intervals. Although there are no clear rules, it has been recommended that factor analytic studies use no less than 100 observations in any analysis (Gorsuch, 1983, p. 332), and that there be at least five observations per parameter being estimated in the model (Loehlin, 1998). Time series of length 100 observations per individual can be considered as a kind of minimum starting point for person-specific dynamic analysis (cf. Wood & Brown, 1994, recommendations for 300+). Further, it is essential that the study design capture variability—the core of any statistical analysis—at the individual level. Key concerns are the time-scale on which the observations were obtained and that the interval between successive occasions is long enough that change can occur but not so long an interval that the progression of the process is entirely missed (Beker, Molenar, & Nesselroade, 2009; Collins, 2005; Shiyko & Ram, 2011).

Similarly, measurement instruments should be sensitive enough to capture the variation produced by the processes of interest. Subtle occasion-to-occasion changes may be lost in the granularity of the response scale. For example, when occasion-to-occasion changes are not so large as to prompt individuals to move their response to the next higher or lower category, the granularity of the response scale can be inadvertently imposing a limit on what constitutes "meaningful" change. This is not to say that one should always strive to use interval scales. Dynamic factor analysis models can be implemented with interval, ordinal, and categorical variables (G. Zhang & Browne, 2010; Zhang & Nesselroade, 2007). The point is that care should be taken that the measurement instruments are well suited to capture the particular process of interest.

Step 3: Variable Selection and Data Preprocessing

Variable Selection. Once individuals’ time series of observations have been collected, it is important to determine whether the data are, in actuality, suitable for application of DFA. Most importantly, there must be reliably measured variation in scores on the specific variables to be analyzed (Comrey & Lee, 1992, p. 238). Variables with no within-person variance across time cannot, by definition, be subjected to analysis of variation and covariation. Various rules of thumb have been used to identify and remove variables that do not have sufficient variance for analysis. These include removing variables with (intra-individual) standard deviations below 0.10, or variables with more than 80% of scores being identical (see, e.g., Lebo & Nesselroade, 1978; Zevon & Tellegen, 1982). The issue becomes complicated when individuals exhibit insufficient variance on different items. Three routes can be taken. (1) Specific items can be excluded from each individual’s analysis, potentially resulting in a different set of variables being analyzed for each individual in the sample. The advantages of this route are that as much information as possible is maintained in the analyses, and idiosyncratic manifestations of the same phenomena can be acknowledged and modeled (Nesselroade et al., 2007). (2) Alternatively, individuals who have insufficient variability on one or more items are excluded from the analyses. The advantage of this approach is that between-person comparisons among the remaining sample are easy. (3) Strike a balance between finding a common set of items and a “common set” of persons by placing equal weight on the selective sampling of persons and selective sampling of items. The idea is to hone in on the subsample of individuals for whom a particular set of items is relevant. This acknowledges the possibility of qualitatively different measurement models (i.e., idiosyncratic interpretation of items), while preserving the benefits of across-person measurement invariant within each subsample of individuals.

Data Preprocessing. Before the main analysis, the data should be examined for suitability for application of the dynamic factor model. In principle, the main objective of the preprocessing is to obtain time series that are weakly stationary. As usual, there are no clear guidelines on what preprocessing steps are most appropriate. Depending on the specifics of the
research question, and how much "nonstationarity" was collected along with the phenomena of interest. Researchers may choose among many possibilities to identify and remove trends (or cycles) and other
anomalous features. This can be accomplished using linear, quadratic, or other polynomial
regressions (e.g., detrending), frequency analysis (i.e., spectral analysis), or through application of various
enabling techniques, filters, or smoothers (see, e.g., Shumway & Stoffer, 2000, for concise review). The
choice of preparations can consider both statistical
criteria for testing whether the resulting time-series are
stationary and the result of evaluation of what processes
should be identified and removed from (and modeled in) the data.

Statistical methods for identifying nonstationarity
include use of evolutionary spectra (Priesley, 1981), the
Augmented Dickey-Fuller Test (Dickey & Fuller, 1979), and fitting of multiple models to the
data. However, there is not always an obvious
way to use these tests or explore possible models for
nonstationarity (see Chartier, 1994, 128-
154). Theoretical considerations revolve around
the idea that each trend, cycle, or other "noise"
component is present in the data as driven by one or more processes (e.g., learning, circularity, measurement error). Modeling and
removing these elements in essence, a procedural
setting aside of particular processes to concentrate on
underlying structures that are independent of those
processes.

Step 3: Empirical Illustration

Data Preparing: Our next goal was to prepare the selected data from these three individuals so that
t hey met stationarity requirements. To this end, we
took the following steps. First, we standardized each
variable for each individual (A – X/SD = 1) to remove
eventual differences in overall variance and
event of the response scale. Second, each individual's
data was inspected visually.
The prepared six-variate time-series for Participants
1 are the data we used for Figure 21.1. By
observation, the item-level trajectories did not show trends
time, nor did we have reason to presume the
presence of systematic trends. Thus, we treated the
data as though they were weakly stationary and con-
ducted post hoc tests to check the viability of the final
models for different portions of the data (e.g., full
half vs. second half) as an approximate indicator or adherence to stationarity assumptions.

Step 4: Testing and Evaluating

Dynamic Factor Models

Dynamic factor analysis can be implemented in
a variety of ways. Methodologists, in recent years,
demonstrated and evaluated the possibilities for using structural equation modelling-based
maximum likelihood approaches (e.g., LISREL, Mplus, OpenMx, SAS PROC CALIS, etc.),
calculating factor approaches (mimic; Bayesian approaches (Win-
bugs), and ordinary least squares (OLS) approaches
(Dybs) as well as provided many resources regarding
the specifications of implementation (Browne &
Zhang, 2007; Chow, Hamaker, & Dolan
2010; Molenaar & Nesselroade, 1998; Nesselroade
et al., 2002; Wood & Brown, 1994; Zhang &
Brown, 2010; Zhang, Hamaker, & Nesselroade
2008). The model is specified in accordance with
temporal expectations (e.g., specific hypothesis about
structure and time dependences, including
the number of lags) and then fitted to the data
using Kalman filter, maximum likelihood, OLS,
Bayesian, or other procedures. Models are fit to each
individual's data separately. Fit statistics (e.g., X2
–2LL) and parameter estimates are obtained and
interpreted in relation to theory and other empirical
evidence. In the context of a person-specific analysis
approach, the fit criterion and solutions can be used
to determine which of the hypothesized structures provide an adequate and/or better description of the
individual's data. Between-person comparisons are
done in a subsequent step.

Step 4: Empirical Illustration

To examine the dynamic structure of well-being and fatigue for each of the three women, we fit con-
firmatory P-technique and DFA models separately
each of the three women's data using mplus (con-
cert C. V. Dolan). For each analysis, the P-data
being analyzed consisted of a 100-plus occasions
(days) X six variables matrix of scores. The measure-
ment portion of the model (Equation 2) was
specified with two factors: a well-being factor (η1) indicated by the items cheerful, happy, and cons-
tented (η = 32) and a fatigue factor (η2) indicated
by the items sluggish, tired, and weary (η = 36). As
in Figure 21.2, the two factors had simple structure
(no cross-loadings) and were allowed to correlate.
The dynamics portion of the model (Equation 3) was
specified in two ways. Specifically, the elements of
B were either (1) constrained to equal zero as in
a typical P-technique factor model assuming inde-
pendent observations, or (2) freely estimated as a
one-lag DFA model. The models were identified by
fixing the variance of the innovations equal to one.
Individual level results are provided in Table 21.1.

The evaluation and interpretation of the other
person-specific solutions proceeded one individual
at a time. With the P-technique model being tested
under the DFA model, we were able to test, using
likelihood ratio tests, whether the DFA model
provided a better statistical fit to the data than the
more parsimonious P-technique model. This test
was done separately for each individual. For
individual 1, the DFA model provided a better fit to the data than the more constrained P-technique
model (Δ = 2LL = 26.64, df = 4, p < 0.05),
meaning that there were indeed time-related
processes to be extracted. Auto-regression parameters
included both for the well-being (β = 0.42) and the fatigue
(β = 0.28) factors were significant, indicating carry-
over in feelings from one day to the next. The
non-significant cross-regression parameters provide
no evidence of buffering. For individual 2, the DFA
model also fit better than the P-technique model
(Δ = 2LL = 17.54, df = 4, p < 0.05). There
was evidence of carryover in fatigue (β = 0.52) but
not in well-being. However, there was a significant
cross-regression, with higher well-being leading to
higher fatigue the following day (β = 0.33),
the opposite of the expected buffering process. Perhaps
this results from increased activity engagement on
the previous day, resulting in some exhaustion on
the present day. For individual 3, the DFA model
again fit the data better than the P-technique model
(Δ = 2LL = 19.16, df = 4, p < 0.05).
However, the only significant lagged effect was an autoregression
for fatigue (β = 0.43), suggesting carryover in
fatigue from one day to the next, but with no evidence of other spillover or buffering processes.

Step 5: Between-Person Differences

The person-specific approach used here main-
tains that the analyses first examine the phenomena of
interest at the individual level. Models are fit
to individual-level P-data and the solutions and
files of those models evaluated one individual at a
time. We have described strategies for person-specific
approaches which also emphasize that the individual
solutions must at some point be integrated for purposes
of generalization (Nesselroade, 2007; Nesselroade
& Molenaar, 1999). For example, working in the
clinical context with subjects suffering from border-
line personality disorder, one might seek to classify
individuals into phenotypes, with some individu-
als' day-to-day changes in negative emotions and
self-destructive behaviors being well characterized
by persisting dynamics and other individuals being
characterized by more random changes. It may also
be of interest to understand how differences
in individuals' dynamic processes are related to
other individual differences. For example, theories
tests of cognitive development suggest that the abilities
and/or differences in other domains (e.g., lifespan,
interindividual differences in age would be related
to interindividual differences in the number of factors
needed to describe fluctuations in performance on
cognitive tasks.

The step-by-step procedures for conducting
dynamic factor analysis are purposely crafted to
maintain the integrity of the individual as the proper
unit of analysis when investigating psychological
processes. Between-person comparisons are made
only after the person-specific solutions have been
obtained. Several approaches have been used to
identify the similarities and differences among
the individual-level models. Solutions from multiple
individuals can be examined with respect to spe-
cific characteristics of the model and its parameters,
including, for example, the number of factors, the
number of factor loadings, and the autoregression
and cross-regression (e.g., Hamaker, Nesselroade,
& Molenaar, 2007). Given the generally small sam-
ple sizes used in DFA studies, identification of similarities and differences in structure can usually
be summarized through qualitative descriptions of
individual-level results. As the sample size increases, these descriptions can be quantified as follows. Similarities and differences among patterns of factor loadings and/or regressions obtained from multiple samples (i.e., individuals) can be quantified using congruence coefficients or other measures of pattern similarity. Modern computing provides the possibility to assess the fit of several a priori models to many individuals' P-data quickly and efficiently. This allows that models from multiple individuals can be integrated and compared through formal statistical tests. In particular, making use of multiple-group equality constraints, it is possible to formally test whether two or more individuals' data can be described by the same factor model parameter. The logic of such pairwise tests exactly follows the logic underlying tests for measurement invariance across multiple groups or occasions (Meredith, 1993). Specifically, observations from each individual are conceptualized as separate groups, with confirmatory models being fit to the multigroup data with and without equality constraints. Nested model comparisons provide evidence that the individual models can be considered equivalent or different. Like individuals can be described by the same model and separated from unlike individuals. We underscore that, as per the bottom-up strategy, identification, description, and testing of between-person similarities and differences should be completed only after individual-level models have been obtained and examined.

**Step 5: Empirical Illustration**

Looking across participants at a global level, each of the three participants' day-to-day experiences were better represented by some type of dynamic process than as a collection of independent states. The factor loadings for the indicator items were generally high for all participants, and there was some type of carryover from one day to the next, most consistently in fatigue, which carried over from day-to-day for all three women. Similarities among the solutions were tested formally using multigroup models. Specifically, treating each individual as a separate group, we tested whether the parameters for each pair of individuals/groups were invariant (i.e., factor loadings and the cross-regression coefficients). Results are given in Table 21.2. None of the three pairwise tests were significant (all $\Delta = 2LL < 11$, critical value [df = 10] = 18.31), suggesting that the women were highly similar. A three-group (person) invariance test was also non-significant ($\Delta = 2LL = 18.62$, critical value [df = 20] = 31.41), indicating that the dynamics within all three participants' data could be represented by the same model. The parameter estimates for this invariant model, shown in the first column of Table 21.2, suggest spillover-type processes that contribute to the ongoing maintenance of both well-being and fatigue, without evidence for buffering of one affective state on the other. In sum, the between-person similarities suggest some homogeneity of maintenance processes across this group of three women. However, before generalizing further, we should not forget that two other women were not included in the analysis because their daily reports on this particular item set did not show sufficient variance for meaningful analysis. Their affective experiences may be characterized by a different set of processes or by the same processes but different indicators (as will be discussed in the following section).

**Future Directions**

The first papers on DFA appeared in the psychological literature in the 1980s (e.g., Molenkamp, 1985). However, despite the recent increase in methodologically oriented papers, the application of these models in substantively oriented studies remains rather limited (cf. Chow et al., 2004). This lack may be rooted in the fact that many areas of social and behavioral science have been focused on modeling of between-person differences (In press in stable traits—e.g., personality). The time-series data needed for focused study of within-person processes have simply not been collected. In other fields, including engineering and economics, where time-series data are obtained as a matter of course, within-person or within-entity modeling traditions hold significant traction (see Chow et al., 2010). In those fields, dynamic models are part of the standard paradigm. As the social and behavioral sciences evolve from use of relatively "static" core representations of phenomena (e.g., traits) toward more "dynamic" representations, dynamic factor models and person-specific modeling approaches will become more pervasive. Already we see substantial movement toward collection of more intensive time-series in the promotion and use of intensive longitudinal, diary, and ecological momentary assessment designs (e.g., Bolger et al., 2003; Skinner et al., 2008; Wals & Schafer, 2008). As mobile technologies become more and more ubiquitous, the data constraints will fall away, and there will be tremendous opportunities for application of
are limited. Non-stationary extensions are needed to adequately capture the changes that are occurring simultaneously at multiple time-scales. Dynamic factor analysis has been extended to accommodate and model non-stationary time series (Molenar, 1994; Molenar, De Gooyer, & Schmitt, 1992). In the P-technique and DFA models presented here, the parameters (e.g., Λ, B) were assumed to be constant over time—time-invariant. In the non-stationary extension, the parameters become time-varying, so that transient or trend-type changes in the parameters accommodate and describe how the structure or process (e.g., factor loadings, autoregressions) changes or transforms over time. We point to some additional state-space models that may provide for useful extensions of the DFA approach.

Kim and Nelson (1999) outlined the use and estimation of multimodal state space models (see also Hamaker, Graimann, & Kamphuis, 2010). The core idea is that ongoing processes may switch among two or more regimes. For one period of time, the process may be well described by one set of parameters. However, after some event, change in context, or on particular occasions, the process is described by a different set of parameters. Consider the one-DMF (as a state-space model) given in Equations 2 and 3, with an additional subscript denoting a time-varying, categorical switching variable, z(t):

\[ y(t) = \Lambda z(t) \xi(t) + e(t), \]

\[ \eta(t) = Bz(t)\eta(t-1) + \zeta(t), \]

where the elements of Λ and B new differ depending on the value of z(t). When z(t) = 0, the evolution of the process is governed by one set of parameters, Λ0 and B0. When z(t) = 1 by a different set of parameters, Λ1 and B1. This family of models accommodates the types of non-stationarity that might accompany discrete changes in context (e.g., experimental conditions) or measured or latent Markov (i.e., sequentially dependent) switching.

For more continuous evolution, the multimodal model can be straightforwardly extended to a model with fully time-varying parameters. Here, all the parameters are time-varying.

\[ y(t) = \Lambda(t)\xi(t) + e(t), \]

\[ \eta(t) = B(t)\eta(t-1) + \zeta(t), \]

with the constraint that the parameter matrices, Λ(t) and B(t), change in a smooth fashion and slowly relative to the states, η(t). Building on applications from engineering, Molenar and colleagues

have recently described and fitted such models to psychological time series using extended Kalman filtering with iterated smoothing (Molenar, Sinclair, Rovine, Ram, & Cornish, 2009; Molenar & Ram, 2009). Initial results have demonstrated the viability and promise of such models. However, further work is needed to establish the limitations of the procedure and the types of processes and changes that can be captured by such models.

Given the prevalence of cyclical trends in bio-behavioral processes (e.g., diurnal and circadian activity), additional forms of change and non-stationarity that accommodate cyclical dynamics may also be useful (see Chow, Hamaker, Fujita, & Boeker, 2009). State-space models that incorporate time-varying parameters that are tied to cyclical components (e.g., sine and cosine functions) of specific frequencies or even time-varying frequencies can be used to model complex nonlinear changes in how processes manifest over time (e.g., Harvey & Sumner, 1998; Young, Pedregal, & Tych, 1999). In principle, the elements of Λ(t) and B(t) are fitted with sinusoidal elements. These models potentially provide a framework for mixing models from the time-domain and the frequency-domain, for modeling changes in amplitude and frequency of oscillatory processes, and discrete shifts in cyclicity or phase (when integrated with the regime shift model given above). Emerging from the econometrics literature, application to bio-behavioral processes is just beginning. Further work is needed to establish the specific processes and data streams to which cyclical versions of the dynamic factor models are best suited. In sum, non-stationarity is a reality of human function to be dealt with. Models that can do so are available and are currently being adapted for use with human data. As that trend (mind the pun) continues, our ability to describe and predict the complex changes that characterize real life will expand.

**Adaptive Guidance**

Moving beyond description and prediction, our scientific goals include the explanation and potential modification of human behavior. Control theory developed in engineering and mathematics as a framework for guiding systems toward desired states or outputs (e.g., Won & Han, 2005). In brief, a controller manipulates time-varying inputs into a system to steer an ongoing process. An additional vector of input variables is introduced into the state space model:

\[ y(t) = \Lambda(t)\xi(t) + e(t) + u(t), \]

where the new input variables are u(t), which can be seen as adaptive guidance.
\[ \eta(t) = \beta_3(t) \eta(t-1) + \beta_4(t) \eta(t-1) + \theta(t) \] (11)

so that the current state, \( \eta(t) \), is now a function of its past states, (random) innovations, and some (external) input. In cases where the input can be controlled, \( \eta(t) \) can be manipulated in such a way that \( \eta(t) \) can be steered toward a desired or optimal level (see Molenaar, 2010; Molenaar & Ram, 2010). For example, inputs of insulin can be used to control an individual’s blood glucose level (Molenaar, Ullbercht, Gold, Rovine, Wang, & Zhou, in press). Ubiquitous in engineering applications (e.g., steering rockets, managing electronic grids, optimizing chemical processes), sophisticated extensions of this simple model allow for analysis and guidance of systems with multiple inputs and outputs, nonlinearities, and complex evolutions over time. Theoretical descriptions of many psychological and behavioral processes make use of adaptive guidance and control-type language (e.g., Carver & Scheier, 1998). As interdisciplinary efforts foster the emergence of a combined engineering and social science, it is likely that the analytical techniques of control theory will be among the tools of this generation. The research between theory and method is strong. For example, in medicine, psychotherapy, and even physical therapy, the main objective is to develop personalized programs that optimize and guide individuals towards healthy states. Models making use of control theory principles and estimation algorithms are at the forefront of the person-specific modeling enterprise.

**Idiographic Filters**

The move toward personalized medicine suggests that person-specific approaches consider further how and in what ways models both generalize across persons and can be tailored to specific persons. In our example, we highlighted the importance of first examining the phenomena of interest at the individual level and only later considering between-person similarities and differences. Among the procedures used were confirmatory tests of model invariance, which provide a statistically rigorous framework for testing similarity. Specifically, we examined invariance in both the factor loadings and autoregression and cross-regression, \( \Lambda_1 = \Lambda \) and \( B_{11} = B_1, \) for all (or each pair of) individuals. Traditionally, invariance tests have concentrated only on the factor loadings (Meredith, 1993). Extending how between-person differences in structure or dynamics are approached, Neelsbrood and colleagues recently proposed that invariance tests should instead consider on similarity of factor correlations or autoregressions and cross-regressions (see Neelsbrood, Gestorff, Hardy, & Ram, 2007, and accompanying commentaries/ critiques). The proposal is that the latent processes or structures may be highly similar or even equivalent across individuals, although the indicator variables may be different. Idiographic filters allow for person-level differences in the manifestation of the same processes. For example, consider the simple dynamic factor model

\[ \gamma(t) = F \theta(t - 1) + \epsilon(t) \] (12)

\[ \eta(t) = \beta_3 \eta(t - 1) + \chi(t) \] (13)

where \( F \) is a matrix filled with 1s and 0s, and \( \epsilon(t) \) denotes the Hadamard product (i.e., elementwise product), \( \epsilon(t) = A \epsilon(t - 1) = B \epsilon(t - 1) \). Through this mechanistic approach, the person specific F matrix serves as a filter that organizes the latent state into particular manifest states—different ways for different people (see also Wildeman & Gráman, 2007).

Pushing these ideas a bit further, person-specific models might allow for filtering or tailoring in many different places—in the state equations, in the measurement equations (as above), in the configurations of dynamic noise, or in the configurations of measurement noise. Although the implications for measurement theory, and whether a new conceptualization is truly needed, remain unclear, evidence is building that highly tailored models hold utility (see above). Further use of tailoring in therapy and prevention efforts. Further work is needed to establish how additional algebraic tools, like the Hadamard product, can be used to expand the repertoire of processes and changes that can be captured by dynamic factor modeling frameworks.

**Synopsys**

The purpose of this chapter was to introduce dynamic factor analysis to researchers interested in modeling within-person processes. We reviewed a "bottom-up," person-specific approach to hypothesis testing and data analysis, wherein the relations among variables are first examined one individual at a time. Applied to within-person data, DFA can be used to identify and describe how a set of variables change together over time and to reveal the parametric structures that may underlie occasion-to-occasion changes and/or "carryover" in an individual’s behavior. Once the dynamic patterns are established at the individual level, they can be compared one person to the next, and the between-person differences and similarities described, quantified, and examined. It seems through recent extensions and developments, DFA is increasingly able to capture the complex and dynamic aspects of human behavior. We hope through our step-by-step illustration we have provided a guide for when and how dynamic factor models may be incorporated into psychological research programs and we take us further along the road toward describing, predicting, explaining, and potentially modifying individuals’ behavior.

**Glossary**

Person-specific analyses: Analyses of single entities’ time series (a single individual; \( N = 1 \))

Person-specific approach: Person-specific approaches seek to articulate the dynamics of the adaptive, regulatory, and other processes that proceed at the individual level.

Time series: observations obtained from the same entity on multiple occasions/repeated measurements obtained from the same person.

Sequential dependencies: Relationships across time between repeated measurements of the same variable (lagged effects) between different variables measured on subsequent occasions (cross-lagged effects).

Dynamic factor analysis: A combination of P-technique factor analysis and time-series analyses. The multivariate state of the individual at a given time is a function of both concurrent influences and influences of past states.

**References**


454 *DYNAMIC FACTOR ANALYSIS: MODELING PERSON-SPECIFIC PROCESS*


