Nesselroade Psyc 882

9

Pooling Lagged Covariance Structures Based on Short, Multivariate Time Series for Dynamic Factor Analysis

John R. Nesselroade & Peter C. M. Molenaar

THE IMPORTANCE of multivariate, intraindividual designs for studying process is well appreciated by students of behavior and behavior change. Unfortunately, optimal research designs are often not implemented because of the difficulty and expense of collecting an abundance of repeated measurements on large, representative samples of participants. We propose a method for modeling multivariate-process data that preserves the benefits of both intraindividual and group analysis while allowing design compromises that bring an encouraging line of inquiry within easier reach. The method involves evaluating statistically the validity of pooling the lagged covariance functions of multiple individuals' short time series for further dynamic analysis (e.g., Molenaar, 1985, 1994; Wood & Brown, 1994). We emphasize a focus on process, using idiographic information to pursue nomothetic laws, and the application of multivariate measurement and latent variable modeling.

A Focus on Process

The collection, analysis, and interpretation of data through which process can be modeled is an important aspect of contemporary behavioral research. Indeed, we find compelling the argument that behavioral phenomena cannot be properly understood until they can be cast in dynamic, change-process terms rather than the static, stability-oriented conceptions that have dominated not only psychology but science in general for the past couple of centuries (Gergen, 1977; Holling, 1973).

One fairly sturdy connotation of the term process is that pertinent data are temporally organized. Certain events precede others, changes in some of a system's parameters occur before others, and "recovery" of the system from perturbing outside forces may be rapid for some manifest variables, gradual for others. The feature of time-relatedness invokes special considerations in modeling repeated measurements.

Idiographic Emphases Within the Pursuit of Nomothetic Laws

A conceptually distinct, but nonetheless related emphasis to that on process involves the use of intensive information about individuals to build strong nomothetic relationships covering groups (Lamiell, 1981; Larsen, 1987; Nesselroade & Ford, 1985; Shoda, Mischel, & Wright, 1994; Zevon & Tellegen, 1982). The exploration of idiographic contributions to nomothetic relationships is consistent with a belief that group-based analyses and statements of relationship often are not very satisfying.

There are at least two reasons for dissatisfaction with group data. Analyses often lead to aggregating data over individuals

who are qualitatively different from each other, distorting the aggregate into an entity that has no parallel in the group (Lamiell, 1988). This can be especially troubling when the sample is composed of distinct subsamples. Second, because of fiscal and temporal constraints, group analyses are often based on relatively superficial attributes of individuals; "important" attributes of individuals are not being represented in the data. A corollary of this sentiment is that characterizations of individuals based on intensive measurement schemes will provide a more fruitful basis for group analyses.

Measuring an individual day after day, for example, and studying the changes can yield a different picture of how behavior "works" than measuring a group of individuals at a single time and analyzing the differences found among them (Lamiell, 1988). The two kinds of "portraits," one based on within-person changes and the other based on among-persons differences, may or may not be mutually consistent.

Familiar tools for analyzing individual time series call for large numbers of repeated observations. Unfortunately, research designs often are not underwritten with sufficient financial resources to permit the measurement of many people on many variables at many occasions of measurement. One seeks to find an optimal configuration of design parameters that meshes well with the research objectives (McArdle & Woodcock, 1995). One can swap persons for occasions of measurements, for example, but there are realistic limits on how far this option can be taken. Thus, practicable longitudinal research designs often do not involve either enough repeated measurements for traditional time series analyses or enough replicate individuals for traditional, large-sample analyses. Methods for extracting process-relevant information rigorously from such data are at a premium.

Multivariate Measurement and Analysis

A third line of emphasis has to do with collecting multivariate data in order to model process in terms of latent variables rather than observable ones (Baltes & Nesselroade, 1973;

Bentler, 1980; Cattell, 1966; Horn & McArdle, 1980). In behavioral research, tools for analyzing data obtained from merging frequently repeated measurements and multivariate observations include multivariate time series analysis (e.g., Holtzman, 1963; Larsen, 1987; West & Hepworth, 1991), stationary components analysis (Millsap & Meredith, 1988), and P-technique factor analysis (Cattell, 1963; Nesselroade & Ford, 1985; Zevon & Tellegen, 1982).

P-technique factor analysis, used somewhat sparingly for its nearly 50 years (e.g., Cattell, Cattell, & Rhymer, 1947), involves fitting the common factor model to one individual's multivariate time series (Cattell, 1963). The difficulty of collecting appropriate data and the limitations of P-technique factor analysis have made some researchers disdainful of using it (Holtzman, 1963; Molenaar, 1985; Steyer, Ferring, & Schmitt, 1992).

Refinements of the basic P-technique factor analysis methodology have involved both design modifications (e.g., studying replicates more or less concurrently to answer questions of generalizability; Jones & Nesselroade, 1990; Lebo & Nesselroade, 1978; Nesselroade & Ford, 1985; Zevon & Tellegen, 1982) and changing factor model specification (e.g., representing the dynamics resident in frequently repeated measurements; McArdle, 1982; Molenaar, 1985; Wood & Brown, 1994). The model specification changes have resulted in some very promising data analysis tools; one of which—dynamic factor analysis (Molenaar, 1985, 1994; Wood & Brown, 1994)—will be discussed subsequently.

Statement of the Problem

Given a high level of interest in time-dependent, multivariate data structures and the infeasibility, in many cases, of collecting hundreds of repeated observations on many participants, a promising research tool is a rationale and procedure for both (a) pooling relatively short time series information across limited numbers of participants and (b) analyzing the pooled information for its dynamic, process-relevant elements. It is these matters that we address.

What Is Needed?

A formal procedure for combining information across experimental units requires both a determination of the appropriateness of pooling and a scheme for analyzing the pooled information. Such a method promises huge dividends by enabling researchers to exploit data that contain information about dynamics and change but which, because of limited numbers of repeated observations, do not permit the application of traditional methods of analyses.

Earlier Work on the Problem

Efforts to combine cross-sectional and time series data can be found in the literature of various disciplines (e.g., Caines, 1988; Shumway, 1988). Within psychology, chain P-technique factor analysis (Cattell, 1963; Cattell & Scheier, 1963) is an early approach to the pooling of intraperson change information across cases. It involves separately standardizing the multivariate time series of two or more individuals, pooling (chaining) the sets of standardized scores into one long multivariate time series for factor analysis. Unfortunately, these techniques do not include a direct test for assessing the propriety of "chaining" information over participants and empirical demonstrations suggest that the blind application of pooling methods can mislead seriously regarding the structural characteristics of the repeated measurements (e.g., Daly, Bath, & Nesselroade, 1974).

Pooling Dynamic Structures Rather Than Individuals' Time Series

In contrast to the "chaining" of individuals' score matrices as in the chain P-technique or averaging time series across several individuals, we focus on first determining the lagged covariance function of each individual's multivariate time series and then, if justified, pooling these lagged covariance functions. Therefore, our use of the term pooling has a rather specialized connotation that is elaborated in the next section.

Assessing the Poolability of Individual Covariance Structures: A Test of Ergodicity

We first present a formal means for assessing the appropriateness of pooling dynamic information across multiple individuals. Subsequently, we discuss a statistical model—dynamic factor analysis—for analyzing the process information in the pooled structural descriptions.

Ergodicity

Assessing the "poolability" of individual's dynamic information is cast in the statistical mechanics terminology *ergodicity* (Arnold & Avez, 1968; see also Molenaar, 1994). Consider a dynamical system that is started up under given initial conditions. The p-variate time series output of this system is a trajectory in some region of p-dimensional Euclidean space called the phase space of the system. The system's dynamics can be characterized by a functional analysis of this trajectory. In particular, if the system is stochastic then averages (moments) can be taken along this trajectory to apprehend the system's dynamics. In contrast, consider a (possibly infinite) set of identical dynamic systems, each of which is started up under some idiosyncratic set of initial conditions (or what amounts to the same situation, a single dynamic system that is repeatedly started up under different initial conditions). One thus obtains a population of trajectories in phase space. In this case, the system dynamics can be apprehended by taking averages over the density of trajectories in phase space.

The key question is: Is the characterization of the system dynamics based on averages along a single trajectory equivalent to that based on averages over the density of multiple trajectories in phase space? Systems for which this equivalence holds are called ergodic. Speaking heuristically, an ergodic system "forgets" the initial conditions from which it started. Consequently, the trajectory describing its output covers the phase space in the same way as the outputs of a collection of identical replicas started from different initial conditions would cover it.

More to the point, is the structure of dynamics at the individual level sufficiently homogeneous across individuals that one can treat these individual dynamics as representing a common structure? A related question is: Do the processes that represent how individuals change also account for how individuals differ from one another at a given point in time? Cattell (1963), for example, argued that factors "... should have a unity of growth (or fluctuation) as well as a unity of structure in terms of static individual differences" (p. 168). It was on this basis that he argued for a coherence of intraindividual change factors derived from P-technique analyses and interindividual difference factors derived from cross-sectional (R-technique) factor analyses.

To construct a test of the appropriateness of pooling information across multiple cases, we focus on the covariance functions of the individual time series. In relation to the preceding distinction, single-subject time series analysis is tantamount to taking averages along a given trajectory, whereas in ordinary longitudinal analysis (e.g., panel data analysis) averages are taken over the density of trajectories associated with the sample of subjects. Pooling of the covariance functions over multiple subjects is only justified if the behavioral system under scrutiny is ergodic.[1] Thus, the basic data with which we are concerned are in the form of N (persons) data matrices of order T_i (occasions for the ith person) by p (variables) and the lagged covariance functions derivable from them.

Lagged Relationships

Lag will be used in two different senses. One has to do with relationships among observed variables. The magnitude of re-

[1] In classical test theory, the true score of a subject is defined as the average over repeated applications of the test to this subject (Lord & Novick, 1968). This is reminiscent of taking an average along a single trajectory. However, it is concluded that such repeated measurement of a single subject is not feasible due to memory effects, fatigue, and so forth. Hence, averages are taken over a single application of the test to a collection of subjects. This is reminiscent of taking an average over phase space. Thus, it seems in practice, at least, that psychometry rests, to some extent, on a conception of ergodicity. Similar arguments could be made in reference to scaling methods in which one-time judgments of several participants are used in place of repeated judgments of a single participant.

lationships between two variables differs depending on whether the observations on one variable are concurrent with, or lagged by one or more occasions of measurement on the observations on the other variable. The data can be "played" by leading and lagging variables on each other to extract more of the information inherent in the repeated measurements. In contrast to this "correlational" sense of lag, there is also the notion that "causal" relationships (e.g., between latent factors and observed variables) are time referenced. A factor's effect on a variable may be immediately large and then dissipate gradually with time or it may be delayed, reaching its greatest magnitude several occasions of measurement later and then dissipating.

The Statistical Test of "Poolability"

The first step in assessing the "poolability" of N individuals' lagged covariance structures involves the construction of N specialized covariance matrices, one from the time series data of each participant. These N specialized covariance matrices are then tested for lack of equality. The covariance matrices, of a form known as block-Toeplitz matrices, are constructed as follows.

Let $\mathbf{z}_i(t)$, $t = 1, 2, 3, \ldots, T_i$, denote the p-variate time series (T_i occasions in length) of the ith participant ($i = 1, 2, \ldots, N$). Let $\mathbf{C}_i(u)$ denote the $p \times p$ matrix-valued covariance function of $\mathbf{z}_i(t)$ at lag u, $u = 0, 1, 2, \ldots, w$. Thus, $\mathbf{C}_i(0)$ is the $p \times p$ covariance matrix for the ith person with no (zero) lagging of the variables on themselves or each other. It is the matrix one would obtain by treating the T_i occasions for person i as though they were $N = T_i$ cases and covarying the p variables across the N cases. This matrix is the one factored in traditional P-technique factor analysis. $\mathbf{C}_i(1)$ is the $p \times p$ matrix of lag 1 covariances between variables (i.e., times t and $t + 1$) for the ith person. This matrix, which although square is ordinarily not symmetric, contains the lag 1 autocovariances of the variables in its principal diagonal. The remaining $\mathbf{C}_i(u)$ up to some maximum lag ($u = w$) are constructed in a similar manner.

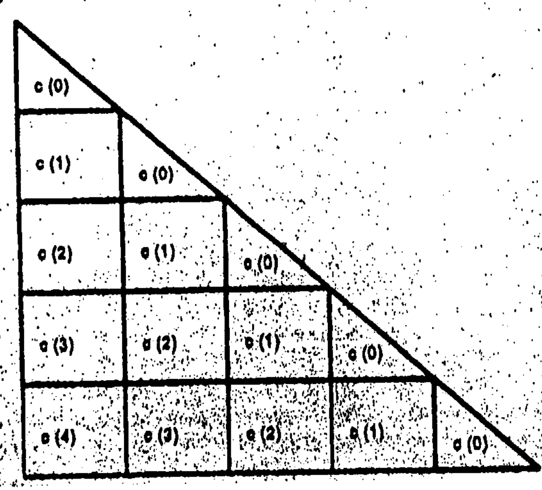
The various $\mathbf{C}_i(u)$ are put together to form a block-Toeplitz covariance matrix, $\mathbf{S}_i(w)$, for each individual as follows:

$$\mathbf{S}_i(w) = \mathbf{C}_i(j - k), \qquad j, k = 0, 1, \ldots, w,$$

with $\mathbf{C}_i(-u) = \mathbf{C}_i'(u)$, and $'$ denoting matrix transposition. Thus, each $\mathbf{S}_i(w)$ is a supermatrix, the submatrices of which are the unlagged and lagged covariance matrices based on the ith participant's data. The construction of a block-Toeplitz matrix for $w = 4$ is illustrated in Figure 1. It contains $w + 1$ lag 0 portions (symmetric), w lag 1 portions (asymmetric), $w - 1$ lag 2 portions (asymmetric), . . ., and one lag 4 portion (asymmetric). The diagonal entries of all submatrices represent autocovariances of the variables for the corresponding number of lags. For additional discussion of the nature of these block-Toeplitz matrices, see Wood and Brown (1994).

Testing the N block-Toeplitz matrices described previously for statistical equivalence requires the calculation of several intermediate values: $M = \sum_i M_i$, where $M_i = (T_i - 1) \times$

Figure 1 Toeplitz-Transformed Covariance Matrix for Lags 0, 1, 2, 3, and 4



$\ln[\det(S(\omega))] - \ln[\det(S_i(\omega))]$, $S(\omega) = \sum_i (T_i - 1)S_i(\omega) / \sum_i (T_i - 1)$, and T_i is the length of the time series, $z_i(t)$, for the ith participant. Also required are the values: $a = \sum_i [1/(T_i - 1)]$, $b = [1/\sum_i (T_i - 1)]$, and

$$c = 1 - \left[\frac{2m^2 + 3m - 1}{6(m + 1)(N - 1)}\right] \cdot \left[\frac{1}{a} - \frac{1}{b}\right].$$

where N is the number of subjects and $m = p(\omega + 1)$, which, by the way, is also the order of the matrices $S(\omega)$ and $S_i(\omega)$.

Under the null hypothesis that the individual participants' lagged covariance functions do not differ from each other, the product $c \cdot M$ is distributed approximately as a chi-squared variable (see Morrison, 1990, p. 297). Ordinarily, the appropriate degrees of freedom would be $m(m + 1)(N - 1)/2$ for $m \times m$ matrices. In this case, however, because of the redundancies in the block-Toeplitz lagged covariance matrix, the appropriate degrees of freedom is computed as $[1/2p(p + 1) + \omega p^2](N - 1)$.

Caution is appropriate regarding the use of chi-squared statistics for testing and model fitting in this context. The work of Taniguchi and Krishnaiah (1987) concerning sampling distributions of the covariance functions of time series and the eigenvalues and eigenvectors of the covariance function at lag 0 suggest the possibility that asymptotically the statistics we use to test poolability are chi-squared distributed if the observed multivariate time series of each subject is Gaussian. However, it is only a suggestion concerning an asymptotic result. Extensive, large-scale simulation work is needed to evaluate this possibility in the special case in which we are interested—relatively short time series for each participant. Such simulations should compare both the use of pseudo-maximum likelihood and asymptotically distribution-free estimation and the use of test statistics that are robust against misspecification of the distribution of the observed time series data (see, e.g., Bentler & Dudgeon, 1996). Lacking the information that simulation studies are expected eventually to provide, we view the results of the empirical example to follow as preliminary and awaiting further evaluation of the proposed test of poolability.

If the value of the test statistic is not significant, pooling the lagged covariance functions for the N participants to provide an estimate of a single-population lagged covariance function is statistically justified. Structural models representing change processes (e.g., the dynamic factor analysis model, Molenaar, 1985; Wood & Brown, 1994) can then be fitted to the pooled covariance function, $S(\omega)$.[2]

When a significant test statistic leads to rejection of the ergodicity hypothesis, one can try to create homogeneous subgroups by sequentially deleting the most "deviant" subject from the sample and reapplying the test. We have explored two ways to do this. The obvious way is to eliminate the person with the largest M_i value (defined previously) since M_i indicates the ith subject's contribution to the overall statistic (properly weighted by n_i). A given person's M_i value, however, is a function of the remaining $N - 1$ participants' data, so changes of only one person in the makeup of the sample can alter the relative size of a participant's M_i value. An alternative way to identify the most deviant subject has proven to be much more effective and is easily implemented. It involves eliminating each subject in turn, recalculating M for each subsample of $N - 1$ individuals. The subsample of $N - 1$ giving the smallest M is retained. If the test statistic is still significant, the procedure is repeated, calculating M for each possible subsample of $N - 2$. This algorithm can be repeated until either the test statistic is no longer significant or the original sample is decimated. When one subset of the sample has been selected as having "poolable" lagged covariance functions that subsample can be removed for dynamic factor analysis

[2] The test of equality of the individual block-Toeplitz matrices can be carried out using the multigroup option in LISREL. Each subject constitutes a group and one simply constrains the block-Toeplitz matrices to be invariant across groups. The chi-square goodness-of-fit statistic thus obtained is equivalent to M defined previously. Premultiplication by c can be carried out a posteriori. A drawback of the multigroup LISREL approach is that, as the number of variables increases, the computer memory requirements quickly become large and computation time is long. The LISREL-based ergodicity test makes clear that the chi-squared distribution for $c \cdot M$ is a likelihood ratio test where the numerator is the likelihood of the model constrained to be equal across groups and the denominator is the likelihood of the unconstrained set of block-Toeplitz matrices.

and the test procedure applied to the remainder of the sample to see if additional homogeneous subsets remain.

Dynamic Factor Analysis of Pooled, Lagged Covariance Functions

Dynamic factor analysis (Molenaar, 1985, 1994; Wood & Brown, 1994) is a merging of two important analytical tools—multivariate time series analysis and the common factor model. It was motivated by the potential value of factor analyzing multivariate time series coupled with the realization that the traditional common factor model did not fully exploit the information inherent in multivariate time series; indeed, that its application to time series data could be misleading with certain kinds of process information (Holtzman, 1963; Molenaar, 1985; Steyer et al., 1992).

The dynamic factor model (DFM) incorporating q factors and s lags of manifest variables on common factors (DFM(q, s)) is specified as

$$z(t) = \Lambda(0)\cdot\eta(t) + \Lambda(1)\cdot\eta(t-1) + \cdots + \Lambda(s-1)\cdot\eta(t-s+1) + \varepsilon(t), \qquad (1)$$

where $z(t)$ is the observed or manifest p-variate time series, $\eta(t)$ is the latent q-variate factor time series, $\varepsilon(t)$ is a p-variate noise time series, and $\Lambda(u)$, $u = 0, 1, \ldots, s-1$, are $p \times q$ matrices of lagged factor loadings. Thus, the various $\Lambda(u)$, $u = 0, 1, \ldots, s-1$, can differ from each other, signifying that the regressions of the variables on the factor vary with the amount of lag.

For the common factor model of lag 0 only, Equation 1 reduces to $z(t) = \Lambda\cdot\eta(t) + \varepsilon(t)$, which is the familiar P-technique factor model if both $\eta(t)$ and $\varepsilon(t)$ are white-noise series. The fuller model (Equation 1) indicates that the observed variables at time t are functions not only of the common factors at time t, but also of those same common factors up to $s - 1$ occasions earlier. In other words, the values of the common factors can influence the values of the observed variables both concurrently and in delayed fashion. Instead of being limited to immediate effects, the common

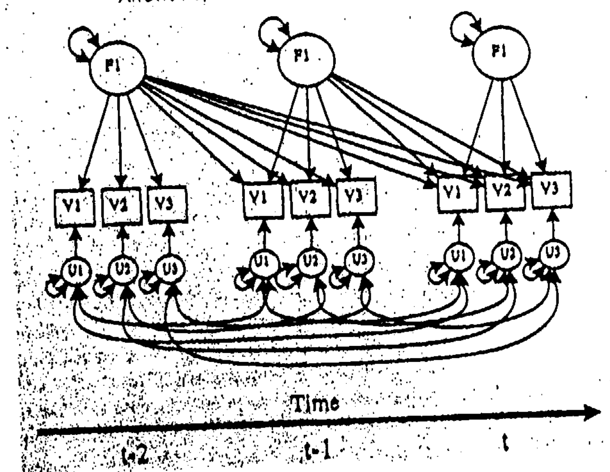
factors influence on the variables can be exerted over several occasions.

Notice that the specific noise series $\varepsilon_j(t)$, $j = 1, \ldots, p$, can be autocorrelated but the lagged cross-correlation between different specific noise series, for example, between $\varepsilon_j(t)$ and $\varepsilon_k(t+u)$, $j \neq k$, is assumed to be zero for all u, including $u = 0$. Thus, analogous to the assumption of uncorrelated unique variances in the traditional common factor model, a lack of correlation between the unique series for all lags, including zero, is assumed.

Alternatives to these structures can be fitted to data, provided one has enough information to identify the parameters. A dynamic factor model for one factor, two manifest variables, and $s = 3$, DFM(1, 3), is schematized in Figure 2.

After a sufficient length of time (lag) all stationary auto- and cross-correlations decay to zero. Thus, if the occasion sam-

Figure 2 Dynamic Factor Model for one Factor (F1) and Lags of 0, 1, and 2 [DFM(1,3)], Shown Here With Three Manifwst Variables (V1, V2, V3). Also Shown Are the Unique Factor (u1, u2, u3) and Their Autocovariances. One-Headed Arrows Represent Factor Loadings. Two-Headed Arrows Represent Variances and Covariances



pling frequency is low enough (i.e., if the time intervals are long enough), then the observed series will become white noise and appropriate for fitting by the traditional P-technique factor model.

An Empirical Example

The test of equality of lagged covariance functions and the follow-up dynamic factor analyses is demonstrated using a subset of data reported elsewhere (Eizenman, Nesselroade, Featherman, & Rowe, 1997; Kim, Nesselroade, & Featherman, 1996).

Data

The data are short, multivariate time series from a sample of 31 older adults. The mean age of the older adults who participated in the study was 77.5 years (SD = 7.2 years). They lived in a retirement community, Cornwall Manor, in central Pennsylvania. The participants, all of whom were volunteers, represent an above-average selection of individuals with regard to health status, education, and other demographic characteristics.

A subset of four cognitive performance and two biomedical variables was selected for the analyses reported here. Included are digit span forward (DSF), digit span backward (DSB), two trials on a delayed spatial recognition task called here Board Game 1 (BG1) and Board Game 2 (BG2), systolic blood pressure (SBP), and diastolic blood pressure (DBP). The SBP and DBP measurements are averages of multiple readings that were taken within each weekly session.

The subjects were measured each week for a period of 25 weeks. Subjects were "platooned" into a Monday, Wednesday, Friday group and a Tuesday, Thursday, Saturday group. Within those 3-day possibilities, measurement days were varied randomly across the 25-week period. Thus, a given individual in the MWF platoon might be measured on Monday of one week, Friday of the next week, and Wednesday of the next week, and so on. Measurements were done by personal interview between trained

testers and participants. Usually, the measurements were made in the living quarters of the participant. For this analysis, we ignore the fact that the weekly measurements were not evenly spaced.

Testing the "Poolability" of the Participants' Covariance Functions

A block-Toeplitz matrix as defined previously was constructed from each individual's data. The values for M, a, b, and c were calculated and the product $c \cdot M$ for $w = 1$ (thus including lags of 0 and 1) was obtained. Due to the relatively short time series, for $w > 1$ many of the block-Toeplitz matrices, $S_i(w)$, were singular and therefore unsuitable for testing ergodicity at lags larger than 1. The test statistic which, under the null hypothesis is distributed as chi-square, was 2,400.23 (df = 1,710), which is statistically significant well beyond the $p < .001$ level. Thus, the verdict was that, as a set, the lagged covariance functions of the N = 31 subjects were sufficiently different from each other that pooling them was not justified.

The iterative "search" analysis described previously identified a subsample of 10 individuals' lagged covariance functions that met the ergodicity criterion (χ^2 = 557.66, df = 513, p = .084). More will be said in the discussion concerning the implications of finding that only 10 of 31 cases' data met the criterion for inclusion in the dynamic factor analysis.

The outcome of the dynamic factor analysis reported next will pertain to both the general ensemble and to 10 individual time series contributing information to the ensemble. On the one hand, the lagged loadings pertain to all 10 time series so, in that sense, the factor loading patterns are invariant across the individual cases. On the other hand, the factor series for each individual can be determined. Thus, when multiple common factors are involved, for instance, it is possible to detect individual differences in the patterning of relationships among factor series over time. Hence, pooling across subjects does not imply that idiographic characterization of each individual is irretrievably lost.

Fitting the Dynamic Factor Model to the Pooled Covariance Functions

A series of dynamic factor models was fitted to the block-Toeplitz matrix $\mathbf{S}(w)$ derived from the pooled covariance function $\mathbf{C}(u)$ of the 10 subjects. First, one-factor dynamic models with increasingly lagged loadings [DFM(1, s), s = 1, 2, 3, 4, 5] were fitted.[3] Subsequently, we fitted a set of confirmatory, dynamic two-factor models with increasingly lagged loadings [DFM(2, s), s = 1, 2, 3, 4, 5] to the block-Toeplitz matrix. For the two-factor models, the first latent factor series was specified with lagged loadings on the four manifest cognitive series and the second latent factor series with lagged loadings on the two manifest blood pressure series. Thus, in all the model fits to be reported later, $w = 4$, meaning that u takes on five values, 0, 1, 2, 3, and 4. Hence, given that the manifest series are six-dimensional, it follows that $S(4)$ is a 30 × 30 matrix.

We fit the series of models using LISREL 8 (Jöreskog & Sörbom, 1993). LISREL 8 requires the user to input the number of observations (NO) on which the input covariance matrix is based. We calculated NO as follows for this situation:

$$NO = \left[\frac{1}{u_{max} + 1} \sum_{i=1}^{N_c} \sum_{u=0}^{u_{max}} (T_i - u) \right] - c,$$

where N_c is the number of cases on which the block-Toeplitz matrix is based, T_i is the number of occasions of measurement for the ith individual, u is the lag, and c is a correction for missing data devised by us for this purpose as the average number of missing values across all p-component series for the N_c subjects.

The rationale for fitting a series of models rather than a single a priori model in dynamic factor analysis was discussed by Molenaar (1985) and Wood and Brown (1994). Suffice it to say that, at this point in the development of dynamic factor analysis, choosing an optimal configuration of the number of factors

[3] The reader is reminded that DFM(1, 5), for example, specifies a one-factor model in which variables are lagged on factors with lags of 0, 1, 2, 3, and 4.

and the number of lags is not a matter of direct calculation but, rather, requires a systematic search process. Search methods certainly involve some risks (e.g., MacCallum, 1986), but, in the present case, the small sample size is compensated for somewhat by the fact that each participant was measured on 25 occasions. Formal characterizations of the roles of the number of participants and the number of occasions of measurement in ascertaining statistical power for this kind of modeling remain to be developed.

The series of model fits was evaluated by chi-square goodness-of-fit statistics and the derived Akaike's information criteria (AIC = chi-square − twice its df; cf. Bollen, 1989). The fit index values are presented in Table 1. It appears that DFM(1, 5) yields the best fit in terms of chi-square relative to its degrees of freedom. Interested readers can consult Molenaar (1985) for a discussion of how to adjust LISREL results for the redundancies in the block-Toeplitz matrices.

Estimates of Noise Series Parameters

Before presenting the factor loadings, we will briefly discuss the noise series that are also a part of this model. Let $\varepsilon_j(t)$ denote the specific noise time series for the jth manifest series and

Table 1 Dynamic Factor Model Fits to Pooled Covariance Functions Varying Number of Factors and Number of Lags (s)

	Dynamic Factor Model					
	DFM(1, s)			DFM(2, s)		
s	χ^2	df	AIC	χ^2	df	AIC
1	336.06	129	78.06	336.06	128	80.06
2	303.39	123	57.39	290.87	122	46.87
3	270.30	117	36.30	242.58	116	10.58
4	229.98	111	7.98	236.92	110	16.92
5	206.52	105	−3.48	205.16	104	−2.84

let $c_j(u)$, $u = 0, 1, \ldots$, denote the lagged autocovariance function of $\varepsilon_j(t)$. Notice that it is assumed that the lagged cross-covariance function between $\varepsilon_j(t)$ and $\varepsilon_k(t-u)$ is zero if $j \neq k$. The estimates of $c_j(u)$, $u = 0, 1, 2, 3, 4$, are presented in Table 2.

The lag 0 autocovariances of the noise series for the manifest cognitive series tend to be larger than those of the noise series for the manifest physiological series. In addition, it can be seen that the noise series specific to the diastolic blood pressure series has larger lagged autocorrelations than the other series. These values are consistent with the systemic role of diastolic blood pressure in the functioning human organism.

Table 2 Estimates of the Specific Noise Series' Lagged Autocovariance Function for Lags of 0, 1, 2, 3, 4 and Factor Loadings for One-Factor, Five-Lag Model

Variable	*Lag(u)*				
	u = 0	*u = 1*	*u = 2*	*u = 3*	*u = 4*
	Specific Noise Series Autocovariances				
Digit span forward	.95*	.18*	.09*	−.06	−.04
Digit span backward	.96*	−.07*	−.15*	−.02	.02
Board Game 1	.77*	−.28*	−.08	−.08	.02
Board Game 2	.97*	.11*	.09*	.04	−.16*
Systolic blood pressure	.76*	.11	−.04	.05	−.10
Diastolic blood pressure	.55*	.26*	.21*	−.36*	−.12
	Factor Loadings				
Digit span forward	.18*	.09*	−.03	−.06	.06
Digit span backward	.00	.11*	.07	.08*	−.13*
Board Game 1	.28*	.26*	.20*	.21*	.05
Board Game 2	.12*	.07	.12*	.04	−.03
Systolic blood pressure	−.13*	−.16*	.31*	−.14*	.29*
Diastolic blood pressure	−.42*	.08	.44*	−.25*	−.08

Note. * indicates that the parameter value is greater than twice its estimated standard error.

Interpretation of the Lagged Factor Loadings

The estimated factor loadings for this model are presented in the lower panel of Table 2. The loading pattern is rather complex, with all lags having at least two statistically significant loadings. The concurrent loadings (lag = 0) tend to be greatest in magnitude for the cognitive variables with the notable exception of DSB. It is instructive to compare the entire loading pattern with the lag = 0 column only. The latter are the relationships captured by traditional P-technique factor analysis. Taken alone, these concurrent loadings reflect an inverse relationship between the cognitive series and the physiological series but the lagged loadings elaborate the nature of the relationships over time in a much more detailed pattern.

In our experience thus far, interpretation of factor loading patterns should not be based only on direct inspection of the loadings. Rather, one should also focus on the auto- and cross-correlation functions derived from those lagged loadings. Therefore, we next examine the analog of the common parts of the variables in the traditional factor model—the communal part of the selected DFM(1, 5). The selected DFM(1, 5) is given by

$$\mathbf{z}(t) = \Lambda(0)\cdot\eta(t) + \Lambda(1)\cdot\eta(t-1) + \cdots + \Lambda(4)\cdot\eta(t-4) + \boldsymbol{\varepsilon}(t).$$

We label the expected covariance function of $\mathbf{z}(t)$ based upon the latent factor series the communal covariance function of $\mathbf{z}(t)$. Hence, the "total" covariance function of $\mathbf{z}(t)$ consists of the sum of the communal covariance function and the covariance function of $\boldsymbol{\varepsilon}(t)$. Let $c(u, j, k)$ denote the communal covariance function at lag u between the jth and the kth component series of $\mathbf{z}(t)$, $j, k = 1, 2, \ldots, 6$, and $u = -4, -3, -2, -1, 0, 1, 2, 3, 4$. Then $c(u, j, k)$ is estimated by

$$\begin{aligned} c(u, j, k) = E\Big[\big[&\lambda(0, j)\eta(t) + \lambda(1, j)\eta(t-1) \\ &+ \cdots + \lambda(4, j)\eta(t-4)\big]\cdot\big[\lambda(0, k)\eta(t+u) \\ &+ \lambda(1, k)\eta(t+u-1) + \cdots + \lambda(4, j)\eta(t+u-4)\big]'\Big]. \end{aligned}$$

Here, $\lambda(u, w)$ denotes the loadings at lag u of the wth component series of $\mathbf{z}(t)$ on the factor series. In the one-factor case, each

$\lambda(u, w)$ and each $\eta(t+u)$ is a scalar. For DFM(2, 5), for example, each $\lambda(u, w)$ would be a 1×2 row vector and each $\eta(t+u)$ would be a 2×1 column vector. For the DFM(2, 5) case, $c(u, j, k)$ is composed of 16 cross terms, but many of these can be zero because $E[\eta(t), \eta'(t+u)] = 0$, when lag $u \neq 0$. Only $E[\eta(t), \eta'(t)] \neq 0$. For DFM(1, 5), it is a scalar. For DFM(2, 5), it is a 2×2 correlation matrix with off-diagonal elements equal to the concurrent correlation between the two factor series. To illustrate specifically, consider $c(-3, i, j)$. This is given by

$$c(-3, j, k) = \lambda(3, j)V(\eta)\lambda'(0, k) + \lambda(4, j)V(\eta)\lambda'(1, k).$$

where $V(\eta) = E[\eta(t), \eta'(t)]$. The 21 communal correlation functions,

$$r(u, j, k) = \frac{c(u, j, k)}{\sqrt{c(0, j, j)}\sqrt{c(0, k, k)}},$$

and the standard deviations $\sqrt{c(0, k, k)}$ and $\sqrt{c(0, j, j)}$ for all j, k combinations are presented in Table 3. Notice that $r(u, j, k) = r(-u, k, j)$.

Several features of the elements in Table 3 are striking. One is the high correlation (+.71) between the communal parts of DSF and DSB, with the former leading the latter by 1 week ($u = 1$). Note that the relationship is not symmetric. When DSB leads DSF by 1 week, the correlation is −.06. The substantial relationship between DSF and DSB suggests that the process represented by the latent factor series that induces a high score (or a low score) in DSF at a given occasion tends to induce a high score (or a low score) in DSB a week later. Also of particular interest is the relationship between BG2, SBP, and DBP. There is a strong positive correlation (+.66) between BG2 and SBP with the former leading the latter by 2 weeks and a strong negative relationship (−.48) between BG2 and DBP with the former trailing the latter by 2 weeks. If the three variables are changing as a "system," one might expect to see a substantial negative correlation between SBP and DBP at 4 weeks with the former leading the latter. Indeed, this is the case: $r = -.37$ at $u = -4$.

Table 3 Standard Deviations (*SD*) and Communal Correlation Functions $r(u, j, k)$ for Variables j and k at Lag u

				Lag(u)								
j	*k*	*Sd(j)*	*Sd(k)*	*u* = −4	*u* = −3	*u* = −2	*u* = −1	*u* = 0	*u* = 1	*u* = 2	*u* = 3	*u* = 4
1	1	.22	.22	.22	−.11	−.26	.24	1.00	.24	−.26	−.11	.22
2	1	.20	.22	.00	.15	−.05	−.06	−.11	.71	.54	.06	−.53
2	2	.20	.20	.00	−.36	−.01	.07	1.00	.07	−.01	−.36	.00
3	1	.48	.22	.16	−.01	−.11	.17	.49	.58	.53	.31	−.08
3	2	.48	.20	−.38	−.12	.15	.39	.68	.33	.20	−.06	.00
3	4	.48	.48	−.06	.20	.43	.67	1.00	.67	.43	.20	−.06
4	1	.19	.22	.17	−.07	−.01	.09	.48	.57	.62	.11	−.13
4	2	.19	.20	−.41	.01	−.04	.59	.61	.36	.06	−.09	.00
4	3	.19	.48	−.06	.24	.36	.74	.93	.57	.41	.04	−.09
4	4	.19	.19	−.10	.08	.38	.56	1.00	.56	.38	.08	−.10

Note. 1 = DSF, 2 = DSB, 3 = BG1, 4 = BG2, 5 = SBP, and 6 = DBP.

Table 3 Continued

				Lag(u)								
j	k	Sd(j)	Sd(k)	u = −4	u = −3	u = −2	u = −1	u = 0	u = 1	u = 2	u = 3	u = 4
5	1	.49	.22	−.07	−.02	.30	−.31	−.20	−.13	.32	.01	.48
5	2	.49	.20	.17	.10	−.63	.18	−.45	.48	.05	.32	.00
5	3	.49	.48	.03	−.08	−.32	.03	−.25	.29	.46	.15	.34
5	4	.49	.19	.04	−.00	−.33	−.12	−.04	−.03	.66	.04	.37
5	5	.49	.49	−.16	−.12	.30	.47	1.00	−.45	.30	−.12	−.16
6	1	.67	.22	−.17	.20	.23	−.56	−.49	.45	.40	−.36	−.10
6	2	.67	.20	.41	−.33	−.60	.20	.22	.18	−.25	−.07	.00
6	3	.67	.48	.06	−.29	−.28	.04	−.18	.22	.13	−.28	−.07
6	4	.67	.19	.10	−.15	−.48	.04	.00	.06	.20	−.28	−.08
6	5	.67	.49	−.37	.25	−.04	−.13	.58	−.45	−.13	.14	.03
6	6	.67	.68	.08	.22	−.54	−.20	1.00	−.20	−.54	.22	.08

In all, the factor loading patterns and corollary information suggest that the relationships represented by this modeling procedure are somewhat more complex than are usually encountered with applications of traditional factor models and their close relatives. Not surprisingly, some of the relationships that the dynamic factor model highlights are not at all evident on mere inspection of the total (raw) correlation functions. Distinguishing between the communal (common) parts and the unique parts of variables is the great strength of the common factor. It is in "prying apart" these very different constituents of observed variables across time that we begin to see order and regularity. Thus, the findings illustrate the potential value of this general approach for building a better understanding of change processes as well as the day-to-day functioning of the organism.

Discussion and Conclusions

Many of our more cherished interindividual differences concepts manifest short-term changes that ought not to be discarded as "noise." However, the limitations of traditional P-technique factor analysis, both strenuous data collection requirements and questions about the use of the common factor model, have severely restricted its application to perplexing issues that might be usefully attacked from an intraindividual variability standpoint. The procedure we have presented here alters this situation in promising ways.

From a research design perspective, our proposals make it feasible to design research on process and change without going to the extreme of using only one participant to generate long series of observations. Investigators are more apt to collect time series data if they can aim for 25–50 repeated observations per experimental unit than if they must collect 100–500 observations. Obviously, the number of subjects and the number of occasions must be played off against each other in optimal ways that will have to be determined in part by trial and error, but the point is that methods such as we have presented make the systematic study of intraindividual variability a much more feasible and manageable prospect than it has been.

Analyses required to "parse" data of the kind used here are neither quick nor easy. However, the computational machinery is available (e.g., Wood & Brown, 1994), and it can be implemented on rather ordinary computing facilities. Today's bigger, faster computers allow the incorporation of more variables and lags in the analyses, so providing adequate representation of variables, occasions, and persons is not a barrier in design and analyses.

Making the pooling of information across participants an explicit, rational procedure eases concerns about aggregation fallacies. The procedures can be applied to a range of repeated measures situations from short multivariate time series to full-scale P-technique data collected on multiple participants. The fact that we found only 10 of 31 cases' data met the test of ergodicity is troubling in some respects, of course, but it may be telling us to search for general lawfulness at a more abstract level than is typically done. More to the point, perhaps, the lack of generality should be equally troubling to investigators who pool information across participants without mustering statistical support for their actions.

Substantively, the outcomes reinforce the notion that linkages between physiological and behavioral domains may not be captured well by looking only at direct, concurrent indications of relationships; rather order and organization may lie in more abstract, complicated, time-lagged relationships. Designing fruitful investigations of more complex relationship patterns will likely involve some trial and error regarding choice of intervals, number of occasions, and so forth, but the theorizing and empirical work of the past couple of decades should help.

Process information is hidden in streams of behavior, rather than in single-occasion or widely separated repeated measurements. It underlies the workings of a collection of manifest variables rather than in the observable action of a few and generality across persons and contexts is a hope yet to be realized. The tools we have presented are aimed at helping us rigorously extract information from relatively short time series of multiple experimental units. They promise to help nudge open a little farther

the door to the understanding of process; a door that now seems only barely ajar.

Author Notes: The John D. and Catherine T. MacArthur Foundation's Research Network on Successful Aging and the Max-Planck-Institute for Human Development, Berlin, Germany, supported the conceptual and empirical work presented here. We also thank Jack McArdle and Chris Hertzog for their valuable comments on an earlier version of the manuscript.

References

Arnold, V. I., & Avez, A. (1968). *Ergodic problems of classical mechanics*. New York: Benjamin.

Baltes, P. B., & Nesselroade, J. R. (1973). The developmental analysis of individual differences on multiple measures. In J. R. Nesselroade & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological issues* (pp. 219–251). New York: Academic Press.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 31, pp. 419–456). Palo Alto, CA: Annual Reviews.

Bentler, P. M., & Dudgeon, P. (1996). Covariance structure analysis: Statistical practice, theory, and directions. *Annual review of psychology* (Vol. 47, pp. 563–592). Palo Alto, CA: Annual Reviews.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Caines, P. E. (1988). *Linear stochastic systems*. New York: Wiley.

Cattell, R. B. (1963). The structure of change by P-technique and incremental R-technique. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 167–198). Madison: University of Wisconsin Press.

Cattell, R. B. (1966). Guest editorial: Multivariate behavioral research and the integrative challenge. *Multivariate Behavioral Research*, *1*, 4–23.

Cattell, R. B., Cattell, A. K. S., & Rhymer, R. M. (1947). P-technique demonstrated in determining psycho-physiological source traits in a normal individual. *Psychometrika, 12,* 267–288.

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press.

Daly, D. L., Bath, K. E., & Nesselroade, J. R. (1974). On the confounding of inter- and intraindividual variability in examining change patterns. *Journal of Clinical Psychology, 30,* 33–36.

Eizenman, D. R., Nesselroade, J. R., Featherman, D. L., & Rowe, J. W. (1997). Intra-individual variability in perceived control of older adults: The MacArthur Successful Aging Studies. *Psychology and Aging, 12,* 489–502.

Gergen, K. J. (1977). Stability, change and chance in understanding human development. In N. Datan & H. Rheese (Eds.), *Life span developmental psychology: Dialectic parspectives* (pp. 135–157). New York: Academic Press.

Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics, 4,* 1–23.

Holtzman, W. H. (1963). Statistical models for the study of change in the single case. In C. W. Harris (Ed.), *Problems in measuring change.* Madison: University of Wisconsin Press.

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980s.* Washington, DC: American Psychological Association.

Jones, C. J., & Nesselroade, J. R. (1990). Multivariate, replicated, single-subject designs and P-technique factor analysis: A selective review of intraindividual change studies. *Experimental Aging Research, 16,* 171–183.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Erlbaum.

Kim, J., Nesselroade, J. R., & Featherman, D. L. (1996). The state component in self-reported world views and religious beliefs of older adults: The MacArthur Successful Aging Studies. *Psychology and Aging, 11,* 396–407.

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist, 36,* 276–289.

Lamiell, J. T. (1988, August). *Once more into the breach: Why individual differences research cannot advance personality theory.* Paper presented at the annual meeting of the American Psychological Association, Atlanta, GA.

Larsen, R. J. (1987). The stability of mood variability: A spectral analytic approach to daily mood assessments. *Journal of Personality and Social Psychology, 52,* 1195–1204.

Lebo, M. A., & Nesselroade, J. R. (1978). Intra-individual differences dimensions of mood change during pregnancy identified in five P-technique factor analyses. *Journal of Research in Personality, 12,* 205–224.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100,* 107–120.

McArdle, J. J. (1982). *Structural equation modeling of an individual system: Preliminary results from "A case study in episodic alcoholism."* Unpublished manuscript, University of Denver.

McArdle, J. J., & Woodcock, R. W. (1995). *Modeling developmental components of change from time-lagged, test-retest data.* Unpublished manuscript, University of Virginia.

Millsap, R. E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika, 53,* 123–134.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time-series. *Psychometrika, 50,* 181–202.

Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 155–180). Thousand Oaks, CA: Sage.

Morrison, D. F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

Nesselroade, J. R., & Ford, D. H. (1985). P-technique comes of age: Multivariate, replicated, single subject designs for research on older adults. *Research on Aging, 7,* 46–80.

Shoda, Y., Mischel, W., & Wright, J. C. (1994). Intra-individual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. *Journal of Personality and Social Psychology, 67,* 674–687.

Shumway, R. H. (1988). *Applied statistical time series analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Steyer, R., Ferring, D., & Schmitt, M. (1992). States and traits in psychological assessment. *European Journal of Psychological Assessment, 8,* 79–98.

Taniguchi, M., & Krishnaiah, P. R. (1987). Asymptotic distributions of functions of the eigenvalues of sample covariance matrix and canonical correlation matrix in multivariate time series. *Journal of Multivariate Analysis*, *22*, 156–176.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality*, *59*, 609–662.

Wood, P. K., & Brown, D. (1994). The study of intraindividual differences by means of dynamic factor models: Rationale, implementation, and interpretation. *Psychological Bulletin*, *116*, 166–186.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology*, *43*, 111–122.

STATISTICAL STRATEGIES FOR SMALL SAMPLE RESEARCH

RICK H. HOYLE

SAGE Publications
International Educational and Professional Publisher
Thousand Oaks London New Delhi