

DFA

**Tools for
Dynamic Factor Analysis
and the
Pooling of Lagged Covariance Structures**

**Institute for Developmental & Health Research Methodology
at
The University of Virginia**

February 2004

**c/o Dr. John R. Nesselroade
PO Box 400400
Department of Psychology
University of Virginia
Charlottesville, Virginia 22904-4400**

Program Notes for DFA1

Introduction

DFA1 is a program that was developed so that researchers might more easily be able to 1) fit dynamic factor analysis (DFA) models and 2) assess the “poolability” of lagged covariance structures. This booklet is intended only as an instruction manual for use of the DFA1 program. We have not presented the theory and technical details underlying DFA models (e.g., Molenaar, 1985, 1994) or the pooling of lagged covariance matrices (Nesselroade & Molenaar, 1999). Users are referred to the references provided at the end of this document for such details (e.g., Nesselroade, McArdle, Aggen & Meyers, 2002). Note, however, that substantial familiarity with these details is a necessary prerequisite to conducting and interpreting the analyses and users are encouraged to digest this material before using the program.

Historically, DFA1 began with some fortran programs written by Peter C. M. Molenaar in the early 1990s to analyze the Cornwall Manor data set (see e.g., Nesselroade & Featherman, 1991; Nesselroade & Molenaar, 1999). Subsequent refinements and integration of the programs included contributions by John R. Nesselroade and Xinxi Wu. More recently, our efforts have been substantially upgraded by the programming skills of Zhiyong (Johnny) Zhang who joined our laboratory in the Fall of 2003. Nilam Ram has actively coordinated the development of this manual.

1. Fitting DFA models with DFA1

DFA models can generally be fit to data using standard structural equation modeling software. However, because of the redundancies involved in fitting the block—Toeplitz lagged covariance functions (see e.g., Nesselroade & Molenaar, 2004), the programming scripts to use available software such as LISREL are extremely tedious to write and likely to be fraught with typographical errors. The DFA1 program was developed to ease this burden and thus make it much easier to fit such models. Using the DFA1 program, researchers can in only a few seconds, and with only a few keystrokes, generate the matrices and LISREL scripts needed to run such analyses. The user simply provides a data file, specifies the number of variables, number of factors, and number of lags in the desired model, the factor loading pattern, and the DFA1 program generates the associated LISREL code. One then only needs to execute the code and interpret the results.

2. Pooling Lagged Covariance Structures

DFA models represent multivariate, multioccasion data for a single individual. Oftentimes, however, it may be necessary to pool information across individuals because the available single time series are not of appropriate length (e.g., > 100 occasions of measurement). Nesselroade and Molenaar (1999) presented a pseudo-formal method for assessing whether or not the lagged covariance functions from different individuals might be reasonably combined. DFA1 uses this method to identify

“poolable” subsets of individuals and compute a lagged covariance matrix based on their combined information. DFA models can then be fit to this matrix. As noted above the LISREL code for such analyses can be generated automatically. In other words, the program has been implemented such that with a data file and a few keystrokes users can find poolable subgroups and run DFA models quickly and efficiently.

Below are instructions on (a) how to structure the initial data file and how to use the DFA1 program to (b) generate LISREL code for conducting DFA and (c), if desired, to find poolable subgroups within multivariate, multioccasion, multiperson data.

A: Preparing the data set

Lagged covariance structures are developed from multivariate, multi-occasion data obtained on multiple individuals (e.g., a 3-dimensional data box, variables x persons x occasions). The data format used here is similar in nature to that of a person-period data set (multiple records per person). However, in the format used here rows are designated by person and variable while columns are designated by occasions (this is different than the usual rows = persons and occasions, columns = variables format). The data file needs to be in an ascii (text) file and formatted as follows:

subject# (numeric)	Variable# (numeric)	ScoreOccasion#1 (numeric)	ScoreOcc#2 (numeric)	...	ScoreOcc#n (numeric)
-----------------------	------------------------	------------------------------	-------------------------	-----	-------------------------

At least one space must separate each entry from the others. Rows are blocked first by variable and second, within those blocks, by individual. Thus, all individuals' Variable1 data are followed by all individuals' Variable 2 data, Variable3, etc... as shown below. Participant order within each variable must remain consistent. Do not include column headings. Missing data should be coded in a numeric manner (e.g., -999). In fact, all cell entries must be numeric (e.g., variables must be given identifying numbers rather than alphanumeric names).

An example file, rawdata.dat, with 3 persons, 2 variables and 10 occasions is shown below.

rawdata.dat

1	1	3	4	6	3	5	4	7	6	4	7
2	1	4	3	6	4	7	5	3	5	3	4
3	1	3	3	4	3	4	4	6	3	5	2
1	2	4	4	2	6	7	6	4	5	2	4
2	2	5	7	3	5	4	5	6	3	4	1
3	2	6	4	7	4	5	7	4	6	5	4

B: Generating a DFA script

The DFA1 program can be used to generate the LISREL script for fitting a user specified dynamic factor model to a set of data. Currently, the program generates a script for the Molenaar (1985) specification of the dynamic factor analysis model. An example script can be found in Appendix B (WNFS model) of Nesselroede, McArdle, Aggen, and Meyers (2001).

1. On the **File** menu, click **New** to bring up a new file for the script to be generated in.
2. On the **Analyze** menu, click **Generate Lisrel Script**.
3. Enter the appropriate **Parameter** information.
 - a. **Number of Occasions** = number of occasions in the data file.
 - b. **Number of Variables** = number of variables in the data file.
 - c. **Number of Lags** = number of lags in the model to be specified. *Note that for a model with only concurrent relationships (i.e., P-technique) the Number of Lags = 0.*
 - d. **Number of Cases** = number of persons in the data file.
 - e. **Number of Factors** = number of factors in the model at time t (i.e., number of factors that are indicated by the variables in the data file at any given occasion t).
 - f. **Pattern** = the pattern of factor loadings. If the number of factors is > 1 the user must supply a comma-delimited string of numbers indicating on which factor a given variable is saliently loaded. Integers, 1,2,3,etc., represent the factors and the serial position in the string represents the variable. Thus, 1,2,2,1 signifies that variable 1 loads on factor 1, variables 2 and 3 load on factor 2, and variable 4 loads on factor 1. The script is generated for the case of each variable loading on only 1 factor. The code must be corrected for other situations.
 - g. **Open Data File** = the location of the data file. For example, C:/DFA/rawdata.dat. Alternatively, clicking on the Open Data File button allows one to browse and select the data file through Windows.
4. Enter the associated **Variable Names**. The program has been implemented such that names will be assigned to the manifest and latent variables in the model, or the user can input their own variable names.
 - i. **Observed Variable Names** = names of the observed variables should be entered in the order in which they appear in the data set, and separate by spaces (e.g., Active Lively Peppy Sluggish Tired Weary).
 - ii. **Factor names** = names of the indicated factors in order, also separated by spaces (e.g., Energy Fatigue).
5. Click **OK**. The program will then generate the corresponding LISREL script. This file should then be checked and saved – on the **File** menu, click **Save As**.
6. Once the file is saved the DFA program can automatically pull up a LISREL window and the just generated script file for immediate fitting.

- a. On the **Analyze** menu, click **Run Lisrel Code**. The user can then work within LISREL as usual.

C: Pooled lagged covariance matrices

Additionally, the DFA1 program can be used to execute a pseudo—formal procedure for combining information across experimental units. This procedure follows the details outlined by Nesselroade and Molenaar (1999). Generally, the objective of the procedure is to identify a sub-group of individuals who have similar lagged covariance matrices and to create a single lagged covariance matrix that is representative of this pooled group. The program computes the lagged covariance structure (block-Toeplitz matrix) for each individual, assesses the appropriateness of pooling information over the individuals (ergodicity), and computes a lagged covariance matrix representative of the “poolable” set of individuals.

The algorithm proceeds as follows: First, the program calculates N block-Toeplitz matrices, one for each individual. Second, to test the “equivalence” of these matrices, the program computes Barlett’s test statistic for every possible sample of size N-1. The N-1 subsample that gives the best agreement (based on individual and total group determinants) is retained. In other words, the individual matrix that is the most different from the others is thrown out. The program then continues, again computing Barlett’s test statistic for every possible N-1 subsamples of the persons left (in actuality, now N-2 size samples). The subsample that gives the best agreement is retained and another individual is thrown out. The cycle is repeated until the remaining pool is not different from the individual matrices.

1. On the **File** menu, click **New** to bring up a new file for the script to be generated in.
2. On the **Analyze** menu, click **Generate Lisrel Script**.
3. Enter the appropriate **Parameter** information.
 - a. **Number of Occasions** = number of occasions in the data file.
 - b. **Number of Variables** = number of variables in the data file.
 - c. **Number of Lags** = number of lags over which to test the poolability of the sample. *Note that for only concurrent relationships the Number of Lags = 0.*
 - d. **Number of Cases** = number of persons in the data file.
 - e. **Open Data File** = the location of the data file. For example, C:/DFA/rawdata.dat. Alternatively, clicking on the Open Data File button allows one to browse and select the data file through Windows.
4. Enter the appropriate **Options** information.
 - a. **Ergodicity Level** = the strictness of the test. Barlett’s test statistic is distributed as a Chi-square, so a .05 level constitutes a more strict test than does .00001.
 - b. **Missing Data** = the numerical value in the data file that designates missing values.

5. Click **OK**. The program will then generate a list of the pooled individuals and the pooled matrix. This file can then be checked and saved – on the **File** menu, click **Save As**.
6. Once the file is saved the DFA program can automatically pull up a LISREL window and the just generated script file for immediate fitting.
 - a. On the **Analyze** menu, click **Run Lisrel Code**. The user can then work within LISREL as usual.

Dynamic Factor Analysis Papers

- Browne, M. W., & Nesselroade, J. R. (in press). Representing psychological processes with dynamic factor models: Some promising uses and extensions of ARMA time series models. In *McDonald Festschrift*.
- Ferrer, E., & Nesselroade, J. R. (2003). Modeling affective process in dyadic relations via dynamic factor analysis. *Emotion, 3*, 344-360.
- Hershberger, S. L. (1998). Dynamic factor analysis. In G. A. Marcoulides (Ed.) *Modern methods for business research: Methodology for business and management* (pp. 217-249). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hershberger, S. L., Molenaar, P. C. M., & Corneal, S. E. (1996). A hierarchy of univariate and multivariate structural time series models. In G. A. Marcoulides (Ed.) *Advances in structural equation modeling: Issues and techniques* (pp. 159-194). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kettunen, J., & Ravaja, N. (2000). A comparison of different time series techniques to analyze phasic coupling: A case study of cardiac and electrodermal activity. *Psychophysiology, 37*, 395-408.
- 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 factor analysis of psychophysiological signals. *Advances in psychophysiology, 5*, 229-302.
- Molenaar, P. C. M. (1994). Dynamic latent variable models in developmental psychology. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis : Applications for developmental research* (pp. 155-180). Thousand Oaks, CA: Sage.
- Molenaar, P. C. M. (1987a). Dynamic assessment and adaptive optimization of the psychotherapeutic process. *Behavioral Assessment, 9*, 389-416.
- Molenaar, P. C. M. (1987b). Dynamic factor analysis in the frequency domain: Causal modeling of multivariate psychophysiological time series. *Multivariate Behavioral Research, 22*, 329-353.
- Molenaar, P. C. M., de Gooijer, J. G., & Schmitz, B. (1993). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika, 57*, 333-349.
- Molenaar, P. C. M., Huizinga, H. M., & Nesselroade, J. R. (2003). The relationship between the structure of inter individual and intra-individual variability: A theoretical and empirical vindication of developmental systems theory. In U. M.

- Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 339-360). Boston: Kluwer Academic Publishers.
- Molenaar, P. C. M., & Nesselroade, J. R. (1998). A comparison of pseudo-maximum likelihood and asymptotically distribution-free dynamic factor analysis parameter estimation in fitting covariance-structure models to Block-Toeplitz matrices representing single-subject multivariate time-series. *Multivariate Behavioral Research*, *33*, 313-342.
- Molenaar, P. C. M., & Nesselroade, J. R. (2001). Rotation in the dynamic factor modeling of multivariate stationary time series. *Psychometrika*, *66*, 99-107.
- Nesselroade, J. R., & Featherman, D. L. (1991). Intraindividual variability in older adults' depression scores: Some implications for developmental theory and longitudinal research. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Törstadius (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 47-66). Cambridge: Cambridge University Press.
- Nesselroade, J. R., & Ghisletta, P. (2003). Structuring and measuring change over the life span. In U. M. Staudinger & U. Lindenberger (Eds.), *Understanding human development* (pp. 319-337). Boston: Kluwer Academic Publishers.
- Nesselroade, J. R., & Ghisletta, P. (2000). Beyond static concepts in modeling behavior. In L. R. Bergman & R. B. Cairns (Eds.) *Developmental science and the holistic approach* (pp. 121-135). Mahwah, NJ: Lawrence Erlbaum Associates.
- Nesselroade, J. R., McArdle, J. J., Aggen, S. H., & Meyers, J. M. (2002). Dynamic factor analysis models for representing process in multivariate time-series. In D. M. Moskowitz & S. L. Hershberger (Eds.), *Modeling intraindividual variability with repeated measures data: Advances and techniques* (pp. 235-265). Mahwah, NJ: Lawrence Erlbaum Associates.
- Nesselroade, J. R., & Molenaar, P. C. M. (2004, in press). Applying dynamic factor analysis in behavioral and social science research. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences*. London: Sage Publications.
- Nesselroade, J. R., & Molenaar, P. (2003). Quantitative models for developmental processes. In J. Valsiner & K. Connolly (Eds.), *Handbook of developmental psychology* (pp. 622-639). London: Sage Publications.
- Nesselroade, J. R., & Molenaar, P. C. M. (1999). Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 223-250). Newbury Park, CA: Sage Publications.

Shifrin, K., Hooker, K. A., Wood, P. K., & Nesselroade, J. R. (1997). The structure and variation in mood in individuals with Parkinson's disease: A dynamic factor analysis. *Psychology and Aging, 12*, 328-339.

Velicer, W. F., & Plummer, B. A. (1998). Time series analysis historiometry: A comment on Simonton. *Journal of Personality, 66*, 477-486.

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

Contributors