

CHAPTER 4

EXAMPLES: EXPLORATORY FACTOR ANALYSIS

Exploratory factor analysis (EFA) is used to determine the number of continuous latent variables that are needed to explain the correlations among a set of observed variables. The continuous latent variables are referred to as factors, and the observed variables are referred to as factor indicators. In EFA, factor indicators can be continuous, censored, binary, ordered categorical (ordinal), counts, or combinations of these variable types. EFA can also be carried out using exploratory structural equation modeling (ESEM) when factor indicators are continuous, censored, binary, ordered categorical (ordinal), and combinations of these variable types. Examples are shown under Confirmatory Factor Analysis.

Several rotations are available using both orthogonal and oblique procedures. The algorithms used in the rotations are described in Jennrich and Sampson (1966), Browne (2001), Bernaards and Jennrich (2005), and Browne et al. (2004). Standard errors for the rotated solutions are available using algorithms described in Jennrich (1973, 1974, 2007). Cudeck and O'Dell (1994) discuss the benefits of standard errors for rotated solutions.

All EFA models can be estimated using the following special features:

- Missing data
- Complex survey data
- Mixture modeling

The default is to estimate the model under missing data theory using all available data. The LISTWISE option of the DATA command can be used to delete all observations from the analysis that have missing values on one or more of the analysis variables. Corrections to the standard errors and chi-square test of model fit that take into account stratification, non-independence of observations, and unequal probability of selection are obtained by using the TYPE=COMPLEX option of the ANALYSIS command in conjunction with the STRATIFICATION,

CHAPTER 4

CLUSTER, and WEIGHT options of the VARIABLE command. The SUBPOPULATION option is used to select observations for an analysis when a subpopulation (domain) is analyzed.

Graphical displays of observed data and analysis results can be obtained using the PLOT command in conjunction with a post-processing graphics module. The PLOT command provides histograms, scatterplots, plots of eigenvalues, individual observed and estimated values, and plots of sample and estimated means and proportions/probabilities. These are available for the total sample, by group, by class, and adjusted for covariates. The PLOT command includes a display showing a set of descriptive statistics for each variable. The graphical displays can be edited and exported as a DIB, EMF, or JPEG file. In addition, the data for each graphical display can be saved in an external file for use by another graphics program.

Following is the set of EFA examples included in this chapter.

- 4.1: Exploratory factor analysis with continuous factor indicators
- 4.2: Exploratory factor analysis with categorical factor indicators
- 4.3: Exploratory factor analysis with continuous, censored, categorical, and count factor indicators*
- 4.4: Exploratory factor mixture analysis with continuous latent class indicators
- 4.5: Two-level exploratory factor analysis with continuous factor indicators
- 4.6: Two-level exploratory factor analysis with both individual- and cluster-level factor indicators

* Example uses numerical integration in the estimation of the model. This can be computationally demanding depending on the size of the problem.

EXAMPLE 4.1: EXPLORATORY FACTOR ANALYSIS WITH CONTINUOUS FACTOR INDICATORS

```
TITLE:      this is an example of an exploratory
             factor analysis with continuous factor
             indicators
DATA:       FILE IS ex4.1a.dat;
VARIABLE:   NAMES ARE y1-y12;
ANALYSIS:   TYPE = EFA 1 4;
OUTPUT:     MODINDICES;
```

In the first part of this example, an exploratory factor analysis with continuous factor indicators is carried out. Rotated solutions with standard errors are obtained for each number of factors. Modification indices are requested for the residual correlations. In the second part of this example, the same exploratory factor analysis for four factors is carried out using exploratory structural equation modeling (ESEM).

```
TITLE:      this is an example of an exploratory
             factor analysis with continuous factor
             indicators
```

The TITLE command is used to provide a title for the analysis. The title is printed in the output just before the Summary of Analysis.

```
DATA:       FILE IS ex4.1.dat;
```

The DATA command is used to provide information about the data set to be analyzed. The FILE option is used to specify the name of the file that contains the data to be analyzed, ex4.1.dat. Because the data set is in free format, the default, a FORMAT statement is not required.

```
VARIABLE:   NAMES ARE y1-y12;
```

The VARIABLE command is used to provide information about the variables in the data set to be analyzed. The NAMES option is used to assign names to the variables in the data set. The data set in this example contains 12 variables: y1, y2, y3, y4, y5, y6, y7, y8, y9, y10, y11, and y12. Note that the hyphen can be used as a convenience feature in order to generate a list of names.

CHAPTER 4

```
ANALYSIS: TYPE = EFA 1 4;
```

The ANALYSIS command is used to describe the technical details of the analysis. The TYPE option is used to describe the type of analysis that is to be performed. By specifying TYPE=EFA, an exploratory factor analysis will be carried out. The numbers following EFA give the lower and upper limits on the number of factors to be extracted. The default rotation is the oblique rotation of GEOMIN. The ROTATION option of the ANALYSIS command can be used to select a different rotation. The default estimator for this type of analysis is maximum likelihood. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator.

```
OUTPUT: MODINDICES;
```

The MODINDICES option is used with EFA to request modification indices and expected parameter change indices for the residual correlations which are fixed at zero in EFA.

```
TITLE:      this is an example of an exploratory
            factor analysis with continuous factor
            indicators using exploratory structural
            equation modeling (ESEM)
DATA:      FILE IS ex4.1b.dat;
VARIABLE:  NAMES ARE y1-y12;
MODEL:    f1-f4 BY y1-y12 (*1);
OUTPUT:    MODINDICES;
```

The difference between this part of the example and the first part is that an exploratory factor analysis for four factors is carried out using exploratory structural equation modeling (ESEM). In the MODEL command, the BY statement specifies that the factors f1 through f4 are measured by the continuous factor indicators y1 through y12. The label 1 following an asterisk (*) in parentheses following the BY statement is used to indicate that f1, f2, f3, and f4 are a set of EFA factors. When no rotation is specified using the ROTATION option of the ANALYSIS command, the default oblique GEOMIN rotation is used. The intercepts and residual variances of the factor indicators are estimated and the residuals are not correlated as the default. The variances of the factors are fixed at one as the default. The factors are correlated under the default oblique GEOMIN rotation. The results are the same as for the four-factor EFA in the first part of the example.

EXAMPLE 4.2: EXPLORATORY FACTOR ANALYSIS WITH CATEGORICAL FACTOR INDICATORS

```
TITLE:      this is an example of an exploratory
            factor analysis with categorical factor
            indicators
DATA:       FILE IS ex4.2.dat;
VARIABLE:   NAMES ARE  u1-u12;
            CATEGORICAL ARE u1-u12;
ANALYSIS:   TYPE = EFA 1 4;
```

The difference between this example and Example 4.1 is that the factor indicators are binary or ordered categorical (ordinal) variables instead of continuous variables. Estimation of factor analysis models with binary variables is discussed in Muthén (1978) and Muthén et al. (1997). The CATEGORICAL option is used to specify which dependent variables are treated as binary or ordered categorical (ordinal) variables in the model and its estimation. In the example above, all twelve factor indicators are binary or ordered categorical variables. Categorical variables can be binary or ordered categorical. The program determines the number of categories for each variable. The default estimator for this type of analysis is a robust weighted least squares estimator. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. With maximum likelihood estimation, numerical integration is used with one dimension of integration for each factor. To reduce computational time with several factors, the number of integration points per dimension can be reduced from the default of 7 for exploratory factor analysis to as few as 3 for an approximate solution. An explanation of the other commands can be found in Example 4.1.

EXAMPLE 4.3: EXPLORATORY FACTOR ANALYSIS WITH CONTINUOUS, CENSORED, CATEGORICAL, AND COUNT FACTOR INDICATORS

```

TITLE:      this is an example of an exploratory
            factor analysis with continuous, censored,
            categorical, and count factor indicators
DATA:      FILE = ex4.3.dat;
VARIABLE:  NAMES = u4-u6 y4-y6 u1-u3 y1-y3;
            CENSORED = y4-y6(b);
            CATEGORICAL = u1-u3;
            COUNT = u4-u6;
ANALYSIS:  TYPE = EFA 1 4;

```

The difference between this example and Example 4.1 is that the factor indicators are a combination of continuous, censored, binary or ordered categorical (ordinal), and count variables instead of all continuous variables. The CENSORED option is used to specify which dependent variables are treated as censored variables in the model and its estimation, whether they are censored from above or below, and whether a censored or censored-inflated model will be estimated. In the example above, y_4 , y_5 , and y_6 are censored variables. The b in parentheses indicates that they are censored from below, that is, have a floor effect, and that the model is a censored regression model. The censoring limit is determined from the data. The CATEGORICAL option is used to specify which dependent variables are treated as binary or ordered categorical (ordinal) variables in the model and its estimation. In the example above, the factor indicators u_1 , u_2 , and u_3 are binary or ordered categorical variables. The program determines the number of categories for each variable. The COUNT option is used to specify which dependent variables are treated as count variables in the model and its estimation and whether a Poisson or zero-inflated Poisson model will be estimated. In the example above, u_4 , u_5 , and u_6 are count variables. The variables y_1 , y_2 , and y_3 are continuous variables.

The default estimator for this type of analysis is maximum likelihood with robust standard errors using a numerical integration algorithm. Note that numerical integration becomes increasingly more computationally demanding as the number of factors and the sample size increase. In this example, the four-factor solution requires four

dimensions of integration. Using the default of 7 integration points per factor for exploratory factor analysis, a total of 2,401 integration points is required for this analysis. To reduce computational time with several factors, the number of integration points per dimension can be reduced from the default of 7 for exploratory factor analysis to as few as 3 for an approximate solution. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. An explanation of the other commands can be found in Example 4.1.

EXAMPLE 4.4: EXPLORATORY FACTOR MIXTURE ANALYSIS WITH CONTINUOUS LATENT CLASS INDICATORS

```
TITLE:      this is an example of an exploratory
             factor mixture analysis with continuous
             latent class indicators
DATA:       FILE = ex4.4.dat;
VARIABLE:   NAMES = y1-y8;
             CLASSES = c(2);
ANALYSIS:   TYPE = MIXTURE EFA 1 2;
```

In this example, an exploratory factor mixture analysis with continuous latent class indicators is carried out. Factor mixture analysis uses a combination of categorical and continuous latent variables. Mixture modeling refers to modeling with categorical latent variables that represent subpopulations where population membership is not known but is inferred from the data. With continuous latent class indicators, the means of the latent class indicators vary across the classes as the default. The continuous latent variables describe within-class correlations among the latent class indicators. The within-class correlations follow an exploratory factor analysis model that varies across the latent classes. This is the mixtures of factor analyzers model discussed in McLachlan and Peel (2000) and McLachlan et al. (2004). Rotated solutions with standard errors are obtained for each latent class. See Example 7.27 for a confirmatory factor mixture analysis.

The CLASSES option is used to assign names to the categorical latent variables in the model and to specify the number of latent classes in the model for each categorical latent variable. In the example above, there is one categorical latent variable c that has two latent classes. The

ANALYSIS command is used to describe the technical details of the analysis. The TYPE option is used to describe the type of analysis that is to be performed. By specifying TYPE=MIXTURE EFA, an exploratory factor mixture analysis will be carried out. The numbers following EFA give the lower and upper limits on the number of factors to be extracted. The default rotation is the oblique rotation of GEOMIN. The ROTATION option of the ANALYSIS command can be used to select a different rotation. The default estimator for this type of analysis is maximum likelihood with robust standard errors. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. An explanation of the other commands can be found in Example 4.1.

EXAMPLE 4.5: TWO-LEVEL EXPLORATORY FACTOR ANALYSIS WITH CONTINUOUS FACTOR INDICATORS

TITLE:	this is an example of a two-level exploratory factor analysis with continuous factor indicators
DATA:	FILE IS ex4.5.dat;
VARIABLE:	NAMES ARE y1-y6 x1 x2 w clus; USEVARIABLES = y1-y6; CLUSTER = clus;
ANALYSIS:	TYPE = TWOLEVEL EFA 1 2 UW 1 1 UB;

In this example, a two-level exploratory factor analysis model with individual-level continuous factor indicators is carried out. Two-level analysis models non-independence of observations due to cluster sampling. An exploratory factor analysis is specified for both the within and between parts of the model. Rotated solutions with standard errors are obtained for both the within and between parts of the model. See Example 9.6 for a two-level confirmatory factor analysis.

The CLUSTER option is used to identify the variable that contains clustering information. The ANALYSIS command is used to describe the technical details of the analysis. The TYPE option is used to describe the type of analysis that is to be performed. By specifying TYPE=TWOLEVEL EFA, a two-level exploratory factor analysis will be carried out. The numbers following EFA give the lower and upper limits on the number of factors to be extracted. The first set of numbers are for the within part of the model. The second set of numbers are for

the between part of the model. In both parts of the model, one- and two-factors solutions and an unrestricted solution will be obtained. The unrestricted solution for the within part of the model is specified by UW and the unrestricted solution for the between part of the model is specified by UB. The within and between specifications are crossed. Factor solutions will be obtained for one factor within and one factor between, two factors within and one factor between, unrestricted within and one factor between, one factor within and unrestricted between, and two factors within and unrestricted between. Rotations are not given for unrestricted solutions. The default rotation is the oblique rotation of GEOMIN. The ROTATION option of the ANALYSIS command can be used to select a different rotation. The default estimator for this type of analysis is maximum likelihood with robust standard errors. The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. An explanation of the other commands can be found in Example 4.1.

EXAMPLE 4.6: TWO-LEVEL EXPLORATORY FACTOR ANALYSIS WITH BOTH INDIVIDUAL- AND CLUSTER-LEVEL FACTOR INDICATORS

TITLE:	this is an example of a two-level exploratory factor analysis with both individual- and cluster-level factor indicators
DATA:	FILE = ex4.6.dat;
VARIABLE:	NAMES = u1-u6 y1-y4 x1 x2 w clus; USEVARIABLES = u1-u6 y1-y4; CATEGORICAL = u1-u6; CLUSTER = clus; BETWEEN = y1-y4;
ANALYSIS:	TYPE = TWOLEVEL EFA 1 2 UW 1 2 UB;
SAVEDATA:	SWMATRIX = ex4.6sw.dat;

The difference between this example and Example 4.5 is that there is a combination of individual-level categorical factor indicators and between-level continuous factor indicators. The exploratory factor analysis structure for the within part of the model includes only the individual-level factor indicators whereas the exploratory factor analysis structure for the between part of the model includes the between part of the individual-level factor indicators and the between-level factor

CHAPTER 4

indicators. Rotated solutions with standard errors are obtained for both the within and between parts of the model.

The BETWEEN option is used to identify the variables in the data set that are measured on the cluster level and modeled only on the between level. Variables not mentioned on the WITHIN or the BETWEEN statements are measured on the individual level and can be modeled on both the within and between levels. The default rotation is the oblique rotation of GEOMIN. The ROTATION option of the ANALYSIS command can be used to select a different rotation. The default estimator for this type of analysis is a robust weighted least squares estimator using a diagonal weight matrix (Asparouhov & Muthén, 2007). The ESTIMATOR option of the ANALYSIS command can be used to select a different estimator. The SWMATRIX option of the SAVEDATA command is used with TYPE=TWOLEVEL and weighted least squares estimation to specify the name and location of the file that contains the within- and between-level sample statistics and their corresponding estimated asymptotic covariance matrix. It is recommended to save this information and use it in subsequent analyses along with the raw data to reduce computational time during model estimation. An explanation of the other commands can be found in Examples 4.1, 4.3, and 4.5.