Indiana University

University Information Technology Services

Confirmatory Factor Analysis using Amos, LISREL, Mplus, SAS/STAT CALIS*

Jeremy J. Albright and Hun Myoung Park

Last modified on April 2009

University Information Technology Services
Center for Statistical and Mathematical Computing
Indiana University
410 North Park Avenue Bloomington, IN 47408
(812) 855-4724 (317) 278-4740
http://www.indiana.edu/~statmath

* The citation of this document should read: "Albright, Jeremy J., and Hun Myoung Park. 2009. *Confirmatory Factor Analysis Using Amos, LISREL, Mplus, and SAS/STAT CALIS*. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University." http://www.indiana.edu/~statmath/stat/all/cfa/index.html

This document summarizes confirmatory factor analysis and illustrates how to estimate individual models using Amos 16.0, LISREL 8.8, Mplus 5.1, and SAS/STAT 9.1.[**]

1. Introduction
2. Exploratory Factor Analysis
3. Confirmatory Factor Analysis
4. Confirmatory Factor Analysis with Missing Data
5. Confirmatory Factor Analysis with Categorical Data
6. Conclusion

1. Introduction

Factor analysis is a statistical method used to find a small set of unobserved variables (also called latent variables, or factors) which can account for the covariance among a larger set of observed variables (also called manifest variables). A factor is an unobservable variable that is assumed to influence observed variables. Scores on multiple tests may be indicators of intelligence (Spearman, 1904); political liberties and popular sovereignty may measure the quality of a country's democracy (Bollen, 1980); or issue emphases in election manifestos may signify a political party's underlying ideology (Gabel & Huber, 2000). Factor analysis is also used to assess the reliability and validity of measurement scales (Carmines & Zeller, 1979).

Principle component analysis also reduces the number of variables, but it differs from *principle factor analysis* (Brown, 2006: 22). A factor (unobserved latent variable) is assumed to exert causal influence on observed variables, while the underlying causal relationship is reversed in principle component analysis; observed variables are linear combinations of latent variables in factor analysis, while principle components are (weighted) linear combinations of observed variables (Hatcher, 1994: 9-10, 69). Principle components account for total variance, while factors account for the common variance (as opposed to unique variance) of a total variance (Brown, 2006: 22; Hatcher, 1994: 69).

1.1. Exploratory versus Confirmatory Factor Analysis

Investigators wish to explore patterns in the data or to test explicitly stated hypotheses. *Exploratory factor analysis* (EFA), corresponding to the former task, imposes no substantive constraints on the data; there is no restrictions on the pattern of relationships between observed and latent variables. EFA is data driven (Brown 2006: 14). Each common factor is assumed to affect every observed variable and that the common factors are either all correlated or uncorrelated. Once model is estimated, factor scores, proxies of latent variables, are calculated and used for follow-up analysis.[1] General purpose statistical software packages such as SPSS, SAS, and Stata can perform EFA.

[**] Jeremy alone wrote the first edition in 2006 and then Hun revised introduction and added exploratory factor analysis and SAS/STAT CALIS sections to later editions.

[1] A factor score is a linear composite of the optimally-weighted observed variables, while a factor-based score is merely a linear composite (e.g., mean or sum) of the variables that demonstrated meaningful factor loadings (Hatcher, 1994: 31).

Confirmatory factor analysis (CFA), on the other hand, is theory- or hypothesis driven. With CFA it is possible to place substantively meaningful constraints on the factor model. Researchers can specify the number of factors or set the effect of one latent variable on observed variables to particular values. CFA allows researchers to test hypotheses about a particular factor structure (e.g., factor loading between the first factor and first observed variable is zero). Unlike EFA, CFA produces many goodness-of-fit measures to evaluate the model but do not calculate factor scores. CFA requires special purpose software packages such as Mplus, LISREL, Amos, EQS, and SAS/STAT CALIS.

Table 1. Explanatory and Confirmatory Factor Analysis

	EFA (Data-driven)	CFA (Theory-driven)
Constraint	N/A	Yes
Unstandardized solution	N/A	Yes
Standardized solution	Yes	Yes
Factor rotation	Yes	N/A
Factor scores	Yes	N/A
Hypothesis test	N/A	Yes
Goodness-of-fit	N/A	Yes
Software package	General purpose software	Mplus, LISREL, Amos, EQS, SAS CALIS

In fact, CFA is a special case of the structural equation model (SEM), also known as the covariance structure (McDonald, 1978) or the linear structural relationship (LISREL) model (Jöreskog & Sörbom, 2004). SEM consists of two components: a *measurement model* linking a set of observed variables to a usually smaller set of latent variables and a *structural model* linking the latent variables through a series of recursive and non-recursive relationships. CFA corresponds to the measurement model of SEM. Table 1 summarizes differences and similarities of EFA and CFA.

1.2. Model Specification and Path Diagram

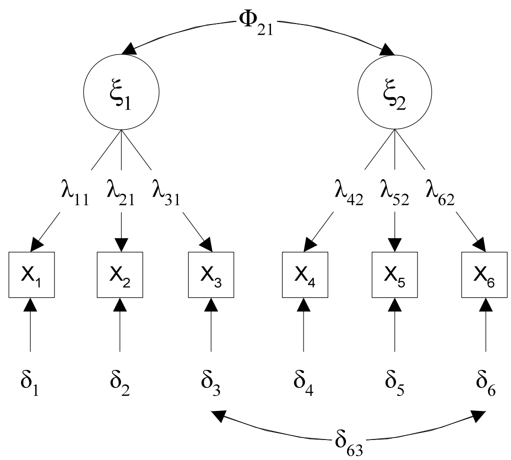
It is common to display confirmatory factor models as path diagrams in which squares represent observed variables and circles represent the latent variables. Figure 1 has two latent variables ξ_1 and ξ_2 in circles that are manifested by six observed variables x_1 through x_6 in squares. Single-headed arrows are used to imply a direction of assumed causal influence, and double-headed arrows represent covariance between two latent variables. Latent variables "cause" the observed variables, as shown by the single-headed arrows pointing away from the circles and towards the manifest variables.

Table 2. Notation for Confirmatory Factor Analysis

Name	Symbol	Matrix Form	Description
Ksi	ξ		Latent variable
x	x	X	Observed variable
Lambda	λ	Λ	Factor loading
Phi	ϕ	Φ	Factor variance and covariance
Theta delta	δ	Θ_δ	Error variance and covariance

The circles labeled ξ (ksi) represent latent variables or *(common) factors*. A factor can point to more than one observed variable; in Figure 1, ξ_1 causes three observed variables x_1 through x_3 and ξ_2 influences x_3 through x_6. The two ξ_i are expected to covary, as represented by ϕ_{21} on the two-headed arrow. *Factor loadings* are represented by λ_{ij}; λ_{31} is, for example, the effect (regression slope) of ξ_1 on x_3. The squared factor loading λ_{ij}^2 is referred to as a *communality* representing the proportion of variance in the *i*th observed variable that is explained by the *j*th latent variable (Brown, 2006: 61). The circles labeled δ_i (delta) represent *unique factors* because they affect only a single observed variable. The δ_i incorporate all the variance in each x_i, such as measurement error, which is not captured by the common factors. Finally, error in the measurement of x_3 is expected to correlate to some extent with measurement error of x_6, as represented by δ_{63}.[2] Table 2 summarizes CFA notation discussed so far.

Figure 1: Path Diagram of a Confirmatory Factor Model



When observed and latent variables are mean centered to have deviations from their means, the confirmatory factor model can be summarized by the equation

$$X = \Lambda\xi + \delta$$

in which *X* is the vector of observed variables, Λ (lambda) is the matrix of factor loadings connecting the ξ_i to the x_i, ξ is the vector of common factors, and δ is the vector of unique factors. It is assumed that the error terms have a mean of zero, $E(\delta) = 0$, and that the common and unique factors are uncorrelated, $E(\xi\delta')=0$. Equation 1 can be rewritten for Figure 1 as:

$$x_1 = \lambda_{11}\xi_1 + \delta_1 \qquad x_2 = \lambda_{21}\xi_1 + \delta_2 \qquad x_3 = \lambda_{31}\xi_1 + \delta_3$$
$$x_4 = \lambda_{42}\xi_2 + \delta_4 \qquad x_5 = \lambda_{52}\xi_2 + \delta_5 \qquad x_6 = \lambda_{62}\xi_2 + \delta_6$$

Here the similarities with regression analysis are evident. Each x_i is a linear function of one or more common factors plus an error term (there is no intercept since the variables are mean centered). The primary difference between these factor equations and regression analysis is that

[2] This may occur, for example, with panel data in which ξ_1 and ξ_2 represent the same concept measured at different points in time; if there is measurement error at t_1 it is likely that there will be measurement error at t_2.

the ξ_i are unobserved in CFA. Consequently, estimation proceeds in a manner distinct from the conventional approach of regressing each *x* on the ξ_i.

1.3. Identification

One essential step in CFA is determining whether the specified model is identified. If the number of the unknown parameters to be estimated is smaller than the number of pieces of information provided, the model is underidentified. For example, the equation 10 = 2x + 3y is not identified because it has two unknowns but only one piece of information (one equation). That is, an infinite number of values for *x* and *y* could make the equation true; the equation is not solvable. To make it just-identified, another independent equation should be provided; for example, adding 3 = x + y ends up with x=-1 and y=4. Provision of more than one independent equation will make it overidentified.

In CFA, a model is identified if all of the unknown parameters can be rewritten in terms of the variances and covariances of the *x* variables.[3] Unknown parameters of the CFA in Figure 1 are ϕ_{21}, six λ_{ij}, six δ_i, and δ_{63}. Information provided is variances and covariances of observed variables including $\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31} \ldots \sigma_{66}$.

$$\begin{matrix} \sigma_{11} \\ \sigma_{21} & \sigma_{22} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \\ \sigma_{51} & \sigma_{52} & \sigma_{53} & \sigma_{54} & \sigma_{55} \\ \sigma_{61} & \sigma_{62} & \sigma_{63} & \sigma_{64} & \sigma_{65} & \sigma_{66} \end{matrix}$$

The number of input information is 21=6(6+1)/2=p(p+1)/2, where p is the number of observed variables. Degrees of freedom are 7=21 (knowns) -14 (unknowns); this CFA is overidentified.

Without introducing some constraints any confirmatory factor model is not identified. The problem lies in the fact that the latent variables are unobserved and hence their scales are unknown. To identify the model, it therefore becomes necessary to set the metric of the latent variables in some manner. The two most common constraints are to set either the variance of the latent variable or one of its factor loadings to one.

1.4. Estimation

When the *x* variables are measured as deviations from their means it is easy to show that the sample covariance matrix for *x*, represented by *S*, can be decomposed as follows:

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Theta}$$

[3] A full discussion of the topic in the context of CFA is available in Bollen (1989, chapter 7), including some necessary and sufficient conditions for identification.

where Φ (phi) represents the covariance matrix of the ξ factors and Θ (theta) represents the covariance matrix of the unique factors δ (Bollen, 1989: 236). Estimation proceeds by finding the parameters $\hat{\mathbf{\Lambda}}$, $\hat{\mathbf{\Phi}}$, and $\hat{\mathbf{\Theta}}$ so that predicted *x* covariance matrix Σ (sigma) is as close to the sample covariance matrix *S* as possible. Several different fitting functions exist for determining the closeness of the implied covariance matrix to the sample covariance matrix, of which maximum likelihood is the most common.

This document includes examples using maximum likelihood estimation (MLE), including Full Information Maximum Likelihood (FIML) for situations in which there are missing values in the raw data file. However, MLE assumes multivariate normality among the observed variables, and preliminary diagnostics of sample data show strong deviations from normality for several of the variables. Alternative estimators exist for cases of non-normal data but for the most part lie outside the limited scope of this document. This document will also describe a weighted least squares (WLS) approach suitable for situations in which the *x* variables are categorical.

1.5. Goodness of Fit

A large class of omnibus tests exists for assessing how well the model matches the observed data. χ^2 is a classic goodness-of-fit measure to determine overall model fit. The null hypothesis is that the implied or predicted covariance matrix Σ is equivalent to the observed sample covariance matrix *S*, Σ=S. A large χ^2 and rejection of the null hypothesis means that model estimates do not sufficiently reproduce sample covariance; the model does not fit the data well. By contrast, a small χ^2 and failure to reject the null hypothesis is a sign of a good model fit. However, the χ^2 test is widely recognized to be problematic (Jöreskog, 1969). It is sensitive to sample size, and it becomes more and more difficult to retain the null hypothesis as the number of cases increases. The χ^2 test may also be invalid when distributional assumptions are violated, leading to the rejection of good models or the retention of bad ones. χ^2 is based on a very stringent hypothesis of Σ=S (Brown 2006: 81).

Due to these drawbacks of χ^2 test many alternative fit statistics have been developed, though each has its own advantages and disadvantages.[4] Another commonly reported statistic is the *Root Mean Square Error of Approximation* (RMSEA), a measure of fit introduced by Steiger and Lind (1980). RMSEA "incorporates a penalty function for poor model parsimony" and thus becomes sensitive to the number of parameters estimated and relatively insensitive to sample size (Brown 2006: 83-84). The Amos User's Guide suggests that "a value of the RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom," although "this figure is based on subjective judgment" and "cannot be regarded as infallible" (Arbuckle, 2005: 496). The Akaike Information Criterion (Akaike, 1987) and Schwarz's Bayesian Information Criterion (Schwartz, 1978) can be also used to compare models with respect to model parsimony.

Comparative fit index (CFI) evaluates "the fit of a user-specified solution in relation to a more restricted, nested baseline model," in which the "covariances among all input indicators are fixed to zero" or no relationship among variables is posited (Brown 2006: 84). CFI ranges from

[4] Appendix C of the Amos User's Guide provides summaries of many different fit measures (Arbuckle, 2005). For a thorough discussion of different tests see Bollen and Long's (1993) edited volume.

0 for a poor fit to 1 for a good fit. *Tucker-Lewis index* (TLI) is another index for comparative fit that "includes a penalty function for adding freely estimated parameters" (Brown 2006: 85). TLI can be interpreted in a similar fashion as CFI, but it can have a value outside of the range of 0 to 1 (p. 86).

There is no single evaluation rule on which everyone agrees. Hu and Bentler (1999) provide rules of thumb for deciding which statistics to report and choosing cut-off values for declaring significance. When RMSEA values are close to .06 or below and CFI and TLI are close to .95 or greater, for example, the model may have a reasonably good fit. Therefore, it is recommended to report not only χ^2 but RMSEA and CFI/TLI.

1.6. Software Issues

This document considers estimating confirmatory factor models using Amos 7.0 (Arbuckle, 2005); LISREL 8.8 (Jöreskog & Sörbom, 2004), and Mplus 5.1 (Muthén & Muthén, 2006). CFA and SEM can also be estimated using the CALIS procedure in SAS. All four programs are supported by the Stat/Math Center at Indiana University, while EQS, another popular SEM program, is currently not supported.

Mplus provides a variety of useful information in a concise manner. Mplus, LISREL, and SAS CALIS need a program describing a model to be estimated, while Amos supports both the point and click method and the program approach. Mplus and SAS CALIS use simple syntax structure, while LISREL (PRELIS, SIMPLIS, and LISREL) and Amos have a relatively abstruse grammar. Amos and LISREL produce a path diagram but Mplus and SAS CALIS do not. LISREL is able to go back and forth between a program and a path diagram, but Amos is not. To sum, Mplus and LISREL are generally recommended for confirmatory factor analysis.

Table 3. Comparison of Amos, LISREL, Mplus, and SAS/STAT CALIS

	Amos	LISREL	Mplus	SAS CALIS
Estimation	Amos Graphics, Program Editor	SIMPLIS, LISREL, path diagram	Program	Program
Path diagram	Yes	Yes	No	No
Data format supported	SPSS	Many formats	ASCII text	Many formats
Syntax (language)	Visual Basic, C#	LISREL, PRELIS, SIMPLIS	Mplus	SAS CALIS
Output	Messy	Normal	Concise	Messy
Platform supported	Windows	Windows, UNIX	Windows	Windows, UNIX
Unstandardized estimates	Yes	Yes	Yes	Yes
Standardized estimates and R^2	Yes w/o s.e.	Yes w/o s.e. on the path diagram only	Yes	Yes w/o s.e.
Covariances of factors	Yes w/o s.e.	Yes w/o s.e.	Yes	Yes w/o s.e.
Correlations of factors	Yes w/o s.e.	Yes w/o s.e.	Yes	Yes w/o s.e.
Goodness-of-fit	Many	Many	Several	Many
Residual (error) variances	Yes	Yes	Yes	Yes
Modification indices	Yes only in Amos Graphics	Yes	Yes	No

This document provides step-by-step examples for conducting a CFA with commonly used statistical software packages: Amos, LISREL, Mplus, and SAS/STAT CALIS. The next section provides an example of EFA with six observed variables. Section 3 begins with two-factor CFA with six observed indicators. Section 4 extends Section 3 to cover cases involving missing data. Section 5 discusses the commonly encountered situation in which the observed variables are categorical rather than continuous. Section 6 provides a brief summary.

2. Exploratory Factor Analysis

Before moving on to CFA, let us discuss sample data used in this document and briefly review exploratory factor analysis to contrast it with confirmatory factor analysis.

2.1 Sample Data

In politics commentators often use the terms left and right to describe the ideological positions of politicians and voters, but it is not always clear what exactly these terms mean. In the United States the political left is generally associated with favoring greater government involvement in the economy while the right is understood to favor market autonomy. Yet on moral issues such as abortion, assisted suicide, and gay marriage it is often the political right that favors a stronger regulatory role for government. Does a single dimension of values underlie Americans' views on both economic and moral issues? Or are there in fact two distinct value dimensions that underlie citizen attitudes?

This example uses data from the American sample of the European Values Survey (European Values Group and World Values Survey Association, 2005) to determine whether a model with one or two common latent factors adequately describes attitudes on economic and moral issues. The survey queried a random sample of 1,200 respondents about their economic, political, and moral values. Three questions summarizing economic attitudes and three questions summarizing moral attitudes, all measured on 10point scales, will be analyzed.

The economic items asked respondents if they felt private ownership of industry should be increased (PRIVTOWN), if the government should take more responsibility to see that all people are provided for (GOVTRESP), and whether competition brings out the best or worst in people (COMPETE). The moral items asked respondents how they felt about homosexuality (HOMOSEX), legalized abortion (ABORTION), and assisted suicide (EUTHANAS).

For this section missing data is handled by listwise deletion (all cases with missing observations on any indicator are removed). Listwise deletion resulted in dropping 40 of the original 1,200 observations, leaving a sample size of 1,160. The data is saved as the SPSS file *values.sav* located in the folder C:\temp\CFA. The data set has six variables and looks like the following:

```
     +-------------------------------------------------------------+
     | privtown   govtresp   compete   homosex   abortion   euthanas   |
     |-------------------------------------------------------------|
  1. |       1          3        2         2          2          2     |
  2. |       2          1        2         1          1          1     |
  3. |       7          3        3         4          3          3     |
  4. |       4          3        1         6          7          1     |
  5. |       8          1        1         1          5          1     |
  6. |       8          3        4         3          4          4     |
  7. |       6          6        6         1          1          1     |
```

2.2 An Explorative Factor Analysis

EFA seeks a smaller number of latent variables to explain variance and covariance among manifest variables. There are many methods to extract factors, such as principal factor (PF), maximum likelihood (ML), weighted least squares (WLS), generalized least squares (GLS), etc., of which PF and ML are most commonly used. Researchers may need to determine the number of factors extracted using eigenvalues calculated from input correlation matrix. As a rule of thumb, count the number of eigenvalues greater than 1 and use it as the number of factors (Brown, 2006: 26; Hatcher, 1994: 22-23). Alternatively, researchers may use scree test and parallel analysis (Brown 2006: 26-30). In the following output, there are two positive eigenvalues only one of which is greater than 1.

```
--------------------------------------------------------------------------
     Factor  |   Eigenvalue   Difference        Proportion   Cumulative
-------------+------------------------------------------------------------
    Factor1  |      1.18970      0.55561          0.9686       0.9686
    Factor2  |      0.63409      0.65191          0.5163       1.4849
    Factor3  |     -0.01782      0.11383         -0.0145       1.4704
    Factor4  |     -0.13165      0.07293         -0.1072       1.3632
    Factor5  |     -0.20458      0.03691         -0.1666       1.1966
    Factor6  |     -0.24149            .         -0.1966       1.0000
--------------------------------------------------------------------------
```

Once factors are extracted, researchers may rotate them to "foster interpretability by maximizing factor loadings close to 1.0 and minimizing factor loadings close to 0" (Brown 2006: 31). Factor rotation is either orthogonal or oblique. Factors are constrained to be uncorrelated in orthogonal rotation but not necessarily in oblique rotation. Varimax (orthogonal) and promax (oblique) rotations are commonly used. Keep in mind that any factor rotation does not alter factor loadings but change views of pattern matrix. Finally, researchers may calculate factor scores for future analysis.

```
-------------------------------------------------
    Variable |  Factor1   Factor2 |   Uniqueness
-------------+--------------------+--------------
    privtown |  -0.0203    0.5494 |      0.6977
    govtresp |   0.2070    0.1619 |      0.9309
     compete |   0.0231    0.5547 |      0.6918
     homosex |   0.6356   -0.0145 |      0.5958
    abortion |   0.6572   -0.0059 |      0.5681
    euthanas |   0.5514    0.0642 |      0.6919
-------------------------------------------------
```

The output above suggests two latent variables underlying six manifest variables. The numbers under first and second columns are factor loadings. Factor loadings of factor 1 and 2 on the perception on competition (COMPETE) are .0231 and .5547. The squared factor loadings, communalities, of $.0005=.0231^2$ and $.3077=.5547^2$ are respectively the proportions of variance in COMPETE that is explained by factor 1 and 2. That is, 31 percent of variance in COMPETE is explained by factor 2 and almost zero percent by factor 1. Unique variance is the proportion of variance that is not explained by any factor. For example, .6918 is calculated as 1-(.0005+.3077); the 69 percent of variance in COMPETE is not explained by two factors.

PRIVTOWN (private ownership) and COMPETE are largely explained by factor 2, while HOMOSEX (homosexuality), ABORTION, and EUTHANAS (assisted suicide) by factor 1. However, neither factor 1 nor 2 can explain GOVTRESP (government responsibility) sufficiently; it has the largest unique variance of 93 percent. Therefore, there appear to be two moral dimensions underlying citizen attitudes.

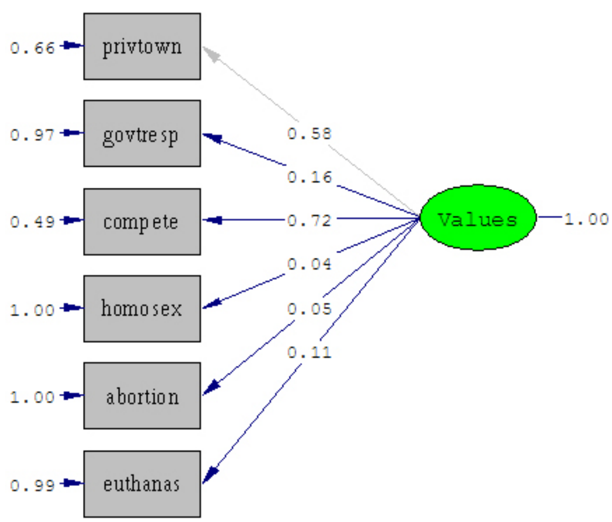
Two factor scores predicted here have zero mean and standard deviations of .7871 and .6616.

3. Confirmatory Factor Analysis

This section begins with one factor model and then moves forward to the two factor model.

3.1 One Factor Model

The following path diagram with standardized solutions illustrates the one factor model, where a latent variable *values* is manifested by six observed variables.[5] An oval and a rectangle represent a latent variable and a manifest variable, respectively. The numbers on arrows from the latent variable to observed variables are standardized factor loadings (regression weights). COMPETE and PRIVTOWN have large factor loadings of .72 and .58; they appear to be the best indicators of *values*. R^2 is a standardized factor loading squared that means the extent that a factor can explain the variance in a manifest variable. For example, the latent variable *values* explains about 51 percent (=$.72^2$) of variance in COMPETE. HOMOSEX, ABORTION, and EUTHANAS have poor factor loadings, suggesting that they appear to indicate other factors.



The following is the LISREL output of this confirmation factor model. The coefficients of *values* listed under the `Measurement Equations` heading are unstandardized factor loadings. The numbers in parentheses are standard errors followed by test statistics. For instance, the factor loading on COMPETE is 1.31; its standard error is.32; and the test statistic is 4.08 (=1.31/.32). R^2 is listed at the end of each equation.

```
                        DATE: 11/18/2008
                          TIME: 13:00

                        L I S R E L  8.80

                              BY

                  Karl G. Jöreskog & Dag Sörbom


            This program is published exclusively by
            Scientific Software International, Inc.
```

[5] See the LISREL section for details about fitting a confirmatory factor model and drawing a path diagram.

```
                    7383 N. Lincoln Avenue, Suite 100
                     Lincolnwood, IL 60712, U.S.A.
          Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
       Copyright by Scientific Software International, Inc., 1981-2006
         Use of this program is subject to the terms specified in the
                     Universal Copyright Convention.
                       Website: www.ssicentral.com

 The following lines were read from file c:\Temp\cfa\values.spj:

 One factor model
 Raw Data from file 'c:\Temp\cfa\values.psf'
 Sample Size = 1160
 Latent Variables Values
 Relationships
 privtown = 1.00*Values
 govtresp = Values
 compete = Values
 homosex = Values
 abortion = Values
 euthanas = Values
 Path Diagram
 End of Problem

 Sample Size =  1160

 Second Two factor model

        Covariance Matrix

           privtown   govtresp   compete   homosex  abortion   euthanas
           --------   --------   --------  --------  --------   --------
 privtown      5.05
 govtresp      0.51       7.20
  compete      2.27       0.67      5.71
  homosex     -0.22       1.39      0.01     10.57
 abortion     -0.11       1.01      0.06      4.81      8.85
 euthanas      0.14       0.72      0.38      3.00      3.05      6.03


 Second Two factor model

 Number of Iterations =  8

 LISREL Estimates (Maximum Likelihood)

        Measurement Equations


 privtown = 1.00*Values, Errorvar.= 3.35 , R² = 0.34
                             (0.44)
                              7.67

 govtresp = 0.33*Values, Errorvar.= 7.01 , R² = 0.026
         (0.083)              (0.30)
          4.00                23.64

  compete = 1.31*Values, Errorvar.= 2.78 , R² = 0.51
         (0.32)               (0.72)
          4.08                3.85

  homosex = 0.11*Values, Errorvar.= 10.55, R² = 0.0020
         (0.094)              (0.44)
          1.18                24.04

 abortion = 0.12*Values, Errorvar.= 8.82 , R² = 0.0030
         (0.086)              (0.37)
          1.44                24.03

 euthanas = 0.21*Values, Errorvar.= 5.96 , R² = 0.012
         (0.073)              (0.25)
```

```
          2.81               23.89


        Variances of Independent Variables

             Values
          --------
              1.70
            (0.44)
              3.84

                        Goodness of Fit Statistics

                          Degrees of Freedom = 9
             Minimum Fit Function Chi-Square = 641.04 (P = 0.0)
     Normal Theory Weighted Least Squares Chi-Square = 718.06 (P = 0.0)
             Estimated Non-centrality Parameter (NCP) = 709.06
         90 Percent Confidence Interval for NCP = (624.84 ; 800.67)

                     Minimum Fit Function Value = 0.55
             Population Discrepancy Function Value (F0) = 0.61
           90 Percent Confidence Interval for F0 = (0.54 ; 0.69)
           Root Mean Square Error of Approximation (RMSEA) = 0.26
          90 Percent Confidence Interval for RMSEA = (0.24 ; 0.28)
            P-Value for Test of Close Fit (RMSEA < 0.05) = 0.00

               Expected Cross-Validation Index (ECVI) = 0.64
          90 Percent Confidence Interval for ECVI = (0.57 ; 0.72)
                      ECVI for Saturated Model = 0.036
                     ECVI for Independence Model = 0.83

   Chi-Square for Independence Model with 15 Degrees of Freedom = 949.26
                         Independence AIC = 961.26
                             Model AIC = 742.06
                           Saturated AIC = 42.00
                         Independence CAIC = 997.59
                            Model CAIC = 814.73
                           Saturated CAIC = 169.18

                        Normed Fit Index (NFI) = 0.32
                     Non-Normed Fit Index (NNFI) = -0.13
                   Parsimony Normed Fit Index (PNFI) = 0.19
                     Comparative Fit Index (CFI) = 0.32
                     Incremental Fit Index (IFI) = 0.33
                      Relative Fit Index (RFI) = -0.13

                          Critical N (CN) = 40.17

                   Root Mean Square Residual (RMR) = 1.45
                          Standardized RMR = 0.17
                     Goodness of Fit Index (GFI) = 0.83
                Adjusted Goodness of Fit Index (AGFI) = 0.60
               Parsimony Goodness of Fit Index (PGFI) = 0.36

 The Modification Indices Suggest to Add an Error Covariance
  Between    and    Decrease in Chi-Square   New Estimate
 compete   privtown        204.2             42.00
 homosex   privtown          8.2             -0.62
 homosex   govtresp         28.1              1.35
 abortion  govtresp         17.0              0.96
 abortion  homosex         285.9              4.80
 euthanas  govtresp         10.3              0.62
 euthanas  homosex         162.9              2.99
 euthanas  abortion        201.6              3.04

                        Time used:   0.000 Seconds
```

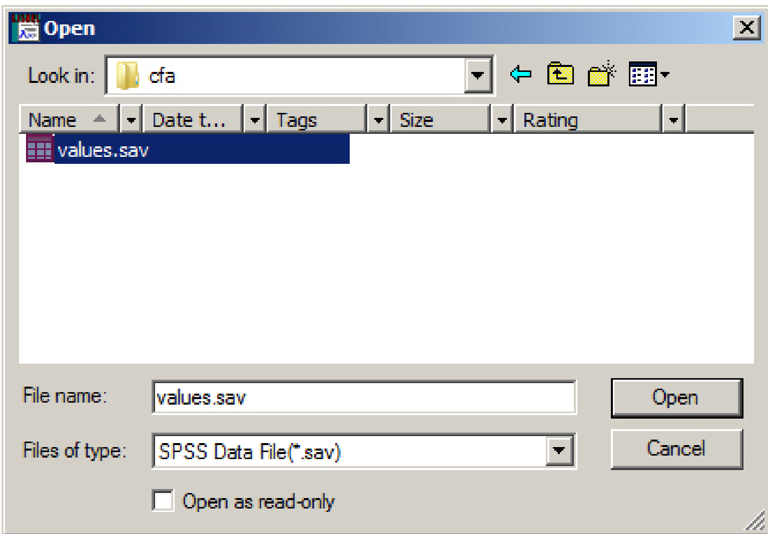
A various goodness-of-fit statistics are listed under `Goodness of Fit Statistics`. χ^2 is 718.06, which is so large that the null hypothesis of a good fit is rejected at the .05 level (p<.000). The degrees of freedom is 9 =21 - 12; there are 21 sample variance and covariance elements and 12 unknown parameters including six λ_{ij} and six δ_i. Root Mean Square Error of Approximation (RMSEA) 0.26 is also large enough to reject the null hypothesis ($p<.000$). Comparative Fit Index (CFI) 0.32 is small. Therefore, this one factor model shows a poor fit and needs to be modified somehow.

Modification indices at the bottom suggest that HOMOSEX, ABORTION, and EUTHANAS are closely related (*moral values*), while PRIVTOWN and COMPETE are grouped together (*economic values*). For instance, if you add covariance between HOMOSEX and ABORTION χ^2 will decrease by 285.9; two variables appear to manifest the same latent variable. Small R^2 of HOMOSEL (.002), ABORTION (.003), and EUTHANAS (.012) support this conclusion since the one factor explains practically no variance in these observed variables. It is likely that a two factor model is more appropriate to describe the economic and moral values of Americans It is not clear, however, whether GOVTRESP falls into *economic values* or *moral values*; let us first connect to *economic values* though.

3.2 Two Factor Model using LISREL

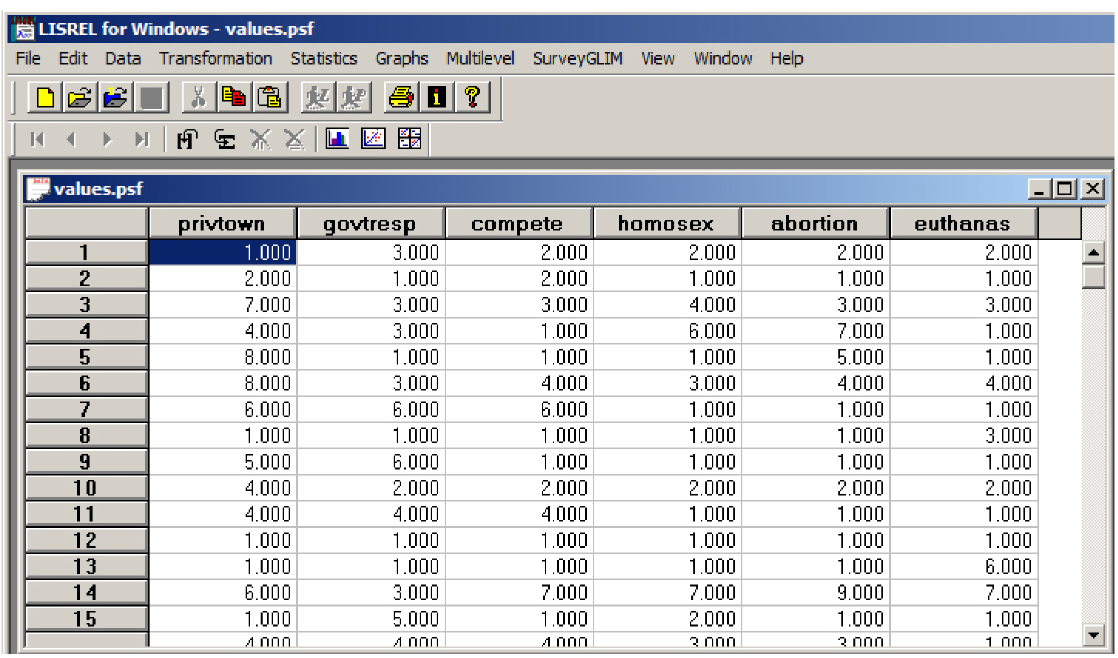
This section demonstrates how to estimate a two factor confirmatory factor model using LISREL. Note that the previous subsection revealed that the single common factor model was a poor fit to the data.

LISREL can be launched from any computer running Windows in the UITS Student Technology Centers by going to **Start** → **All Programs** → **Departmentally Sponsored** → **Statistics-Math** → **LISREL 8.80** → **LISREL 8.80**.



First, you need to import the data file into LISREL and save it as a .psf (PRELIS system) file. PRELIS, the pre-processor to LISREL, can read data files from a number of statistical programs, including SAS, Stata, and SPSS. To open the SPSS file *values.sav* saved in the C:\temp\CFA folder, go to **File** → **Import Data**. The **Open** dialog box opens. Change **Files of Type** to *SPSS Data File(*.sav)*, navigate to the correct folder, and click on *values.sav*.

Click **Open.** You will then be prompted to save the data as a .psf file. Name the file *values* and click **Save** to store it in the working directory. A spreadsheet with the raw data will display.



	privtown	govtresp	compete	homosex	abortion	euthanas
1	1.000	3.000	2.000	2.000	2.000	2.000
2	2.000	1.000	2.000	1.000	1.000	1.000
3	7.000	3.000	3.000	4.000	3.000	3.000
4	4.000	3.000	1.000	6.000	7.000	1.000
5	8.000	1.000	1.000	1.000	5.000	1.000
6	8.000	3.000	4.000	3.000	4.000	4.000
7	6.000	6.000	6.000	1.000	1.000	1.000
8	1.000	1.000	1.000	1.000	1.000	3.000
9	5.000	6.000	1.000	1.000	1.000	1.000
10	4.000	2.000	2.000	2.000	2.000	2.000
11	4.000	4.000	4.000	1.000	1.000	1.000
12	1.000	1.000	1.000	1.000	1.000	1.000
13	1.000	1.000	1.000	1.000	1.000	6.000
14	6.000	3.000	7.000	7.000	9.000	7.000
15	1.000	5.000	1.000	2.000	1.000	1.000

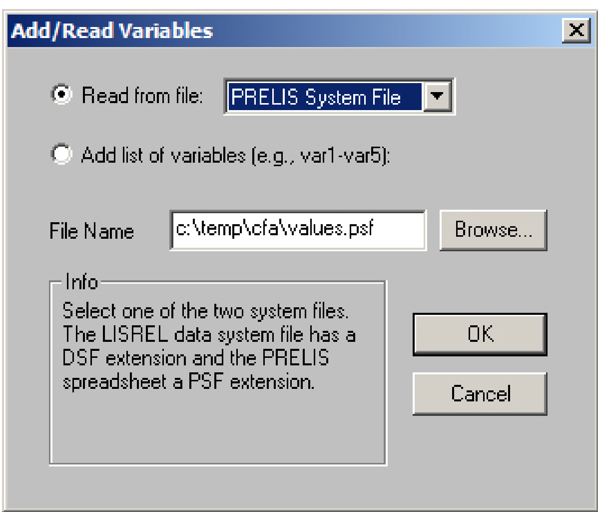
LISREL allows you to construct a path diagram of the model and then generate a PRELIS or LISREL program based on the path diagram. Go to **File** → **New** and choose **Path Diagram**. You will be immediately prompted to save the path diagram. Name the file *values* and click **Save** (the .pth extension will differentiate this file from the other files named *values* in the working directory). An empty window opens where you will eventually draw the diagram.

The next step is to name the variables that will be in the model. Go to **Setup** → **Title and Comments** to open the **Title and Comments** dialog box. Enter *First Two Factor Model* in the **Title** field and click **Next**.

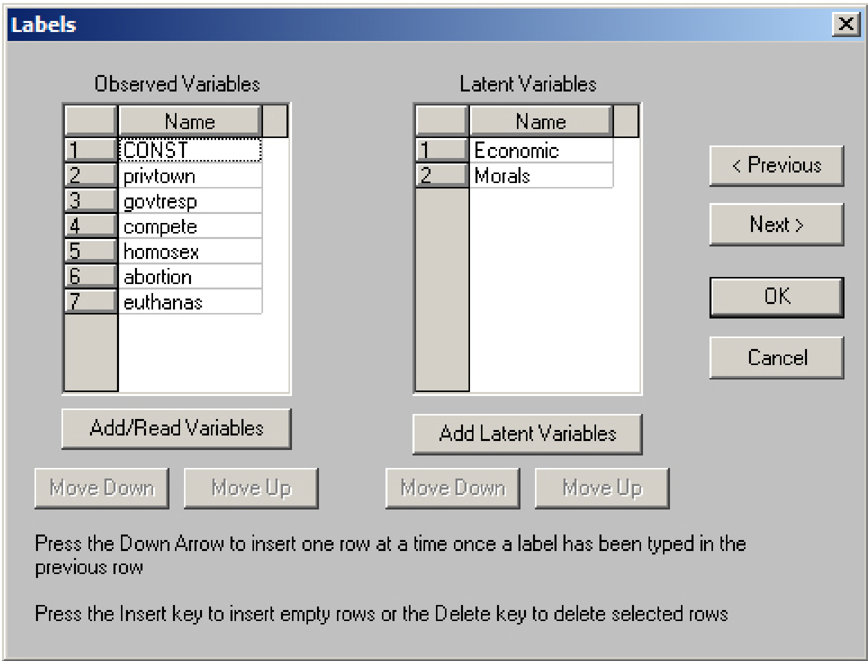
The **Group Names** box opens, which is used to label different groups when comparing models for multiple independent samples. Because we are interested only in the single sample of American respondents we can skip this box by clicking **Next**.

The **Labels** dialog box then opens, which is used to identify the latent and observed variables to be analyzed. Currently no variables have been selected. To choose variable names click on **Add/Read Variables**.

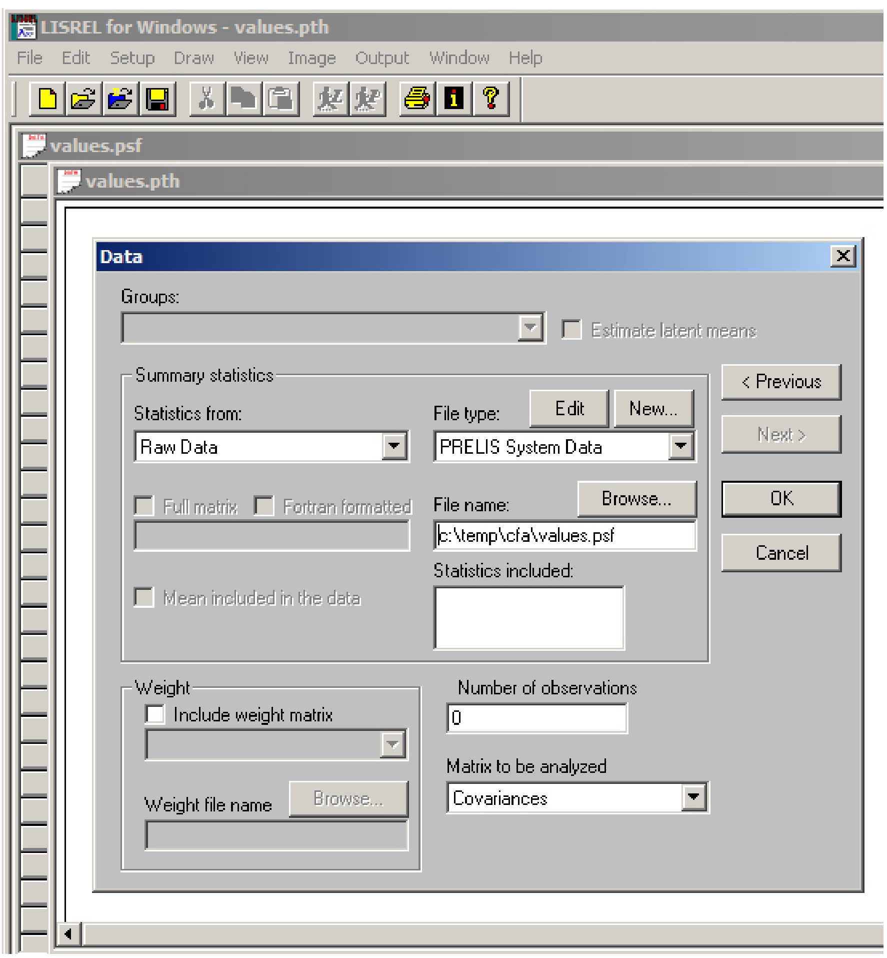
This opens a new dialog box used to locate the PRELIS system file. Verify that the **Read from file** radio button is chosen and pick **PRELIS System File** from the drop-down menu. Then click **Browse** to choose the PRELIS system file created earlier. Click **OK**.



The names of the observed variables are now listed in the **Labels** box. Note that the default variable CONST appears on the list of observed variables. To add the names of the latent variables click **Add Latent Variables**. Enter ECONOMIC in the box that opens. Repeat to enter the name of the second common factor MORALS. Click **OK**.



Click **Next**, and a final dialog box opens.

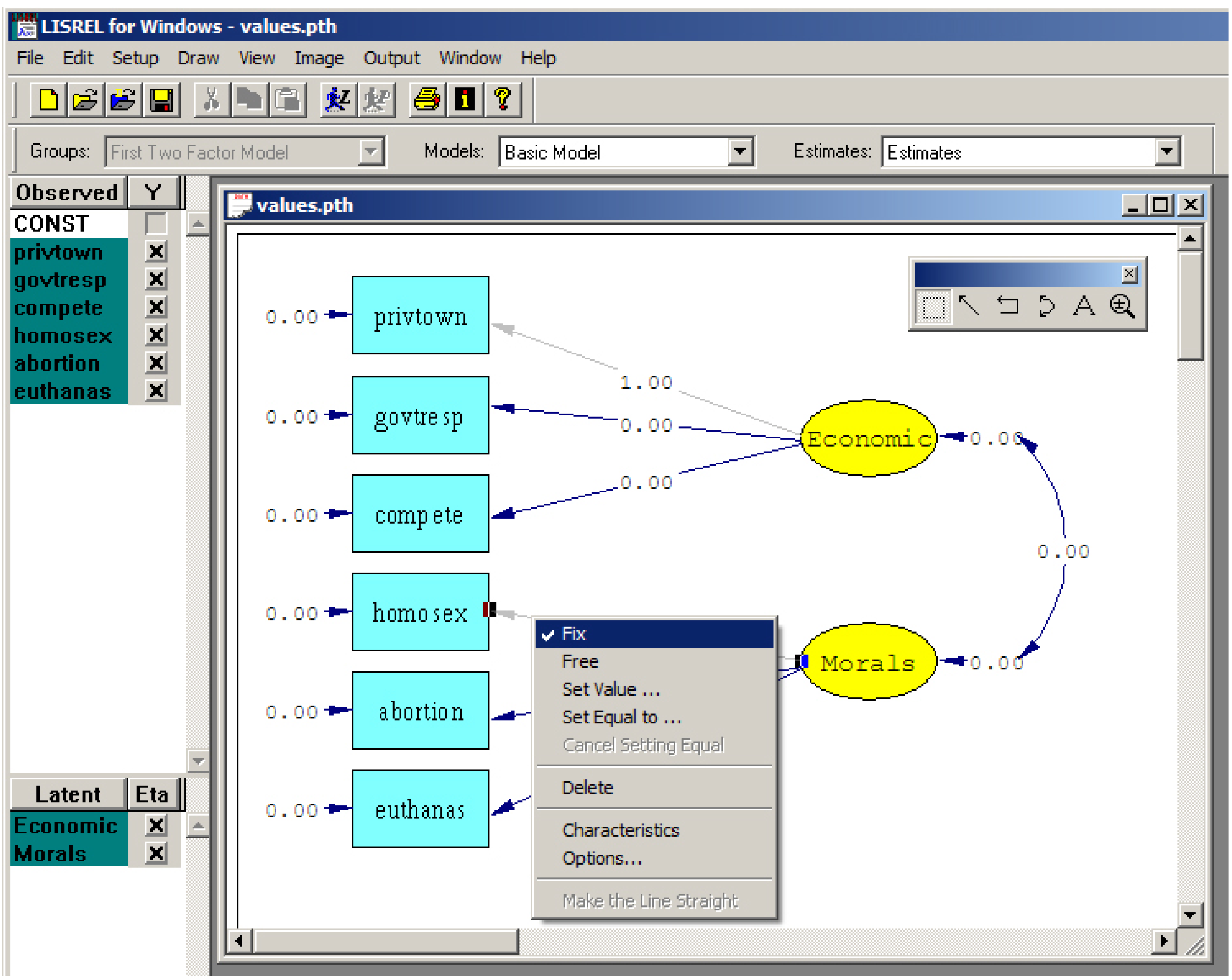


Raw data from a PRELIS system file will be analyzed. If desired, the data can be viewed and edited by clicking on the **Edit** button. Because this system file already contains information about the sample size it is not necessary to make further changes. Click **OK**.

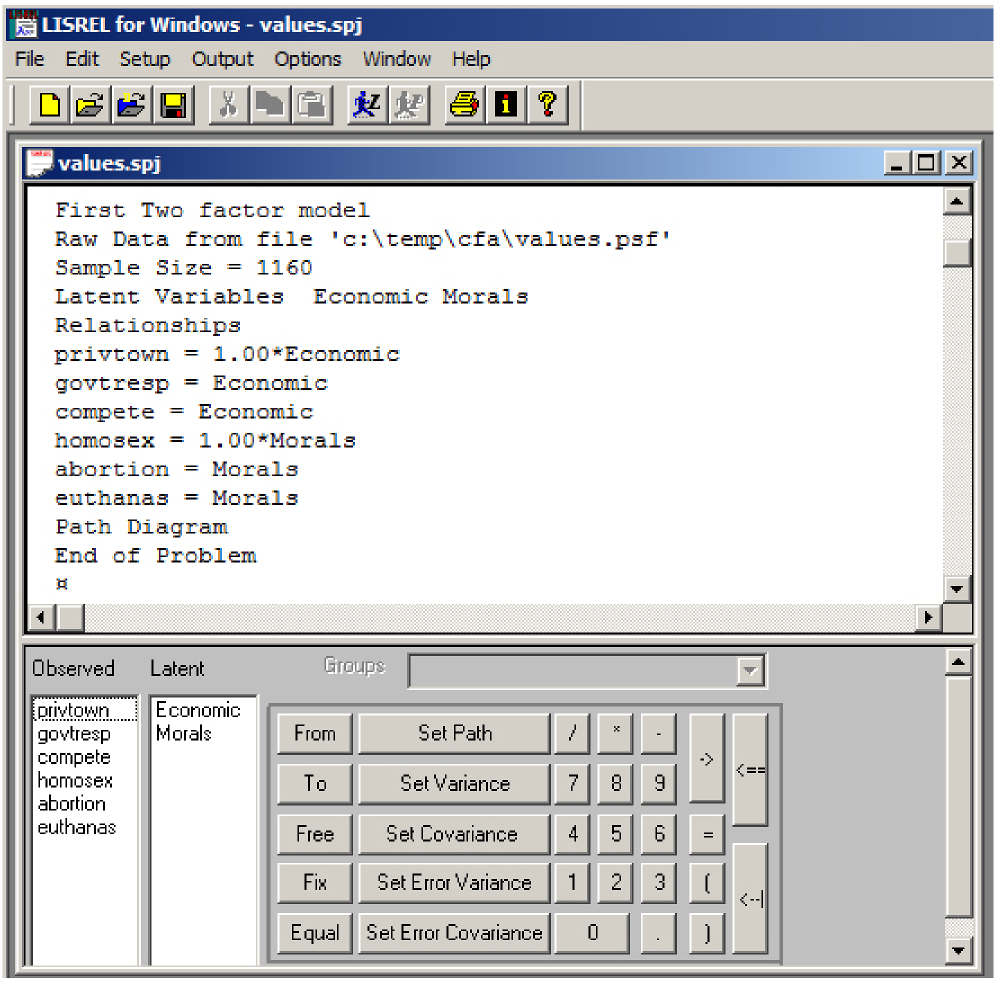
It is now possible to begin drawing the path diagram. The names of the observed and latent variables appear on the left side of the screen. Drag all of the observed variables to the drawing pad along with the latent variables ECONOMIC and MORALS.

Next click on the single-headed arrow on the tool bar and connect the ECONOMIC factor to PRIVTOWN, GOVTRESP, and COMPETE. Also draw arrows from MORALS to HOMOSEX, ABORTION, and EUTHANAS. Because the usual assumption is that the latent variables "cause" the observed variables, the arrows should point towards the six indicators. Finally, draw a two-headed arrow connecting each latent variable.

In LISREL, it is not necessary to draw the unique factors representing measurement error for each of the observed variables. LISREL includes these by default and automatically sets their scales by constraining the loadings to one. To set the scale of ECONOMIC, constrain the regression weight of the PRIVTOWN variable to one. Double-click on the line at the point where 0.00 appears and change the loading to 1.00. LISREL will not recognize this constraint, however, unless you then right-click on the loading and choose **Fix**. Do the same for the path connecting MORALS to HOMOSEX to set the metric for the second common factor.



The final step before estimation is to build from the path diagram the corresponding syntax LISREL uses for estimation. There are actually two languages that LISREL understands: LISREL syntax and SIMPLIS syntax. As its name suggests, SIMPLIS is more straightforward and easy to read than LISREL syntax. A SIMPLIS syntax file can be built from the path diagram by choosing **Setup → Build SIMPLIS syntax**. This opens an editor displaying the SIMPLIS commands needed to estimate the model.



Click the **Run LISREL** button  to begin estimation. Each time the **Run LISREL** button is clicked, a text output file is written to the working directory (extension .out) and a path diagram is presented. The text output contains unstandardized estimates, test statistics, goodness-of-fit statistics, modification indices, and other additional information; standardized estimates do not appear on the text output. It is always a good idea to inspect the output file for any error messages and, in some cases, warnings that a model may not be identified. For this model the output file is the following:

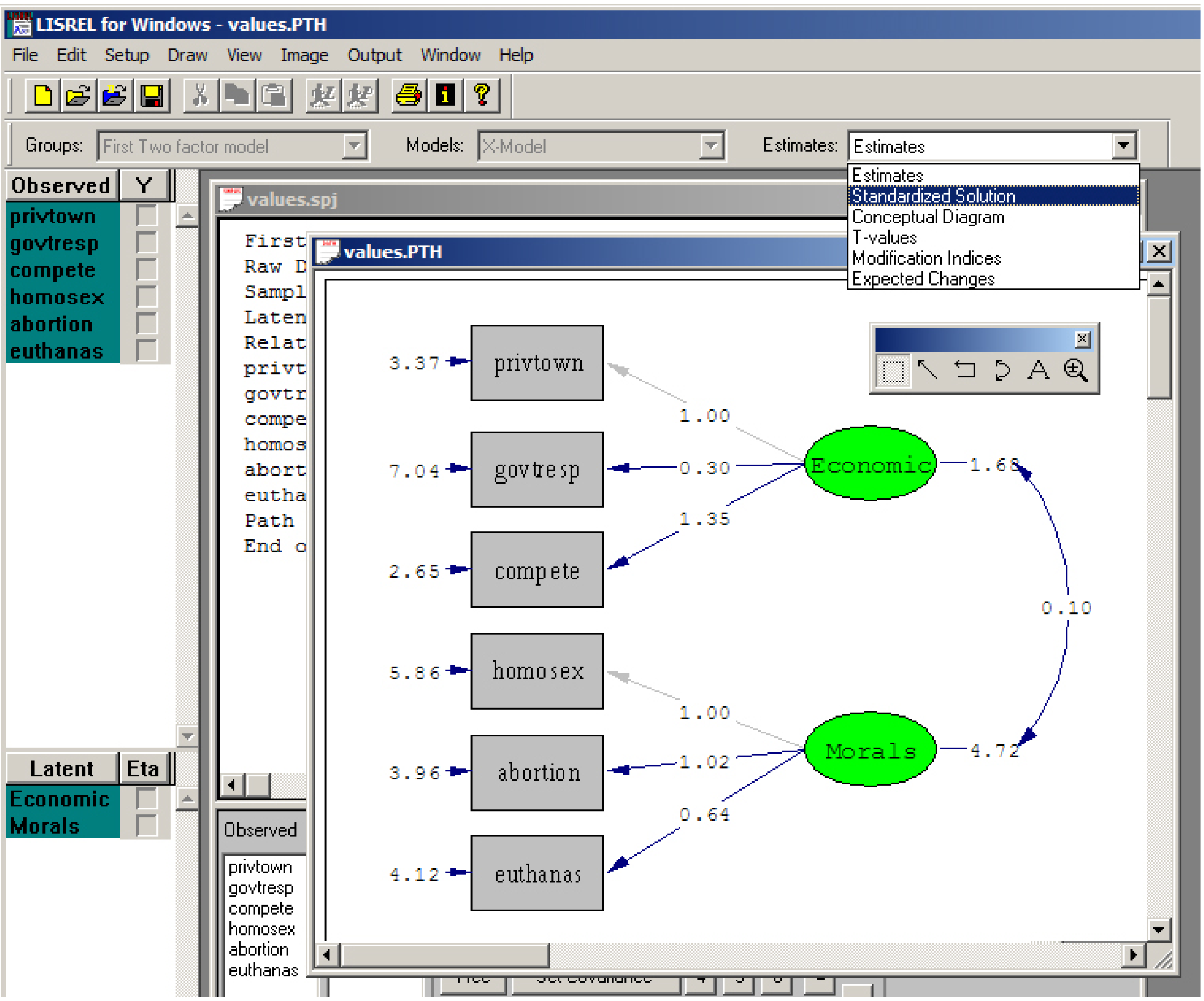
```
                                DATE: 11/18/2008
                                  TIME:  1:33
```

```
                         L I S R E L  8.80

                               BY

                   Karl G. Jöreskog & Dag Sörbom



              This program is published exclusively by
              Scientific Software International, Inc.
                 7383 N. Lincoln Avenue, Suite 100
                  Lincolnwood, IL 60712, U.S.A.
         Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
     Copyright by Scientific Software International, Inc., 1981-2006
      Use of this program is subject to the terms specified in the
                  Universal Copyright Convention.
                   Website: www.ssicentral.com

 The following lines were read from file c:\temp\cfa\values.spj

 First Two factor model
 Raw Data from file 'c:\temp\cfa\values.psf'
 Sample Size = 1160
 Latent Variables  Economic Morals
 Relationships
 privtown = 1.00*Economic
 govtresp = Economic
 compete = Economic
 homosex = 1.00*Morals
 abortion = Morals
 euthanas = Morals
 Path Diagram
 End of Problem

 Sample Size =  1160

 First Two factor model

        Covariance Matrix

           privtown   govtresp   compete   homosex  abortion  euthanas
           --------   --------   --------  --------  --------  --------
 privtown      5.05
 govtresp      0.51       7.20
  compete      2.27       0.67       5.71
  homosex     -0.22       1.39       0.01     10.57
 abortion     -0.11       1.01       0.06      4.81      8.85
 euthanas      0.14       0.72       0.38      3.00      3.05      6.03


 First Two factor model

 Number of Iterations =  5

 LISREL Estimates (Maximum Likelihood)

        Measurement Equations


 privtown = 1.00*Economic, Errorvar.= 3.37 , R² = 0.33
                                    (0.57)
                                     5.88

 govtresp = 0.30*Economic, Errorvar.= 7.04 , R² = 0.022
          (0.084)                  (0.30)
           3.63                     23.70

  compete = 1.35*Economic, Errorvar.= 2.65 , R² = 0.54
          (0.45)                   (1.02)
```

```
         3.01                  2.60

 homosex = 1.00*Morals, Errorvar.= 5.86 , R² = 0.45
                              (0.40)
                               14.49

abortion = 1.02*Morals, Errorvar.= 3.96 , R² = 0.55
        (0.076)               (0.37)
         13.39                 10.63

euthanas = 0.64*Morals, Errorvar.= 4.12 , R² = 0.32
        (0.046)               (0.22)
         13.73                 19.13


       Covariance Matrix of Independent Variables

         Economic    Morals
         --------   --------
Economic      1.68
           (0.58)
             2.91
  Morals      0.10      4.72
           (0.13)    (0.49)
             0.78      9.67


                    Goodness of Fit Statistics

                        Degrees of Freedom = 8
              Minimum Fit Function Chi-Square = 42.11 (P = 0.00)
      Normal Theory Weighted Least Squares Chi-Square = 41.53 (P = 0.00)
              Estimated Non-centrality Parameter (NCP) = 33.53
           90 Percent Confidence Interval for NCP = (17.01 ; 57.57)

                     Minimum Fit Function Value = 0.036
              Population Discrepancy Function Value (F0) = 0.029
           90 Percent Confidence Interval for F0 = (0.015 ; 0.050)
           Root Mean Square Error of Approximation (RMSEA) = 0.060
          90 Percent Confidence Interval for RMSEA = (0.043 ; 0.079)
             P-Value for Test of Close Fit (RMSEA < 0.05) = 0.16

               Expected Cross-Validation Index (ECVI) = 0.058
          90 Percent Confidence Interval for ECVI = (0.044 ; 0.079)
                      ECVI for Saturated Model = 0.036
                     ECVI for Independence Model = 0.83

     Chi-Square for Independence Model with 15 Degrees of Freedom = 949.26
                         Independence AIC = 961.26
                            Model AIC = 67.53
                          Saturated AIC = 42.00
                         Independence CAIC = 997.59
                           Model CAIC = 146.26
                          Saturated CAIC = 169.18

                       Normed Fit Index (NFI) = 0.96
                     Non-Normed Fit Index (NNFI) = 0.93
                  Parsimony Normed Fit Index (PNFI) = 0.51
                     Comparative Fit Index (CFI) = 0.96
                     Incremental Fit Index (IFI) = 0.96
                      Relative Fit Index (RFI) = 0.92

                          Critical N (CN) = 553.99


                   Root Mean Square Residual (RMR) = 0.41
                          Standardized RMR = 0.052
                     Goodness of Fit Index (GFI) = 0.99
                Adjusted Goodness of Fit Index (AGFI) = 0.97
               Parsimony Goodness of Fit Index (PGFI) = 0.38
```

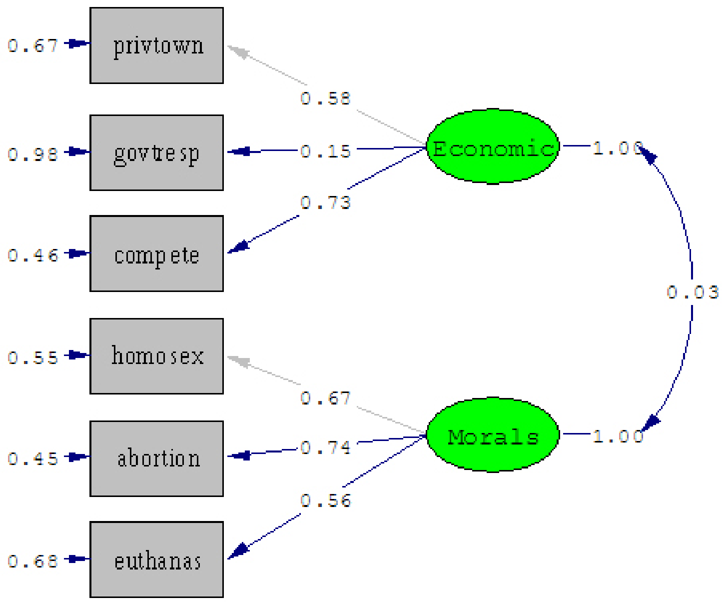
```
      The Modification Indices Suggest to Add the
 Path to  from      Decrease in Chi-Square    New Estimate
govtresp  Morals             31.6               0.24

The Modification Indices Suggest to Add an Error Covariance
 Between    and     Decrease in Chi-Square    New Estimate
compete   privtown           31.6              86.63
homosex   govtresp           12.4               0.76

                      Time used:    0.016 Seconds
```

The χ^2 statistic for model fit is 42.11, which is large enough to reject the null that the model is a good fit to the data. We report the Minimum Fit Function χ^2 in order to be consistent with the output from other software packages. The degrees of freedom is 8 = 21 -13; there are 13 unknown parameters including six λ_{ij}, six δ_i , and ϕ_{21}. The Root Mean Square Error of Approximation (RMSEA) is .060. Using a cut-off rule of .05, the RMSEA is too high to indicate a good fit. CFI of .96 indicates a moderate fit.



The text output presents unstandardized estimates and their standard errors. It is possible to ascertain the statistical significance of the estimates by comparing the unstandardized loadings displayed in the equations under the `Measurement Equations` heading in the output file with their standard errors displayed in parentheses. When the unstandardized loadings are at least twice the size of the standard errors the estimates are significant at the .05 level. In this case each of the unconstrained estimates is significant.

The unstandardized estimates also appear in the path diagram by default (see left diagram below). To view the standardized estimates choose **Standardized Solution** from the **Estimates** drop-down menu (see above screenshot). Note that factor loadings of PRIVTOWN and HOMOSEX are set 1 in the above diagram for unstandardized solutions, while variances of two factors are set 1 in the following diagram for standardized solutions.

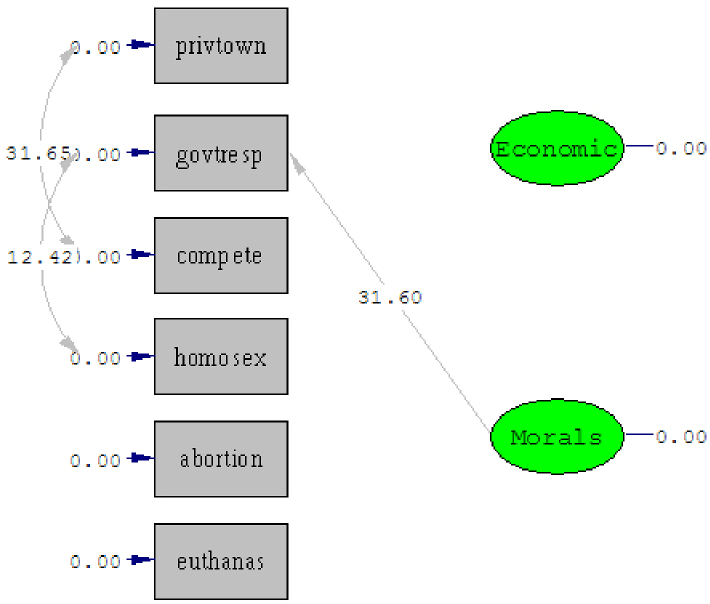


The standardized loadings represent the correlation between each observed variable and the corresponding factor. Considering first the indicators of ECONOMIC, they are .58 for PRIVOWN, .15 for GOVTRESP, and .73 for COMPETE. Considering the indicators of MORALS, the standardized loadings are .67 for HOMOSEX, .74 for ABORTION, and .56 for EUTHANAS.

A good deal of the variance in each observed variable, with the exception of GOVTRESP, is accounted for. The R^2 for PRIVTOWN is .33 (=$.58^2$); for COMPETE it is .54 (=$.73^2$); for HOMOSEX it is .45; for ABORTION it is .65; and for EUTHANAS it is .32. Only GOVTRESP, with its R^2 of .022, does not fit in well with the model. It may be the case that this survey question taps some kind of value dimension distinct from the economic dimension measured by the PRIVTOWN and COMPETE variables.

LISREL reports modification indices, both in the path diagram (by choosing **Modification Indices** from the **Estimation** menu) and in the output. These indices make suggestions about loosening certain model parameters in order to improve the overall model fit. As long as any decisions made on the basis of modification indices are theoretically meaningful and do not result in an unidentified model they can be helpful in improving model specification. Three suggestions are given in the output: add a path from GOVTRESP to MORALS; add error covariances between HOMOSEX and GOVTRESP, and between PRIVTOWN and COMPETE.

GOVTRESP appears to have something in common with the morality dimension, either by sharing measurement error with HOMOSEX and COMPETE or as a direct indicator of the latent morality dimension. The modification index 31.60 says that adding an arrow from MORALS to GOVTRESP will reduce χ^2 by 31.60. Because the standardized loading of GOVTRESP on ECONOMIC was so low, it is possible that the item is actually tapping a different values dimension.



The final model therefore adds a path from MORALS to GOVTRESP. This modification results in the following standardized solution:

```
Second Two factor model
Raw Data from file 'c:\Temp\cfa\values.psf'
Sample Size = 1160
Latent Variables  Economic Morals
Relationships
privtown = 1.00*Economic
govtresp = Economic Morals
compete = Economic
homosex = 1.00*Morals
abortion = Morals
euthanas = Morals
Path Diagram
End of Problem
```

The following output is selective, ignoring less informative parts.

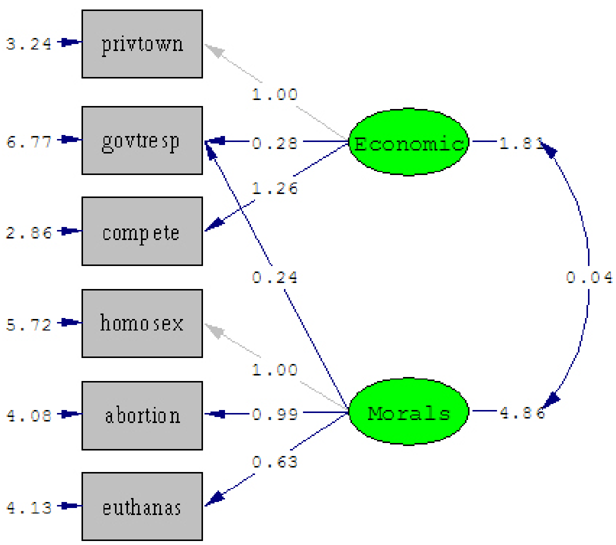
```
        Measurement Equations

 privtown = 1.00*Economic, Errorvar.= 3.24 , R² = 0.36
                                      (0.62)
                                       5.23

 govtresp = 0.28*Economic + 0.24*Morals, Errorvar.= 6.77 , R² = 0.060
          (0.082)        (0.043)               (0.29)
           3.46           5.56                  23.33
```

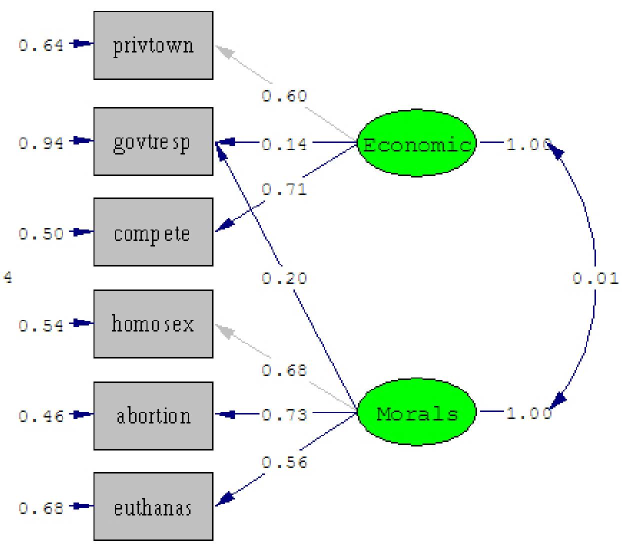
```
 compete = 1.26*Economic, Errorvar.= 2.86 , R² = 0.50
        (0.42)                 (0.96)
         2.98                   2.97

 homosex = 1.00*Morals, Errorvar.= 5.72 , R² = 0.46
                            (0.40)
                             14.33

abortion = 0.99*Morals, Errorvar.= 4.08 , R² = 0.54
        (0.072)              (0.36)
         13.83                11.46

euthanas = 0.63*Morals, Errorvar.= 4.13 , R² = 0.32
        (0.045)              (0.21)
         13.80                19.28


       Covariance Matrix of Independent Variables

         Economic    Morals
         --------   --------
Economic      1.81
           (0.63)
             2.89
  Morals      0.04      4.86
           (0.13)    (0.49)
             0.32      9.93


                      Goodness of Fit Statistics

                         Degrees of Freedom = 7
            Minimum Fit Function Chi-Square = 9.88 (P = 0.20)
     Normal Theory Weighted Least Squares Chi-Square = 9.89 (P = 0.19)
             Estimated Non-centrality Parameter (NCP) = 2.89
           90 Percent Confidence Interval for NCP = (0.0 ; 15.41)

                  Minimum Fit Function Value = 0.0085
            Population Discrepancy Function Value (F0) = 0.0025
           90 Percent Confidence Interval for F0 = (0.0 ; 0.013)
          Root Mean Square Error of Approximation (RMSEA) = 0.019
          90 Percent Confidence Interval for RMSEA = (0.0 ; 0.044)
            P-Value for Test of Close Fit (RMSEA < 0.05) = 0.98

               Expected Cross-Validation Index (ECVI) = 0.033
         90 Percent Confidence Interval for ECVI = (0.030 ; 0.043)
                     ECVI for Saturated Model = 0.036
                    ECVI for Independence Model = 0.83

    Chi-Square for Independence Model with 15 Degrees of Freedom = 949.26
                       Independence AIC = 961.26
                           Model AIC = 37.89
                         Saturated AIC = 42.00
                       Independence CAIC = 997.59
                          Model CAIC = 122.68
                        Saturated CAIC = 169.18

                      Normed Fit Index (NFI) = 0.99
                    Non-Normed Fit Index (NNFI) = 0.99
                 Parsimony Normed Fit Index (PNFI) = 0.46
                    Comparative Fit Index (CFI) = 1.00
                    Incremental Fit Index (IFI) = 1.00
                     Relative Fit Index (RFI) = 0.98

                        Critical N (CN) = 2167.38

                  Root Mean Square Residual (RMR) = 0.12
                        Standardized RMR = 0.018
                    Goodness of Fit Index (GFI) = 1.00
                Adjusted Goodness of Fit Index (AGFI) = 0.99
```

```
Parsimony Goodness of Fit Index (PGFI) = 0.33

        Time used:    0.000 Seconds
```

This model fits the data well. The χ^2 measure of model fit is 9.88, which is too small to reject the null hypothesis of a good fit (p<.20). Adding λ_{22} reduces degrees of freedom by 1 (df=7). The RMSEA has declined to .019, which is small enough to indicate a good fit, while CFI is almost (as oppose to numerically) 1.

The unconstrained loadings are all statistically significant at the .05 level, having estimates that are more than twice the size of their standard errors (see the output above and left diagram below). GOVTRESP continues to have a low correlation with the ECONOMIC factor (.14) and has a similarly low correlation with MORALS (.20) (see right diagram below). However, the remaining standardized loadings range from .56 (EUTHANAS) to .73 (ABORTION). In between are PRIVTOWN (.60), HOMOSEX (.68), and COMPETE (.71).





Despite receiving a path from both common factors GOVTRESP continues to have by far the smallest R^2 (.060). The remaining statistics are moderately well accounted for by the corresponding factors. The R^2 values are, in order of increasing magnitude, .32 for EUTHANAS, .36 for PRIVTOWN, .46 for HOMOSEX, .50 for COMPETE, and .54 for ABORTION. Finally, the correlation between ECONOMIC and MORALS is a negligible -.01 (right path diagram above) and their covariance is .04 (left diagram above).

The conclusion from this analysis is that two nearly orthogonal dimensions underlie the economic and moral values of American citizens. Additionally it is unclear whether the GOVTRESP item is tapping either dimension. Future surveys should incorporate more reliable measures of economic values.

3.3 Two Factor Model using Mplus

This section demonstrates how to estimate a confirmatory factor model using Mplus 5.2. Because the subsection covering Amos revealed that the single common factor model was a poor fit to the data, this section will also begin with the two factor model.

To launch Mplus from any Windows machine in the UITS Student Technology Centers go to **Start → All Programs → Departmentally Sponsored → Statistics-Math → Mplus5.2 → Mplus Editor**. This will open Mplus and display the program's built-in syntax editor.

Unlike Amos and LISREL, Mplus does not allow you to simply draw a path diagram and estimate the model; you must write the syntax yourself. The Language Generator under the Mplus menu, however, can make this task a little easier. Mplus reads only ASCII text files with free or fixed formatted; Mplus cannot directly read a SPSS data set. Nonetheless, Mplus is an extremely powerful program for estimating a much wider range of models than is possible with Amos and LISREL, and this example will only scratch the surface of what Mplus can do.

The first model to be examined consists of two common factors and the six observed indicators of economic and moral values. The Mplus syntax for estimating this model is the following:

```
TITLE:      First Two Factor Model;
DATA:       FILE IS c:\temp\cfa\values.dat;
VARIABLE:   NAMES ARE privtown govtresp compete
            homosex abortion euthanas;
MODEL:      economic BY privtown govtresp compete;
            morals BY homosex abortion euthanas;
OUTPUT:     STANDARDIZED;
            MODINDICES;
```

Save as an input file under the name *values1.inp* in the same folder as the *values.dat* file.

The TITLE statement provides a label for the particular analysis that will be run. The DATA statement specifies where the data file is located at. Absolute path names are only necessary if the syntax file is located in a directory different from where the data is saved. The VARIABLE statement provides names for the six observed variables in the raw data file in the order in which they appear. The MODEL statement specifies the particular model to be estimated. In this case ECONOMIC is assumed to cause the three observed variables PRIVTOWN, GOVTRESP, and COMPETE; and MORALS is assumed to cause HOMOSEX, ABORTION, and EUTHANAS. The OUTPUT statement requests that standardized parameter estimates and modification indices be included in the output file.

There are a few things to keep in mind when creating Mplus syntax. First, all commands end with a semicolon; omitting the semicolon will lead to error messages. Second, Mplus cannot read more than 80 characters in a line. One way to limit this problem is to use very short names for variables, such as x_1, x_2, y_1, y_2 (longer names are used here to be consistent with the Amos and LISREL examples later). Commands can take up more than one line, as the semicolon marks the command end. Finally, Mplus is not case sensitive; capital and lowercase letters can be used interchangeably.

It is also important to know that the default behavior for setting the scale of the common latent variable is to constrain the loading for the first variable (in this case PRIVTOWN and HOMOSEX) to one. This option can be overridden but will not be altered here to keep the example consistent with the Amos and LISREL examples above.

After entering the syntax and saving it as an Mplus input (.inp) file, estimate the model by clicking the **Run** button 

. This produces a text output (.out) file stored in the working directory with the results. For this model the output file looks like the following:

```
Mplus VERSION 5.1
MUTHEN & MUTHEN
11/17/2008   9:58 PM

INPUT INSTRUCTIONS

  TITLE:     First Two Factor Model;
  DATA:       FILE IS C:\Temp\CFA\values.dat;
  VARIABLE:  NAMES ARE privtown govtresp compete
              homosex abortion euthanas;
  MODEL:      economic BY privtown govtresp compete;
              morals BY homosex abortion euthanas;
  OUTPUT:    STANDARDIZED;
              MODINDICES;


INPUT READING TERMINATED NORMALLY

Two Factor Model 1;

SUMMARY OF ANALYSIS

Number of groups                                         1
Number of observations                                1160

Number of dependent variables                              6
Number of independent variables                            0
Number of continuous latent variables                        2

Observed dependent variables

  Continuous
   PRIVTOWN    GOVTRESP    COMPETE     HOMOSEX     ABORTION    EUTHANAS

Continuous latent variables
   ECONOMIC    MORALS

Estimator                                               ML
Information matrix                                OBSERVED
Maximum number of iterations                          1000
Convergence criterion                            0.500D-04
Maximum number of steepest descent iterations           20

Input data file(s)
  C:\Temp\CFA\values.dat

Input data format  FREE

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

        Value                        42.147
```

```
        Degrees of Freedom                     8
        P-Value                           0.0000

Chi-Square Test of Model Fit for the Baseline Model

        Value                            891.990
        Degrees of Freedom                    15
        P-Value                           0.0000

CFI/TLI

        CFI                                0.961
        TLI                                0.927

Loglikelihood

        H0 Value                      -16217.034
        H1 Value                      -16195.961

Information Criteria

        Number of Free Parameters             19
        Akaike (AIC)                   32472.068
        Bayesian (BIC)                 32568.135
        Sample-Size Adjusted BIC       32507.785
          (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                           0.061
        90 Percent C.I.                    0.043  0.079
        Probability RMSEA <= .05           0.147

SRMR (Standardized Root Mean Square Residual)

        Value                              0.046


MODEL RESULTS
                                                    Two-Tailed
                Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           1.000      0.000    999.000    999.000
    GOVTRESP           0.304      0.083      3.675      0.000
    COMPETE            1.350      0.433      3.120      0.002

 MORALS   BY
    HOMOSEX            1.000      0.000    999.000    999.000
    ABORTION           1.018      0.076     13.432      0.000
    EUTHANAS           0.637      0.046     13.725      0.000

 MORALS   WITH
    ECONOMIC           0.098      0.124      0.784      0.433

 Intercepts
    PRIVTOWN           3.553      0.066     53.876      0.000
    GOVTRESP           4.312      0.079     54.775      0.000
    COMPETE            3.440      0.070     49.048      0.000
    HOMOSEX            4.781      0.095     50.095      0.000
    ABORTION           4.352      0.087     49.845      0.000
    EUTHANAS           2.638      0.072     36.590      0.000

 Variances
    ECONOMIC           1.677      0.557      3.011      0.003
    MORALS             4.715      0.486      9.692      0.000

 Residual Variances
    PRIVTOWN           3.366      0.553      6.090      0.000
    GOVTRESP           7.034      0.297     23.685      0.000
    COMPETE            2.645      0.983      2.692      0.007
```

```
   HOMOSEX          5.851      0.403     14.527      0.000
   ABORTION         3.957      0.372     10.648      0.000
   EUTHANAS         4.119      0.215     19.116      0.000


STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                  Two-Tailed
                Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN         0.577      0.094      6.136      0.000
    GOVTRESP         0.147      0.038      3.825      0.000
    COMPETE          0.732      0.118      6.223      0.000

 MORALS   BY
    HOMOSEX          0.668      0.028     23.883      0.000
    ABORTION         0.743      0.029     26.028      0.000
    EUTHANAS         0.563      0.028     20.215      0.000

 MORALS   WITH
    ECONOMIC         0.035      0.044      0.782      0.434

 Intercepts
    PRIVTOWN         1.582      0.044     35.908      0.000
    GOVTRESP         1.608      0.044     36.171      0.000
    COMPETE          1.440      0.042     34.366      0.000
    HOMOSEX          1.471      0.042     34.720      0.000
    ABORTION         1.463      0.042     34.637      0.000
    EUTHANAS         1.074      0.037     29.136      0.000

 Variances
    ECONOMIC         1.000      0.000    999.000    999.000
    MORALS           1.000      0.000    999.000    999.000

 Residual Variances
    PRIVTOWN         0.667      0.108      6.157      0.000
    GOVTRESP         0.978      0.011     86.785      0.000
    COMPETE          0.464      0.172      2.690      0.007
    HOMOSEX          0.554      0.037     14.820      0.000
    ABORTION         0.448      0.042     10.541      0.000
    EUTHANAS         0.683      0.031     21.796      0.000


STDY Standardization

(skip output)


STD Standardization

(skip output)

R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    PRIVTOWN         0.333      0.108      3.068      0.002
    GOVTRESP         0.022      0.011      1.913      0.056
    COMPETE          0.536      0.172      3.111      0.002
    HOMOSEX          0.446      0.037     11.941      0.000
    ABORTION         0.552      0.042     13.014      0.000
    EUTHANAS         0.317      0.031     10.107      0.000


QUALITY OF NUMERICAL RESULTS

     Condition Number for the Information Matrix              0.720E-03
```

```
      (ratio of smallest to largest eigenvalue)

MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index    10.000

                       M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

BY Statements

MORALS   BY GOVTRESP      31.631     0.240      0.522       0.195

WITH Statements

COMPETE  WITH PRIVTOWN    31.719    86.626     86.626      29.028
HOMOSEX  WITH GOVTRESP    12.426     0.760      0.760       0.118


    Beginning Time:  21:58:43
       Ending Time:  21:58:43
      Elapsed Time:  00:00:00


MUTHEN & MUTHEN
3463 Stoner Ave.
Los Angeles, CA  90066

Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support: Support@StatModel.com

Copyright (c) 1998-2008 Muthen & Muthen
```

The overall model fit is not great, with a χ^2 statistic of 42.147 (df=8) large enough to reject the null of a good fit. The RMSEA (Root Mean Square Error of Approximation) is .061, which is higher than the cutoff value of .05 (significance level) chosen to indicate a good fit. CFI (Comparative Fit Index) is .961.

Under the `MODEL RESULTS` heading the unstandardized loadings appear along with standard errors, ratios of the estimates to their standard errors, and p-values. The ratios under `Est./S.E.` column are test statistics that are equivalent to *z* scores. If the absolute value of the number in this column is greater than 1.96, the estimate can be interpreted as significant at the .05 level. Alternatively, if a p-value is smaller than .05, you may reject the null hypothesis at the same significance level that the parameter is zero. In this case all of the unconstrained loading estimates are significant.

Standardized factor loadings are presented under the `STANDIARDIZED MODEL RESULTS` heading.[6] ABORTION has the highest standardized factor loading .743; ABORTION appears to be a reliable indicator of moral values. By contrast, GOVTRESP has the lowest .147, suggesting that it is an unreliable indicator of economic values. Unstandardized and standardized delta values appear under `Residual Variances`.

The squared multiple correlations under `R-SQUARE` provide information on how much variance the common factors account for in the observed variables. R^2 is a standardized factor loading

[6] The previous version presents standardized factor loadings under the `StdYX` column of the `MODEL RESULTS`.

squared; for example, .535824=.732^2 for COMPETE. Economic values, for example, explains 53.6 percent of the total variation in COMPETE and moral values account for 55.2 percent of variance of ABORTION. The low R^2 of .022 suggests that economic values can explain only a small portion of variation of GOVTRESP.

The correlation between the two common factors (economic and moral values) is a very small .035 ($p<.434$), and the covariance estimate of .098 is not statistically discernable from zero ($p<.433$). You may find these statistics appear respectively on standardized and unstandardized results. Two factors do not appear to be closely related each other. The variances of two common factors on unstandardized results are 1.677 and 4.714, respectively. Note that standardized variances of common factors are set 1.

The `MODINDICES` of OUTPUT command produces model modification indices and gives you some hints about model specification. Under the `MODEL MODIFICATION INDICES` heading Mplus makes three suggestions: 1) adding covariance between COMPETE and PRIVTOWN will reduce χ^2 by 31.719, 2) adding a path from the MORALS latent variable to GOVTRESP by 31.631, and 3) adding a covariance between HOMOSEX and GOVTRESP by 12.426. The GOVTRESP item has something in common with the morality dimension, either by sharing measurement error with the HOMOSEX variable or as a direct indicator of the latent morality dimension. Because the standardized loading of GOVTRESP on ECONOMIC was so low, it is possible that the item is actually tapping a different values dimension. Thus the second suggestion makes theoretical sense and will be estimated.

Add GOVTRESP to moral values in the MODEL command and then save the input file.

```
TITLE:        Second Two Factor Model;
DATA:         FILE IS values.dat;
VARIABLE:     NAMES ARE privtown govtresp compete
              homosex abortion euthanas;
MODEL:        economic BY privtown govtresp compete;
              morals BY homosex abortion euthanas govtresp;
OUTPUT:       STANDARDIZED;
              MODINDICES;
```

Mplus returns the following output. Pay attention to the parts in red. Note that the degrees of freedom decrease from 8 to 7.

```
Mplus VERSION 5.1
MUTHEN & MUTHEN
11/17/2008  11:38 PM

INPUT INSTRUCTIONS

(skip output)


INPUT READING TERMINATED NORMALLY

(skip output)


THE MODEL ESTIMATION TERMINATED NORMALLY
```

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit

        Value                              9.893
        Degrees of Freedom                     7
        P-Value                           0.1947

Chi-Square Test of Model Fit for the Baseline Model

        Value                            891.990
        Degrees of Freedom                    15
        P-Value                           0.0000

CFI/TLI

        CFI                               0.997
        TLI                               0.993

Loglikelihood

        H0 Value                      -16200.907
        H1 Value                      -16195.961

Information Criteria

        Number of Free Parameters             20
        Akaike (AIC)                   32441.814
        Bayesian (BIC)                 32542.938
        Sample-Size Adjusted BIC       32479.411
          (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                           0.019
        90 Percent C.I.                    0.000  0.044
        Probability RMSEA <= .05           0.985

SRMR (Standardized Root Mean Square Residual)

        Value                              0.016


MODEL RESULTS

                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           1.000      0.000    999.000    999.000
    GOVTRESP           0.285      0.082      3.488      0.000
    COMPETE            1.256      0.442      2.839      0.005

 MORALS   BY
    HOMOSEX            1.000      0.000    999.000    999.000
    ABORTION           0.991      0.072     13.787      0.000
    EUTHANAS           0.627      0.046     13.662      0.000
    GOVTRESP           0.239      0.043      5.610      0.000

 MORALS   WITH
    ECONOMIC           0.043      0.136      0.317      0.752

 Intercepts
    PRIVTOWN           3.553      0.066     53.876      0.000
    GOVTRESP           4.312      0.079     54.775      0.000
    COMPETE            3.440      0.070     49.048      0.000
    HOMOSEX            4.781      0.095     50.096      0.000
    ABORTION           4.352      0.087     49.845      0.000
    EUTHANAS           2.638      0.072     36.590      0.000
```

```
 Variances
    ECONOMIC           1.806      0.654      2.762      0.006
    MORALS             4.855      0.490      9.906      0.000

 Residual Variances
    PRIVTOWN           3.238      0.648      4.998      0.000
    GOVTRESP           6.760      0.290     23.312      0.000
    COMPETE            2.856      1.007      2.836      0.005
    HOMOSEX            5.711      0.400     14.269      0.000
    ABORTION           4.078      0.356     11.465      0.000
    EUTHANAS           4.123      0.213     19.312      0.000


STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           0.598      0.107      5.608      0.000
    GOVTRESP           0.143      0.038      3.785      0.000
    COMPETE            0.707      0.125      5.658      0.000

 MORALS   BY
    HOMOSEX            0.678      0.028     24.613      0.000
    ABORTION           0.734      0.028     26.548      0.000
    EUTHANAS           0.562      0.028     20.437      0.000
    GOVTRESP           0.196      0.034      5.810      0.000

 MORALS   WITH
    ECONOMIC           0.015      0.047      0.312      0.755

 Intercepts
    PRIVTOWN           1.582      0.044     35.908      0.000
    GOVTRESP           1.608      0.044     36.171      0.000
    COMPETE            1.440      0.042     34.366      0.000
    HOMOSEX            1.471      0.042     34.721      0.000
    ABORTION           1.463      0.042     34.637      0.000
    EUTHANAS           1.074      0.037     29.136      0.000

 Variances
    ECONOMIC           1.000      0.000    999.000    999.000
    MORALS             1.000      0.000    999.000    999.000

 Residual Variances
    PRIVTOWN           0.642      0.128      5.028      0.000
    GOVTRESP           0.940      0.017     55.747      0.000
    COMPETE            0.501      0.177      2.836      0.005
    HOMOSEX            0.540      0.037     14.476      0.000
    ABORTION           0.461      0.041     11.365      0.000
    EUTHANAS           0.684      0.031     22.099      0.000


STDY Standardization

(skip output)


STD Standardization

(skip output)


R-SQUARE

    Observed                                        Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    PRIVTOWN           0.358      0.128      2.804      0.005
```

```
    GOVTRESP          0.060      0.017      3.538      0.000
    COMPETE           0.499      0.177      2.829      0.005
    HOMOSEX           0.460      0.037     12.307      0.000
    ABORTION          0.539      0.041     13.274      0.000
    EUTHANAS          0.316      0.031     10.218      0.000


QUALITY OF NUMERICAL RESULTS

     Condition Number for the Information Matrix           0.646E-03
       (ratio of smallest to largest eigenvalue)


MODEL MODIFICATION INDICES

Minimum M.I. value for printing the modification index   10.000

                       M.I.     E.P.C.  Std E.P.C.  StdYX E.P.C.

No modification indices above the minimum value.

(skip output)
```

Skipped are some parts of the output, which are redundant or less informative.
The χ^2 test yields a value of 9.893 (df=7), which does not reject the null hypothesis of an overall good fit (p<.1947). The RMSEA .019 (p<.985) and CFI .997 indicate that this model fits the data well.

All unstandardized and standardized factor loadings are statistically significant (p<.000). Individual standardized loading values remain almost unchanged. The factor loading .196 suggests that GOVTRESP manifests both economic and moral values significantly. The covariance and correlation of two factors are respectively .043 and .015, which are smaller than those of model 1. Adding a relation between GOVTRESP and moral values appears to make the model fit better. Model modification indices do not suggest any change in this model.

The conclusion from this analysis is that two nearly orthogonal dimensions underlie the economic and moral values of American citizens. It is not clear, however, whether the GOVTRESP item is tapping either dimension. Future surveys should incorporate more reliable measures of economic values.

3.4 Two Factor Model using SAS CALIS

This section demonstrates how to estimate a confirmatory factor model using the SAS CALIS procedure. This procedure supports various linear models including structural equation model, explanatory and confirmatory factor analysis, multivariate linear regression, path analysis, simultaneous equation model, and canonical correlation. The following CALIS procedure fits the first two factor model discussed above.

```
LIBNAME cfa 'c: \temp\cfa';

DATA cfa.values;
      INFILE 'c: \temp\cfa\values.dat';
      INPUT privtown govtresp compete homosex abortion euthanas;
RUN;
```

The LIBNAME statement above defines a library `cfa` that refers a physical space *c:\temp\cfa*; SAS looks for data sets in the directory once the library `cfa` is specified. The DATA step above reads six variables from an ASCII text file *values.dat* and save them into a SAS data set *values.sas7bdat* in the library *cfa*.

```
ODS HTML;

PROC CALIS DATA=cfa.values METHOD=ML PALL;
VAR privtown govtresp compete homosex abortion euthanas;
LINEQS      privtown = 1.0 f1 + e1,
      govtresp = lambda2 f1 + e2,
      compete = lambda3 f1 + e3,
      homosex = 1.0 f2 + e4,
      abortion = lambda5 f2 + e5,
      euthanas = lambda6 f2 + e6;
STD   f1-f2 = phi1-phi2,
      e1-e6 = theta1-theta6;
COV   f1-f2 = phi3;
RUN;

ODS HTML CLOSE;
```

PROC CALIS reads a data set *values.sas7bdat* from a library *sas* (c:\temp\cfa). The METHOD=ML uses the maximum likelihood method and the PALL option reports all possible output. The VAR statement lists the variables used. The LINEQS statement specifies relationships between latent and manifest variables in equation form. Note that *lambda1* and *lambda4* are set 1 for model identification. STD and COV statements are place where standard deviations and covariances are defined. Finally, ODS HTML redirects the SAS output into the HTML format.

The following is selected from the SAS output.

```
The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation
```

Fit Function	0.0363

Goodness of Fit Index (GFI)	0.9882
GFI Adjusted for Degrees of Freedom (AGFI)	0.9690
Root Mean Square Residual (RMR)	0.0519
Parsimonious GFI (Mulaik, 1989)	0.5270
Chi-Square	42.1104
Chi-Square DF	8
Pr > Chi-Square	<.0001
Independence Model Chi-Square	891.22
Independence Model Chi-Square DF	15
RMSEA Estimate	0.0607
RMSEA 90% Lower Confidence Limit	0.0434
RMSEA 90% Upper Confidence Limit	0.0793
ECVI Estimate	0.0589
ECVI 90% Lower Confidence Limit	0.0445
ECVI 90% Upper Confidence Limit	0.0798
Probability of Close Fit	0.1472
Bentler's Comparative Fit Index	0.9611
Elliptic Corrected Chi-Square	38.0261
Pr > Elliptic Corrected Chi-Square	<.0001
Normal Theory Reweighted LS Chi-Square	41.5308
Akaike's Information Criterion	26.1104
Bozdogan's (1987) CAIC	-22.3390
Schwarz's Bayesian Criterion	-14.3390
McDonald's (1989) Centrality	0.9854
Bentler & Bonett's (1980) Non-normed Index	0.9270
Bentler & Bonett's (1980) NFI	0.9527
James, Mulaik, & Brett (1982) Parsimonious NFI	0.5081
Z-Test of Wilson & Hilferty (1931)	4.6039
Bollen (1986) Normed Index Rho1	0.9114
Bollen (1988) Non-normed Index Delta2	0.9614
Hoelter's (1983) Critical N	428

Manifest Variable Equations with Estimates

```
privtown = 1.0000    f1       + 1.0000    e1
govtresp = 0.2546  * f1       + 1.0000    e2
Std Err    0.0701    lambda2
t Value    3.6310
compete  = 1.2699  * f1       + 1.0000    e3
Std Err    0.4224    lambda3
t Value    3.0064
homosex  = 1.0000    f2       + 1.0000    e4
abortion = 1.1127  * f2       + 1.0000    e5
Std Err    0.0831    lambda5
t Value    13.3876
euthanas = 0.8426  * f2       + 1.0000    e6
Std Err    0.0614    lambda6
t Value    13.7307
```

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
f1	phi1	0.33257	0.11439	2.91
f2	phi2	0.44621	0.04613	9.67
e1	theta1	0.66743	0.11357	5.88
e2	theta2	0.97844	0.04129	23.70
e3	theta3	0.46369	0.17864	2.60
e4	theta4	0.55379	0.03822	14.49
e5	theta5	0.44750	0.04211	10.63
e6	theta6	0.68319	0.03571	19.13

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f1	f2	phi3	0.01336	0.01721	0.78

Manifest Variable Equations with Standardized Estimates

```
privtown  = 0.5767    f1      + 0.8170    e1
govtresp  = 0.1468  * f1      + 0.9892    e2
                      lambda2
compete   = 0.7323  * f1      + 0.6809    e3
```

```
                    lambda3
homosex  = 0.6680   f2       + 0.7442    e4
abortion = 0.7433 * f2       + 0.6690    e5
                    lambda5
euthanas = 0.5629 * f2       + 0.8266    e6
                    lambda6
```

```
               Squared Multiple Correlations
     Variable   Error Variance   Total Variance   R-Square
 1   privtown  0.66743          1.00000          0.3326
 2   govtresp  0.97844          1.00000          0.0216
 3    compete  0.46369          1.00000          0.5363
 4    homosex  0.55379          1.00000          0.4462
 5   abortion  0.44750          1.00000          0.5525
 6   euthanas  0.68319          1.00000          0.3168
```

```
    Correlations Among Exogenous Variables
  Var1    Var2     Parameter        Estimate
   f1      f2         phi3     0.03469
```

```
                 Predicted Moments of
                   Latent Variables
                  f1                    f2
 f1  0.33257                 0.01336
 f2  0.01336                 0.44621
```

This model has a large χ^2 42.1104 (p<.0001), RMSEA .0607 (p<.1472), and CFI .9611, which indicates a poor fit. This χ^2 is slightly smaller than 42.147 that Mplus produced.

Unstandardized factor loadings of SAS CALIS are different from those Mplus returned, although z scores are very similar. For example, the factor loading of GOVTRESP is .2546 and its test statistic is 3.6320 (=.2546/.0701) in SAS CALIS. Corresponding statistics were respectively .304 and 3.675 (=.304/.083) in Mplus. Similarly, the covariance of two factors is .01336 and its test statistic is .78 (=.01336/.01721); Mplus reported .098 and .784, respectively.

However, standardized factor loadings, R^2, and correlation of two factors are almost the same as corresponding statistics that Mplus computed. For instance, SAS CALIS and Mplus respectively report the factor loading .7433 and .743 for ABORTION. R^2 are .5525 (=$.7433^2$) and .552, respectively. The correlation of two factors is .03469 in SAS CALIS and .035 in Mplus.

The second two factor model is estimated by the following SAS codes. Note that `lambda7 f2` was added to the GOVTRESP equation.

```
ODS HTML;

PROC CALIS DATA=sas.values METHOD=ML PALL;
VAR privtown govtresp compete homosex abortion euthanas;
LINEQS       privtown = 1.0 f1 + e1,
             govtresp = lambda2 f1 + lambda7 f2 + e2,
             compete = lambda3 f1 + e3,
             homosex = 1.0 f2 + e4,
             abortion = lambda5 f2 + e5,
             euthanas = lambda6 f2 + e6;
STD          f1-f2 = phi1-phi2,
             e1-e6 = theta1-theta6;
```

```
COV          f1-f2 = phi3;
RUN;

ODS HTML CLOSE;
```

The output is similar to that of the first two factor model but you need to pay attention to *lambda7* in the GOVTRESP equation, which represents the factor loading of ECONOMIC.

```
The CALIS Procedure
Covariance Structure Analysis: Maximum Likelihood Estimation
```

Fit Function	0.0085
Goodness of Fit Index (GFI)	0.9972
GFI Adjusted for Degrees of Freedom (AGFI)	0.9915
Root Mean Square Residual (RMR)	0.0178
Parsimonious GFI (Mulaik, 1989)	0.4653
Chi-Square	9.8844
Chi-Square DF	7
Pr > Chi-Square	0.1952
Independence Model Chi-Square	891.22
Independence Model Chi-Square DF	15
RMSEA Estimate	0.0189
RMSEA 90% Lower Confidence Limit	.
RMSEA 90% Upper Confidence Limit	0.0436
ECVI Estimate	0.0328
ECVI 90% Lower Confidence Limit	.
ECVI 90% Upper Confidence Limit	0.0437
Probability of Close Fit	0.9847
Bentler's Comparative Fit Index	0.9967
Elliptic Corrected Chi-Square	8.9257
Pr > Elliptic Corrected Chi-Square	0.2580
Normal Theory Reweighted LS Chi-Square	9.8904
Akaike's Information Criterion	-4.1156
Bozdogan's (1987) CAIC	-46.5088
Schwarz's Bayesian Criterion	-39.5088
McDonald's (1989) Centrality	0.9988
Bentler & Bonett's (1980) Non-normed Index	0.9929
Bentler & Bonett's (1980) NFI	0.9889
James, Mulaik, & Brett (1982) Parsimonious NFI	0.4615
Z-Test of Wilson & Hilferty (1931)	0.8623
Bollen (1986) Normed Index Rho1	0.9762
Bollen (1988) Non-normed Index Delta2	0.9967
Hoelter's (1983) Critical N	1651

```
Manifest Variable Equations with Estimates
privtown = 1.0000     f1      + 1.0000    e1
govtresp = 0.2387   * f1      + 0.2893  * f2      + 1.0000    e2
Std Err    0.0691     lambda2   0.0520    lambda7
t Value    3.4558               5.5599
compete  = 1.1809   * f1      + 1.0000    e3
Std Err    0.3967     lambda3
t Value    2.9770
homosex  = 1.0000     f2      + 1.0000    e4
abortion = 1.0828   * f2      + 1.0000    e5
Std Err    0.0783     lambda5
t Value    13.8252
euthanas = 0.8295   * f2      + 1.0000    e6
Std Err    0.0601     lambda6
t Value    13.8008
```

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
f1	phi1	0.35809	0.12399	2.89
f2	phi2	0.45951	0.04626	9.93
e1	theta1	0.64191	0.12276	5.23
e2	theta2	0.94033	0.04030	23.33
e3	theta3	0.50067	0.16838	2.97

```
           Variances of Exogenous Variables
 Variable   Parameter   Estimate   Standard   t Value
                                    Error
    e4       theta4   0.54049    0.03772    14.33
    e5       theta5   0.46127    0.04025    11.46
    e6       theta6   0.68382    0.03548    19.28

          Covariances Among Exogenous Variables
 Var1   Var2   Parameter   Estimate   Standard   t Value
                                       Error
  f1     f2       phi3    0.00589    0.01814    0.32

Manifest Variable Equations with Standardized Estimates
privtown = 0.5984    f1       + 0.8012    e1
govtresp = 0.1428 *  f1       + 0.1961 *  f2       + 0.9697    e2
                     lambda2              lambda7
compete  = 0.7066 *  f1       + 0.7076    e3
                     lambda3
homosex  = 0.6779    f2       + 0.7352    e4
abortion = 0.7340 *  f2       + 0.6792    e5
                     lambda5
euthanas = 0.5623 *  f2       + 0.8269    e6
                     lambda6

                Squared Multiple Correlations
     Variable   Error Variance   Total Variance   R-Square
 1   privtown  0.64191          1.00000          0.3581
 2   govtresp  0.94033          1.00000          0.0597
 3    compete  0.50067          1.00000          0.4993
 4    homosex  0.54049          1.00000          0.4595
 5   abortion  0.46127          1.00000          0.5387
 6   euthanas  0.68382          1.00000          0.3162

      Correlations Among Exogenous Variables
   Var1      Var2        Parameter        Estimate
    f1        f2           phi3       0.01453

  Predicted Moments of
    Latent Variables
           f1         f2
 f1  0.35809    0.00589
 f2  0.00589    0.45951
```

This second model has a smaller χ^2 9.8844 ($p<.1952$) and RMSEA .0189 ($p<.0328$), which do not reject the null hypothesis at the .05 level. CFI .9967 also indicates a good fit. Government responsibility (GOVTRESP) is significantly explained by both economic and moral values.

As discussed above, unstandardized factor loadings of SAS CALIS are different from those of Mplus, but their test statistics are similar. The covariance of two factors is .00589, slightly larger than .043 in Mplus, but its test statistic is .32 (=.00589/.01814) that is very close to .317 Mplus returned. Standardardized factor loadings and correlation of two factors reported by both software packages are virtually same. SAS CALIS reports .1961 as a standardized factor loading of moral values on GOVTRESP, which is the same as .196 that Mplus returned.

3.5 Two Factor Model using Amos

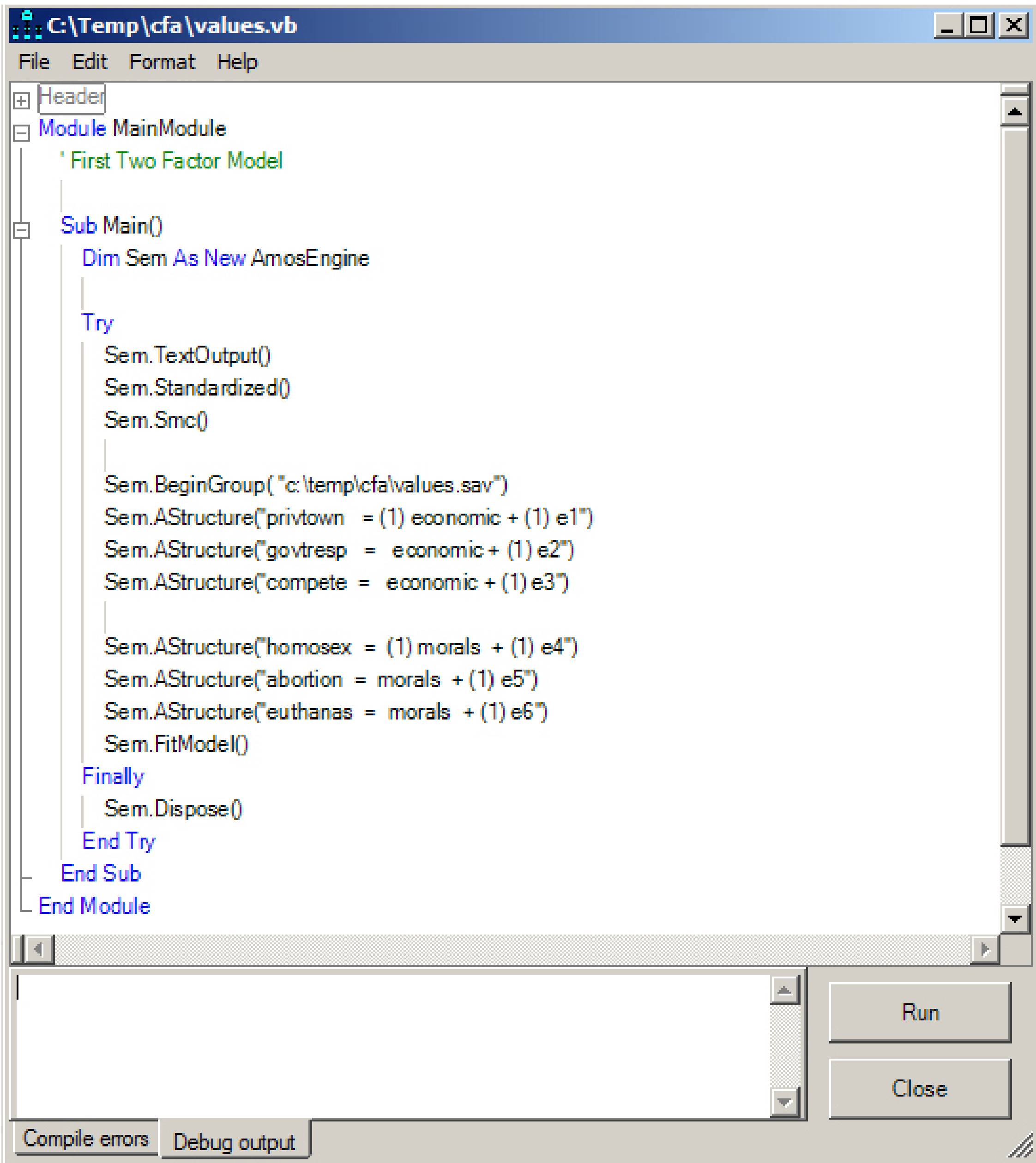
Amos consists of several applications including Amos Graphics, Program Editor, File manager, and Seed Manager. Most people begin with Amos Graphics to fit a model because it provides a point-and-click GUI environment, which makes it easy for beginners to specify models. This section, however, uses Program Editor to take advantage of its simplicity at the expense of burdensome programming. See Appendix if you want to estimate a model using Amos Graphics.

Amos Program Editor can be launched from any computer running Windows in the UITS Student Technology Centers by going to **Start** → **All Programs** → **Departmentally Sponsored** → **Statistics-Math** → **Amos 16** → **Program Editor**. At Program Editor, click **File** → **Open** to open a dialog box, and then select a sample program *Ex08.vb* listed in the default directory *Examples*. The following screen will display.

```
C:\Program Files\AMOS 16.0\Examples\Ex08.vb
File  Edit  Format  Help
Header
Module MainModule
    ' Example 8:
    ' Factor analysis
    '
    ' Holzinger and Swineford (1939) Grant-White sample.
    ' Intelligence factor study. Raw data of 73 female
    ' students from the Grant-White high school, Chicago.
    '
    Sub Main()
        Dim Sem As New AmosEngine

        Try
            Sem.TextOutput()
            Sem.Standardized()
            Sem.Smc()

            Sem.BeginGroup(AmosEngine.AmosDir & "Examples\Grnt_fem.sav")
            Sem.AStructure("visperc  = (1) spatial + (1) err_v")
            Sem.AStructure("cubes    =     spatial + (1) err_c")
            Sem.AStructure("lozenges =     spatial + (1) err_l")

            Sem.AStructure("paragrap = (1) verbal  + (1) err_p")
            Sem.AStructure("sentence =     verbal  + (1) err_s")
            Sem.AStructure("wordmean =     verbal  + (1) err_w")
            Sem.FitModel()
        Finally
            Sem.Dispose()
        End Try
    End Sub
End Module
```

We are going to copy its syntax structure, which is the same as that of two factor models that we want to fit. Amos uses Visual Basic (VB.NET) or C# language to specify a model. The beginning of the program describes a program title and backgrounds. Main program begins with `Sub Main()` and ends with `End Sub`. Pay attention to `Sem.BeginGroup()` and **`Sem.AStructure()`** functions. The former tells where a data file is located, while the latter specifies model to estimate. `Sem.TextOutput()` creates an output file and `Sem.Standardized()` computes standardized estimates.

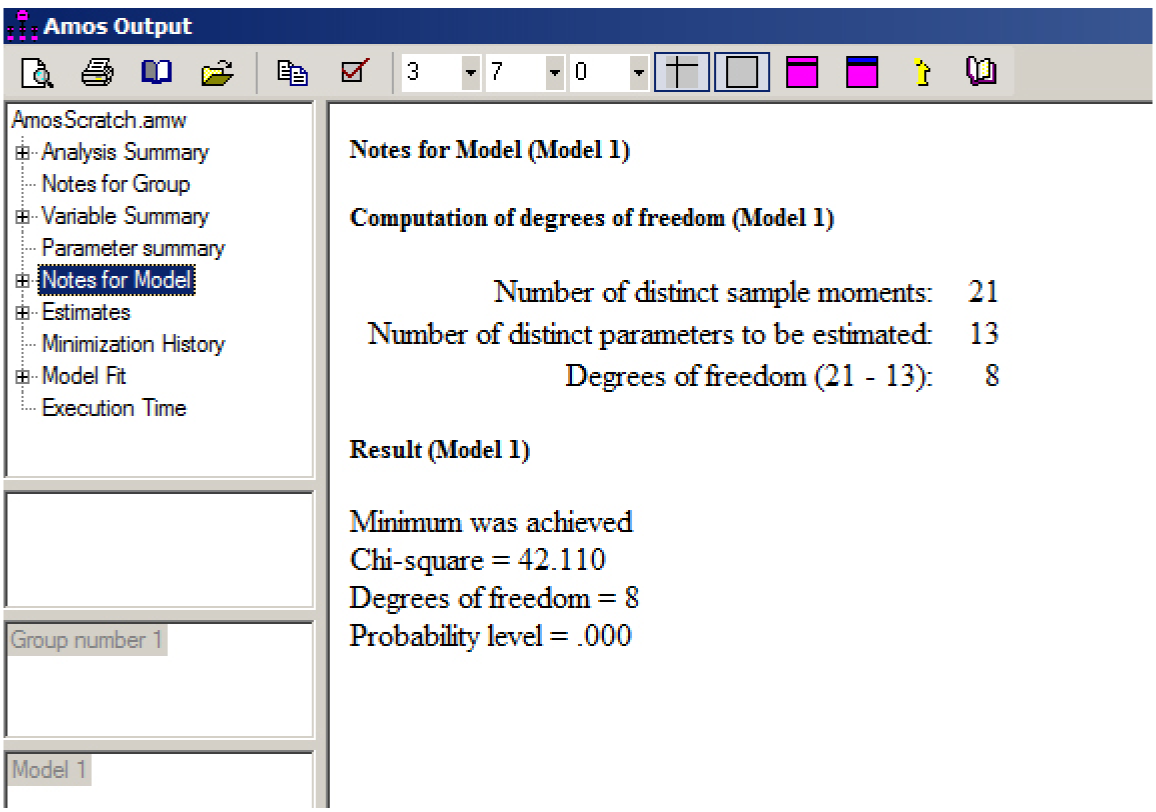
Let us rename the program file first. Click **File → Save As…** to open Save As dialog box. Provide a name *values* and then click **Save**. Now, the file name became *values.vb*.



Look at the comments in green. Replace *Example 8:* with *First Two Factor Model* and then remove other comments (see the screenshot below). Change `Sem.BeginGroup()` to look like `Sem.BeginGroup("c:\temp\cfa\values.sav")`. Amos reads data set *`values.sav`* from `c:\temp\cfa`. In the six `SEM.AStructure()` functions, replace `spatial` with `economic` and `verbal` with `morals`. Replace `visperc` with `privtown`, `cubes` with `govtresp`, and so on. Finally, change error terms from `e1` through `e6` as shown in the above screenshot.

The first `SEM.AStructure()` tells that the observed variable `privtown` is explained by a latent variable `economic` and error `e1`. Note that the estimate of factor loading is set 1 for the purpose of identification.

Click **File** → **Run** or click **Run** button on the right bottom to fit this model. Amos Output pops up if no error is found. You may see the list of results. Click **Notes for Model** on the left pane to get the chi-square test of model fit (see following screenshot). Chi-square χ^2 of 42.110 (df=8) is so large as to reject the null hypothesis of a good fit (p<.000).



Now click **Model Fit** on the left pane to see the other goodness-of-fit statistics. RMSEA .061 and CFI .961, which are the same as what Mplus and LISREL returned, suggests that the fit of the model is questionable.

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	13	42.110	8	.000	5.264
Saturated model	21	.000	0		
Independence model	6	891.221	15	.000	59.415

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.411	.988	.969	.376
Saturated model	.000	1.000		
Independence model	1.557	.786	.700	.561

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.953	.911	.961	.927	.961
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.533	.508	.513
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	34.110	17.433	58.302
Saturated model	.000	.000	.000
Independence model	876.221	782.083	977.751

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.036	.029	.015	.050
Saturated model	.000	.000	.000	.000
Independence model	.769	.756	.675	.844

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.061	.043	.079	.147
Independence model	.225	.212	.237	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	68.110	68.268	133.841	146.841
Saturated model	42.000	42.255	148.180	169.180
Independence model	903.221	903.294	933.558	939.558

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.059	.044	.080	.059
Saturated model	.036	.036	.036	.036
Independence model	.779	.698	.867	.779

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	427	553
Independence model	33	40

Click **Estimates** on the pane to check parameter estimates. Under the `Regression Weights` heading the unstandardized loadings appear along with standard errors, a critical ratios, and p-values. The standardized estimates under `Standardized Regression Weights` can be interpreted as the correlation between the observed variable and the corresponding common factor. These unstandardized and standardized estimates are the same as what Mplus and LISREL produced. However, standard errors are slightly different from Mplus and LISREL counterparts.

ABORTION (.743) and COMPETE (.732) have highest standardized factor loadings, GOVTRESP has a lowest factor loading of .147. Squared multiple correlations (R^2) corresponding to the six observed variables, which are arranged in an arbitrary order, indicate that the respective factor explains a respectable portion of the variance. Economic values (ECONOMIC) can explain only 2.2 percent (=$.147^2$) of variation in GOVTRESP although regression weights are all significant though. GOVTRESP does not seem to tap the same values dimension as the other two economics questions.

Estimates (Group number 1 - Model 1)

Scalar Estimates (Group number 1 - Model 1)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Model 1)

			Estimate	S.E.	C.R.	P	Label
privtown	<---	economic	1.000				
govtresp	<---	economic	.304	.084	3.631	***	
compete	<---	economic	1.351	.449	3.006	.003	
homosex	<---	morals	1.000				
abortion	<---	morals	1.018	.076	13.388	***	
euthanas	<---	morals	.637	.046	13.731	***	

Standardized Regression Weights: (Group number 1 - Model 1)

			Estimate
privtown	<---	economic	.577
govtresp	<---	economic	.147
compete	<---	economic	.732
homosex	<---	morals	.668
abortion	<---	morals	.743
euthanas	<---	morals	.563

Covariances: (Group number 1 - Model 1)

			Estimate	S.E.	C.R.	P	Label
economic	<-->	morals	.098	.126	.777	.437	

Correlations: (Group number 1 - Model 1)

			Estimate
economic	<-->	morals	.035

Variances: (Group number 1 - Model 1)

	Estimate	S.E.	C.R.	P	Label
economic	1.677	.577	2.907	.004	
morals	4.715	.487	9.674	***	
e1	3.366	.573	5.877	***	
e2	7.034	.297	23.698	***	
e3	2.645	1.019	2.596	.009	
e4	5.851	.404	14.491	***	
e5	3.957	.372	10.627	***	

	Estimate	S.E.	C.R.	P	Label
e6	4.119	.215	19.132	***	

Squared Multiple Correlations: (Group number 1 - Model 1)

	Estimate
euthanas	.317
abortion	.553
homosex	.446
compete	.536
govtresp	.022
privtown	.333

Unlike LISREL, an Amos program written in Program Editor is not able to produce a path diagram and Amos graphics does not generate a VB or C# program on the basis of a path diagram drawn.

GOVTRESP was only weakly accounted for by the ECONOMIC variable, hinting that the survey item was not tapping the same values dimension as the other two economic values indicators. An alternative possibility is that GOVTRESP is also tied to the morality dimension. Let us set a relationship between MORALS to GOVTRESP by adding `+ morals` to the second `Sem.AStructure()` below. It is like adding an arrow from MORALS to GOVTRESP on a path diagram. Look at the right part in red.

```
Header
Module MainModule
       ' Second Two Factor Model

   Sub Main()
   Dim Sem As New AmosEngine

       Try
          Sem.TextOutput()
          Sem.Standardized()
          Sem.Smc()

          Sem.BeginGroup( "c:\temp\cfa\values.sav")
          Sem.AStructure("privtown   = (1) economic + (1) e1")
          Sem.AStructure("govtresp   =   economic + morals + (1) e2")
          Sem.AStructure("compete  =   economic + (1) e3")

          Sem.AStructure("homosex  =  (1) morals  + (1) e4")
          Sem.AStructure("abortion  =  morals  + (1) e5")
          Sem.AStructure("euthanas  =  morals  + (1) e6")
                      Sem.FitModel()
       Finally
          Sem.Dispose()
       End Try

   End Sub
End Module
```

Run this program by clicking **File → Run** and then click **Notes for Model** in the Amos Output. The overall model fit appears quite good. χ^2 (df=7) decreases down to 9.884, which is too small to reject the null hypothesis of a good fit (p<.195). A small RMSEA of .019 and a large CFI of .997 indicate a good fit of this model.

Notes for Model (Model 1)

Computation of degrees of freedom (Model 1)

```
             Number of distinct sample moments:   21
Number of distinct parameters to be estimated:   14
                 Degrees of freedom (21 - 14):    7
```

Result (Model 1)

```
Minimum was achieved
Chi-square = 9.884
Degrees of freedom = 7
Probability level = .195
```

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	14	9.884	7	.195	1.412
Saturated model	21	.000	0		
Independence model	6	891.221	15	.000	59.415

RMR, GFI

Model	RMR	GFI	AGFI	PGFI
Default model	.119	.997	.991	.332
Saturated model	.000	1.000		
Independence model	1.557	.786	.700	.561

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.989	.976	.997	.993	.997
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.467	.461	.465
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	2.884	.000	15.404
Saturated model	.000	.000	.000
Independence model	876.221	782.083	977.751

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.009	.002	.000	.013
Saturated model	.000	.000	.000	.000
Independence model	.769	.756	.675	.844

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.019	.000	.044	.985
Independence model	.225	.212	.237	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	37.884	38.055	108.671	122.671
Saturated model	42.000	42.255	148.180	169.180
Independence model	903.221	903.294	933.558	939.558

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.033	.030	.043	.033
Saturated model	.036	.036	.036	.036
Independence model	.779	.698	.867	.779

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	1650	2167
Independence model	33	40

Amos reports the factor loading of MORALS on GOVTRESP on the third row. Unstandardized and standardized estimates are respectively .239 and .196, which are smaller than those of other estimates. GOVTRESP appears an unreliable indicator of both economic and moral values. Other standardized factor loadings range from .562 (EUTHANAS) to .734 (ABORTION). Corresponding squared multiple correlations (R^2) range from .060 (GOVTRESP) to .499 (COMPETE) and .539 (ABORTION). Latent variables ECONOMIC and MORALS account for 6 percent of the total variation in GOVTRESP. Covariance and correlation of two factors are .043 and .015 but they are not statistically discernable ($p<.7$). Finally, the correlation between the two common factors is a very small -.011, and the covariance estimate of -.030 is not statistically discernable from zero ($p<.808$).

Estimates (Group number 1 - Model 1)

Scalar Estimates (Group number 1 - Model 1)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Model 1)

			Estimate	S.E.	C.R.	P	Label
privtown	<---	economic	1.000				
govtresp	<---	economic	.285	.082	3.456	***	
govtresp	<---	morals	.239	.043	5.560	***	
compete	<---	economic	1.256	.422	2.977	.003	
homosex	<---	morals	1.000				
abortion	<---	morals	.991	.072	13.825	***	
euthanas	<---	morals	.627	.045	13.801	***	

Standardized Regression Weights: (Group number 1 - Model 1)

			Estimate
privtown	<---	economic	.598
govtresp	<---	economic	.143
govtresp	<---	morals	.196
compete	<---	economic	.707
homosex	<---	morals	.678
abortion	<---	morals	.734
euthanas	<---	morals	.562

Covariances: (Group number 1 - Model 1)

			Estimate	S.E.	C.R.	P	Label
economic	<-->	morals	.043	.132	.325	.745	

Correlations: (Group number 1 - Model 1)

			Estimate
economic	<-->	morals	.015

Variances: (Group number 1 - Model 1)

	Estimate	S.E.	C.R.	P	Label
economic	1.806	.625	2.888	.004	
morals	4.855	.489	9.933	***	
e1	3.237	.619	5.228	***	
e2	6.760	.290	23.330	***	
e3	2.857	.960	2.974	.003	
e4	5.711	.399	14.329	***	
e5	4.078	.356	11.459	***	
e6	4.123	.214	19.275	***	

Squared Multiple Correlations: (Group number 1 - Model 1)

	Estimate
euthanas	.316
abortion	.539
homosex	.460
compete	.499
govtresp	.060
privtown	.358

In summary, there appear to be two orthogonal dimensions which underlie American attitudes on a number of different issues: one representing economic values and the other representing moral values. It is unclear which dimension the GOVTRESP item was tapping, however, and future surveys should employ a more reliable measure.

3.6 Summary

Tables 4 and 5 compare the unstandardized and standardized factor loadings from each software package for both two-factor models. The tables also present standard errors in parentheses, χ^2, Root Mean Square Error of Approximation (RMSEA), and Comparative Fit Index (CFI).

Table 4 Comparison of Estimates: First Two Factor Model (N=1,160, DF=8)

	Amos		LISREL		Mplus		SAS CALIS	
	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard
F1 → PRIVTOWN	1.00	.577	1.00	.58	1.000	.577 (.094)	1.0000	.5767
F1 → GOVTRESP	.304 (.084)	.147	.30 (.084)	.15	.304 (.083)	.147 (.038)	.2546 (.0701)	.1468
F1 → COMPETE	1.351 (.449)	.732	1.35 (.45)	.73	1.350 (.433)	.732 (.118)	1.2699 (.4224)	.7323
F2 → HOMOSEX	1.00	.668	1.00	.67	1.000	.668 (.028)	1.0000	.6680
F2 → ABORTION	1.018 (.076)	.743	1.02 (.076)	.74	1.018 (.076)	.743 .029)	1.1127 (.0831)	.7433
F2 → EUTHANAS	.637 (.046)	.563	.64 (.046)	.56	.637 (.046)	.563 (.028)	.8426 (.0614)	.5629
F1 ←→ F2	.098[a] (.126)	.035[b]	.10[a] (.13)	.03[b]	.098[a] (.124)	.035[b] (.044)	.0134[a] (.0172)	.0347[b]
χ^2, RMSEA, CFI	42.110 .061 .961			42.11 .060 .96		42.147 .061 .961		42.1104 .0607 .9611

Standard errors appear in parentheses
[a] covariance; [b] correlation between F1 (economic values) and F2 (moral values)

Table 5 : Comparison of Estimates: Second Two Factor Model (N=1,160, DF=7)

	Amos		LISREL		Mplus		SAS CALIS	
	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard
F1 → PRIVTOWN	1.00	.598	1.00	.60	1.000	.598 (.107)	1.0000	.5984
F1 → GOVTRESP	.285 (.082)	.143	.28 (.082)	.14	.285 (.082)	.143 (.038)	.2387 (.0691)	.1428
F1 → COMPETE	1.256 (.422)	.707	1.26 (.42)	.71	1.256 (.442)	.707 (.125)	1.1809 (.3967)	.7066
F2 → HOMOSEX	1.00	.678	1.00	.68	1.000	.678 (.028)	1.0000	.6779
F2 → ABORTION	.991 (.072)	.734	.99 (.072)	.73	.991 (.072)	.734 .028)	1.0828 (.0783)	.7340
F2 → EUTHANAS	.627 (.045)	.562	.63 (.045)	.56	.627 (.046)	.562 (.028)	.8295 (.0601)	.5623
F2 →GOVTRESP	.239 (.043)	.196	.24 (.043)	.20	.239 (.043)	.196 (.034)	.2893 (.0520)	.1961
F1 ←→ F2	.043[a] (.132)	.015[b]	.04[a] (.13)	.01[b]	.043[a] (.136)	.015[b] (.047)	.0059[a] (.0181)	.0145[b]
χ^2, RMSEA, CFI		9.884 .019 .997		9.88 .019 1.00		9.893 .019 .997		9.8844 .0189 .9967

Standard errors appear in parentheses
[a] covariance; [b] correlation between F1 (economic values) and F2 (moral values)

Mplus, LISREL, and Amos produce almost same statistics. SAS/STAT CALIS reports different unstandidized factor loadings but same standardized statistics. Mplus reports key goodness-of-fit statistics, while other software packages provide various fit statistics including χ^2, RMSEA, and CFI.

4. Confirmatory Factor Analysis with Missing Data

This section shows how to estimate the two-factor model when the raw data matrix includes missing observations. The data to be analyzed has been saved as an SPSS file named *values_full.sav* in the C:\temp\CFA folder. All missing observations have been coded as system missing (.) in SPSS.

4.1 Missing Data Issue

Missing data is a pervasive problem in the social sciences. A subject may fail to complete a test in an experimental setting, refuse to give an answer to a particular survey item, or drop out of a panel. In many cases, researchers choose to drop all observations from subjects that have missing observations on any of the items included in the model. This approach to handling missing data is referred to as *listwise deletion* and is the default in programs such as SPSS and Stata. Unfortunately dropping incomplete cases results in sacrificing information from the sample and can lead to biased estimates when the data is not missing completely at random. Another approach is *pairwise deletion* that removes observations with missing data in any one of two variables when computing their covariance. This method of dealing with missing data can use all available data but each element of covariance matrix may be based on different observations.

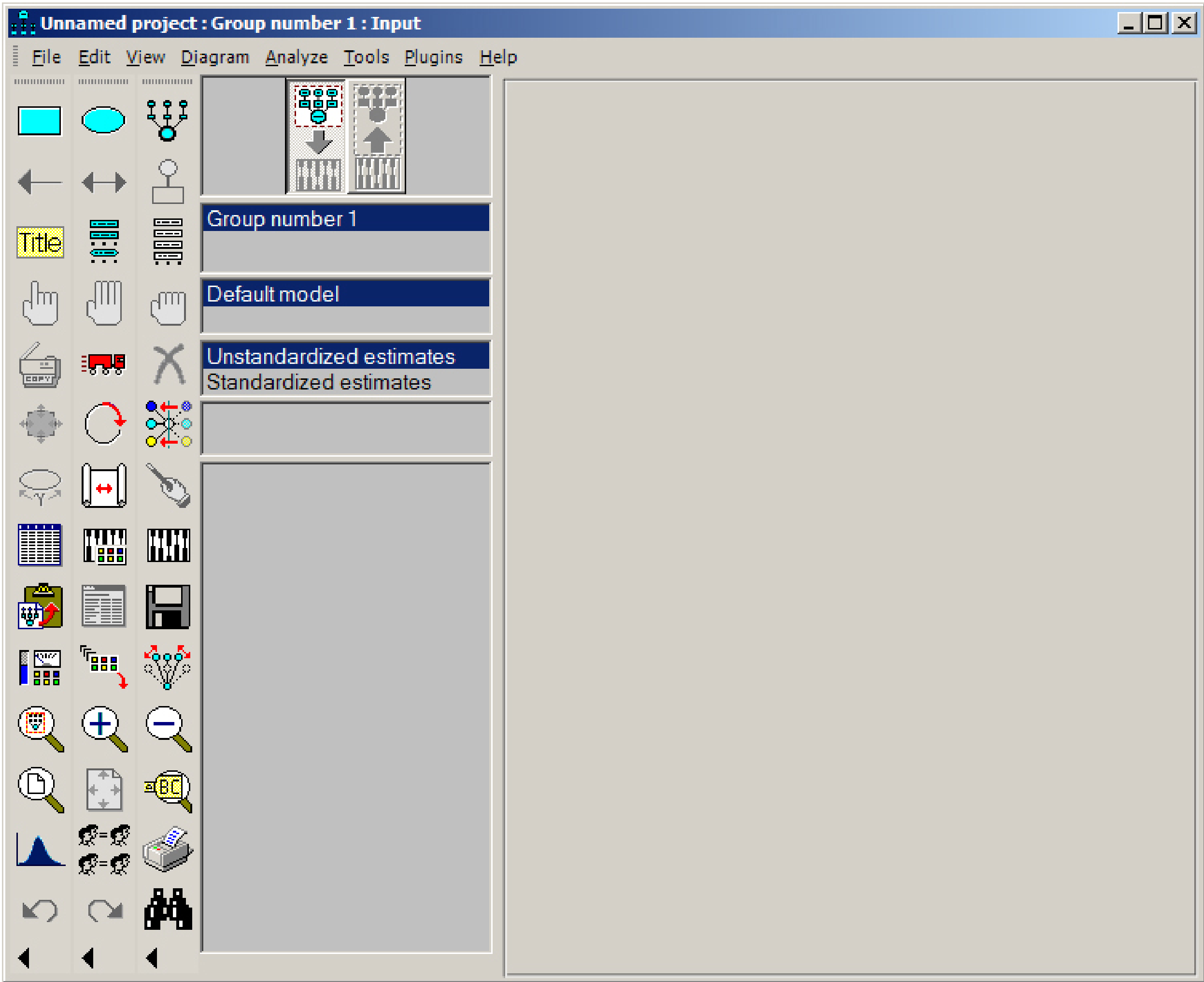
Over the last 30 years more sophisticated means have emerged for dealing with missing data, many of which have been incorporated into structural equation modeling software. Because it is available in Amos, LISREL, and Mplus, this document will consider Full Information Maximum Likelihood (FIML) estimation which makes maximal use of all data available from every subject in the sample. Other approaches to dealing with missing data, such as multiple imputation via Bayesian simulation, may also be available depending on the specific software packages. A non-technical overview of different methods for handling missing data in the context of structural equation models is available in Enders (2001), though the description of capabilities of specific computer packages is already dated.

4.2 CFA with Missing Data using Amos

This section fits the model using Amos Graphics instead of Program Editor. Amos can be launched from any computer running Windows in the UITS Student Technology Centers by going to S**tart → All Programs → Departmentally Sponsored → Statistics-Math →Amos 16 → Amos Graphics**. The following screen will display:

On the far left pane appear the different tools that can be used to create path diagrams. Just to the right of the toolbar buttons is a column that will display information about the model after estimates have been calculated. The remainder of the screen contains the area where the path diagram will be drawn.

Click **File →Data Files** to load data. After the **Data Files** dialog box opens, click on **File Name**. Navigate to the C:\temp\CFA folder and choose *values_full.sav* containing missing values. Click **Open**, Then **Okay**.



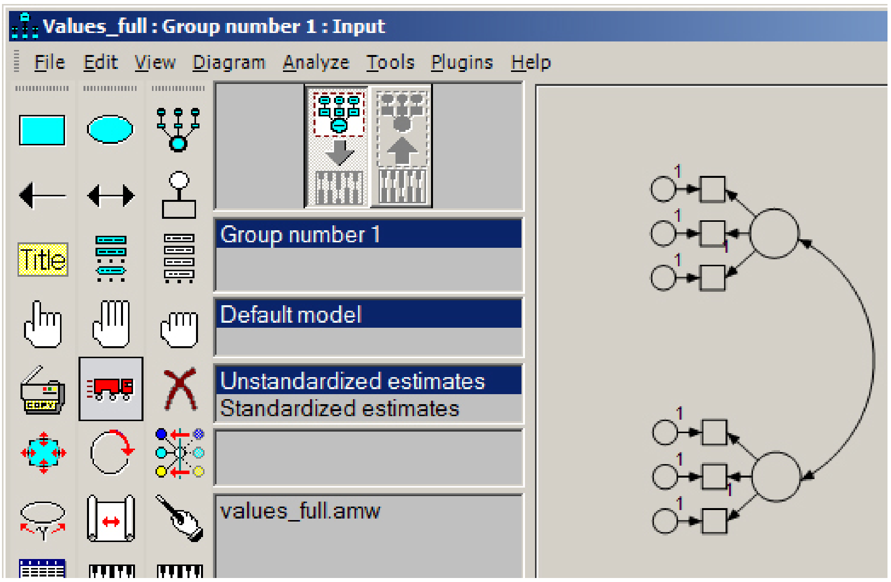
On the left pane, click on the **Draw unobserved variable** button in the tool box (alternatively click on **Diagram → Draw Unobserved**) and click some place on the right workspace to draw an oval representing a latent variable. Move the mouse pointer just below the oval and click once to create a second oval. Add a covariance between the two latent variables (common factors) by choosing the **Draw Covariances** button and clicking and dragging a two-headed arrow from one factor to the other (see the screenshot on the next page).

Click on the **Draw a latent variable button or add an indicator to a latent variable** button . Click three times inside each oval to add a total of six indicators and their respective error terms (see the following screenshot).[7] By default Amos sets the metric of each error term

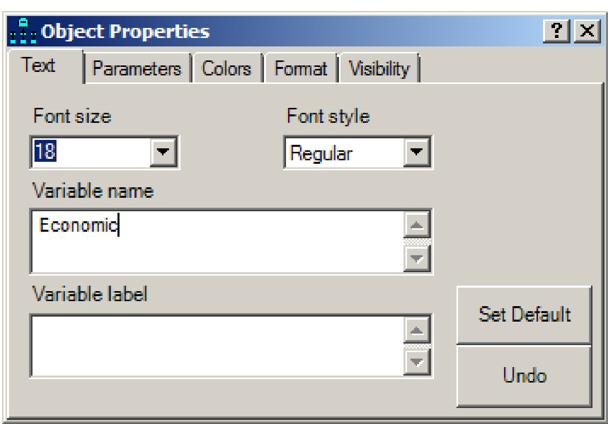
---

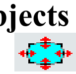
[7] Alternatively, you may draw ovals and boxes and then link them with appropriate arrows manually. Click on the blue rectangle button of **Draw Observed Variable** (alternatively click on **Diagram → Draw Observed**). Then click six times to create a total of six boxes for observed variables. Select the blue oval and click six times on the left of six boxes to represent measurement error specific to each of the observed indicators. Select the **Draw Paths** button for drawing single headed arrows. Click and drag from the common and unique factors to the appropriate observed variable.

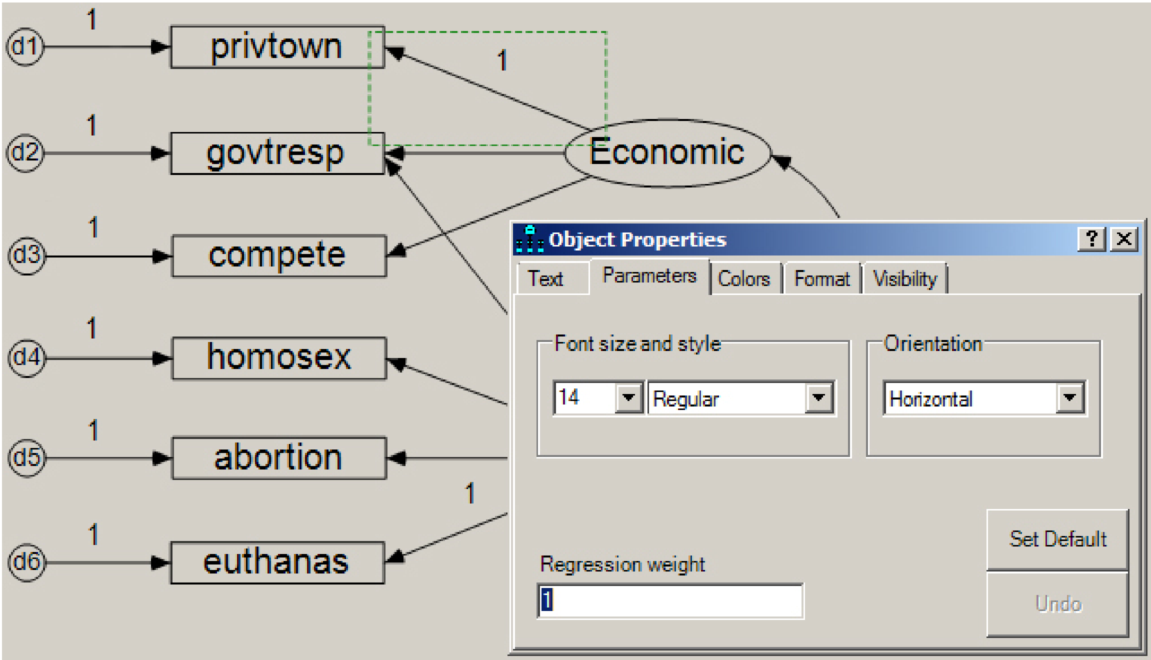
by constraining the path parameters to one. The factor loading of the first indicator for each latent variable is also set to one. Rotate each latent variable by choosing the **Rotate the indicators of a latent variable** button  and clicking each factor three times until you are satisfied with the appearance so that the path diagram should now look like the following.



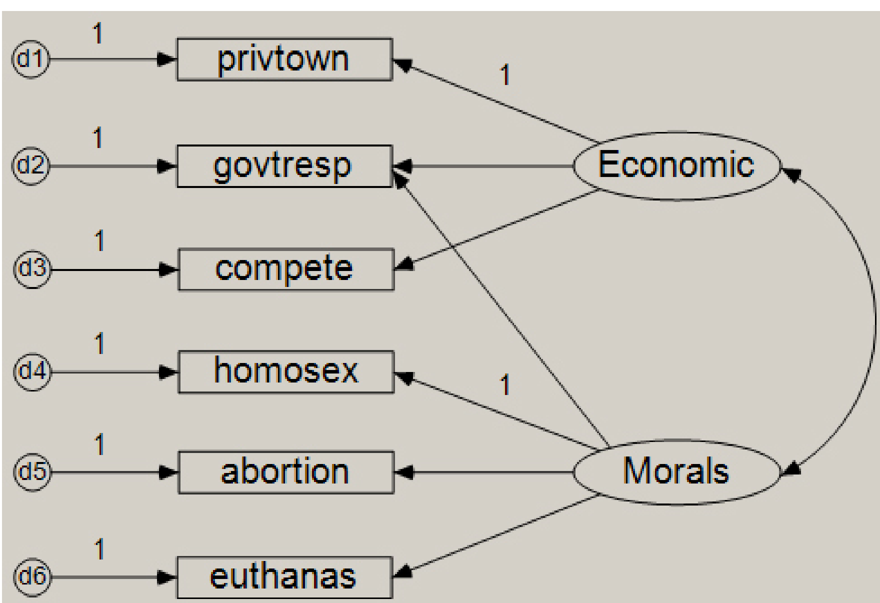
To label the latent variables, right-click in the first oval and choose **Object Properties**. When the **Object Properties** dialog box opens, choose the **Text** tab. Name the variable ECONOMIC and then click the second oval to name it MORALS. Click the first error oval and name it d_1 and adjust the font size to 12 point. Do the same to name the error terms d_2 through d_6. Click **X** in the upper right hand corner to close the **Object Properties** dialog box.



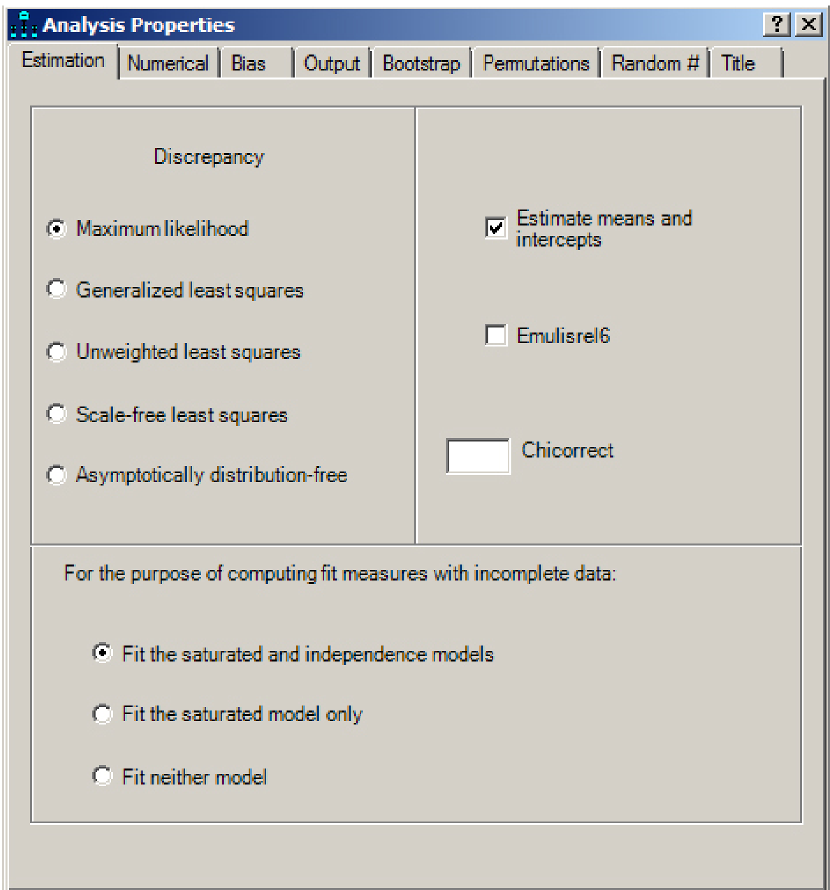
To name the observed variables, choose **View → Variables in Dataset**. Click and drag the names of each variable to the appropriate box in the path diagram. If the names do not fit, you can change the position of parts of the diagram after choosing the **Move objects** button  or resize the box after clicking on the **Change the shape of objects** button . Finally, add an arrow from the MORALS latent variable to the GOVTRESP indicator as suggested in the earlier analysis. The path diagram now should look something like the following:

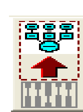


For comparison with LISREL and Mplus, let us set factor loading on PRIVTOWN and HOMOSEX to one and factor loadings on COMPETE and EUTHANAS free. Right-click on the arrow from ECONOMIC to COMPETE to open **Object Properties** dialog box and then click on the **Paramaters** tab and delete 1 in the field labeled **Regression weight**. Click on the arrow from ECONOMIC to PRIVTOWN and enter 1 in **Object Properties** dialog box (see screenshot above). Repeat the same to set the factor loading on HOMOSEX to 1 after set factor loadings on EUTHANAS free. Click **X** to close the **Object Properties** dialog box. The final path diagram should look like this:



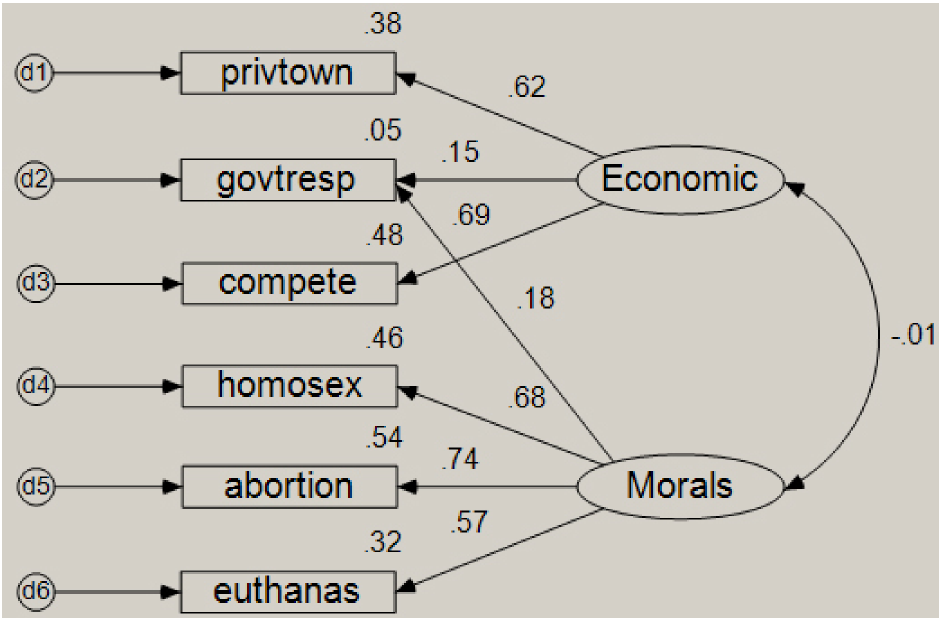
When missing values are present, it is necessary to request that Amos estimate means and intercepts (required for FIML estimation), which is not the default. Choose **View →Analysis Properties**, click the **Estimation** tab in the **Analysis Properties** dialog box, and select **Estimate means and intercepts**. Next click on the **Output** tab. **Minimization History** is checked by default. Also place checks next to **Standardized Estimates** and **Squared Multiple Correlations**.



Now, we are ready to fit the model. Go to **Analyze → Calculate Estimates**. To see the results in the path diagram click on the **View the output path diagram** button  on the second pane.

The unstandardized estimates are displayed by default. The path diagram for unstandardized estimates is often hard to read especially when the diagram does not have enough space for parameter estimates. To bring up the standardized estimates, click on the **Standardized estimates** option in the column between the tools and the workspace.

Amos now displays the standardized factor loadings on arrows and the squared multiple correlation coefficient for each observed variable. All factor loadings except for GOVTRESP are large, while the correlation between two latent variables is negligible (-.01). Note that for some models with many parameters and missing data, Amos (and all SEM software) may require a large number of iterations to estimate a χ^2 statistic.



To see more detail about the results, go to **View** → **Text Output**. A selected portion of the output is the following:

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

```
             Number of distinct sample moments:   27
 Number of distinct parameters to be estimated:   20
                  Degrees of freedom (27 - 20):    7
```

Result (Default model)

```
Minimum was achieved
Chi-square = 9.911
Degrees of freedom = 7
Probability level = .194
```

Estimates (Group number 1 - Default model)

Scalar Estimates (Group number 1 - Default model)

Maximum Likelihood Estimates

Regression Weights: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
compete <--- Economic	1.200	.369	3.255	.001	
govtresp <--- Economic	.302	.082	3.695	***	
privtown <--- Economic	1.000				
euthanas <--- Morals	.633	.045	14.009	***	
abortion <--- Morals	.992	.071	13.937	***	
homosex <--- Morals	1.000				
govtresp <--- Morals	.215	.042	5.070	***	

Standardized Regression Weights: (Group number 1 - Default model)

	Estimate
compete <--- Economic	.691
govtresp <--- Economic	.154
privtown <--- Economic	.615

			Estimate
euthanas	<---	Morals	.566
abortion	<---	Morals	.737
homosex	<---	Morals	.677
govtresp	<---	Morals	.176

Intercepts: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
compete	3.442	.069	49.743	***	
govtresp	4.312	.078	55.366	***	
privtown	3.541	.065	54.587	***	
euthanas	2.645	.071	37.076	***	
abortion	4.360	.086	50.760	***	
homosex	4.774	.095	50.472	***	

Covariances: (Group number 1 - Default model)

			Estimate	S.E.	C.R.	P	Label
Morals	<-->	Economic	-.016	.134	-.118	.906	

Correlations: (Group number 1 - Default model)

			Estimate
Morals	<-->	Economic	-.005

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
Economic	1.902	.604	3.150	.002	
Morals	4.850	.485	10.000	***	
d3	2.998	.847	3.542	***	
d2	6.872	.289	23.745	***	
d6	4.115	.213	19.361	***	
d5	4.021	.353	11.407	***	
d4	5.732	.396	14.481	***	
d1	3.118	.596	5.235	***	

Squared Multiple Correlations: (Group number 1 - Default model)

	Estimate
homosex	.458
abortion	.543
euthanas	.321
privtown	.379
govtresp	.054
compete	.477

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	20	9.911	7	.194	1.416
Saturated model	27	.000	0		
Independence model	6	910.164	21	.000	43.341

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.989	.967	.997	.990	.997
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Default model	.333	.330	.332
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Default model	2.911	.000	15.448
Saturated model	.000	.000	.000
Independence model	889.164	794.150	991.574

FMIN

Model	FMIN	F0	LO 90	HI 90
Default model	.008	.002	.000	.013
Saturated model	.000	.000	.000	.000
Independence model	.759	.742	.662	.827

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.019	.000	.043	.987
Independence model	.188	.178	.198	.000

AIC

Model	AIC	BCC	BIC	CAIC
Default model	49.911	50.146		
Saturated model	54.000	54.317		
Independence model	922.164	922.235		

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Default model	.042	.039	.052	.042
Saturated model	.045	.045	.045	.045
Independence model	.769	.690	.855	.769

HOELTER

Model	HOELTER .05	HOELTER .01
Default model	1702	2235
Independence model	44	52

The overall model fit appears quite good. The χ^2 test yields a statistic of 9.911 (df=7), which has a corresponding p-value of .194. This p-value is too high to reject the null hypothesis of a good fit. The RMSEA of .019 and CFI of .997 also suggest that the model fits the data well.

Under the `Regression Weights` heading, the unstandardized loadings appear along with standard errors, critical ratios (test statistics), and p-values. A critical ratio greater than 1.96 or a p-value smaller than .05 signifies the parameter is statistically discernable from zero at the .05 significance level. Three asterisks (***) indicate that the p-value is smaller than .001. In this case all of the unconstrained estimates are statistically significant. Notice that the order of variables listed in the table is not the same as what is specified in the path diagram; it is a bit confusing and inconvenient to read results.[8]

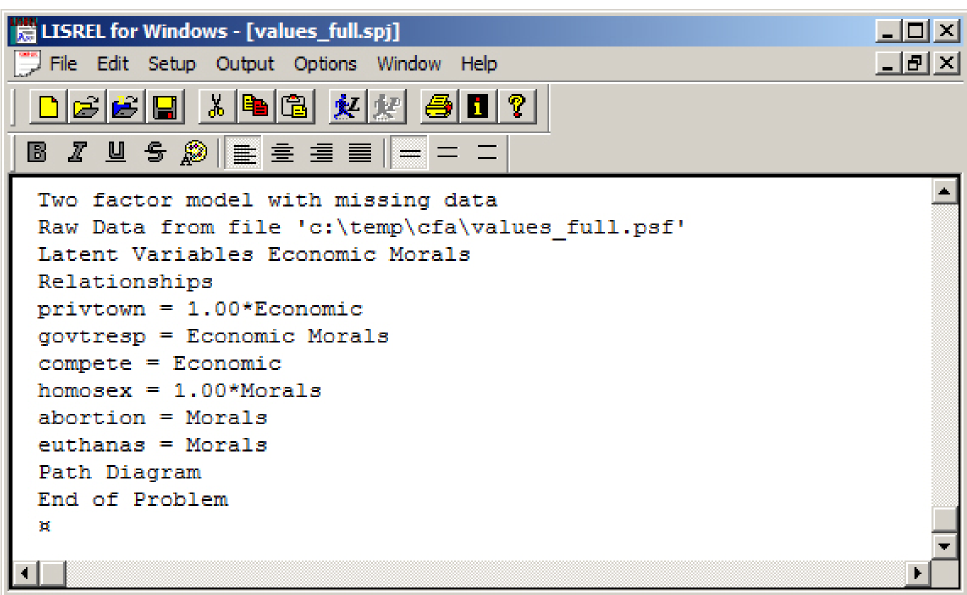
[8] Variables are recognized from the left when created by clicking **Draw a latent variable button or add an indicator to a latent variable** button to right, but we rotated the path diagram so that manifest variables are vertically arranged.

The unstandardized weights are highly sensitive to model constraints, whereas the standardized regression weights provide more intuitive information about the strength of factor loadings. The GOVTRESP has low standardized loadings on both factors (.154 for ECONOMIC and .176 for MORALS), suggesting that it is an unreliable indicator of both economic and moral values. However, the other indicators have moderate to strong standardized loadings. For EUTHANAS the factor loading is .566, for PRIVTOWN it is .615, for HOMOSEX it is .677, for COMPETE it is .691, and for ABORTION it is .737.

The squared multiple correlations provide information about how much variance of an observed variable the factors can account for. Despite receiving a path from both latent variables, GOVTRESP has a very low R^2 of only .054. The remaining R^2 statistics are, in order of increasing magnitude, .321 (EUTHANAS), .379 (PRIVTOWN), .458 (HOMOSEX), .477 (COMPETE), and .543 (ABORTION). Finally, the covariance between the two common factors is -.012 and their correlation is -.005, which is not statistically distinguishable from zero.

4.3 CFA with Missing Data using LISREL

Launch LISREL and open the file *values_full.sav* by choosing **File → Import Data**. When **Open** dialog box opens, change **Files of type** to *SPSS Data File(*.sav)*, navigate to the folder C:\temp\CFA, and choose *values_full.sav*. Click **Open** to import the data set. A prompt appears immediately to save the file as a PRELIS system file (.psf). Enter *values_full* and click **Save**. A spreadsheet will open displaying the data. Notice that missing observations are coded -999999.0.[9]

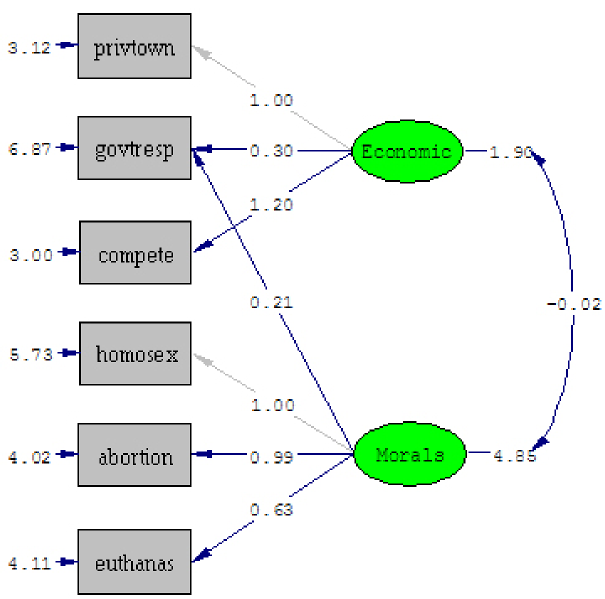


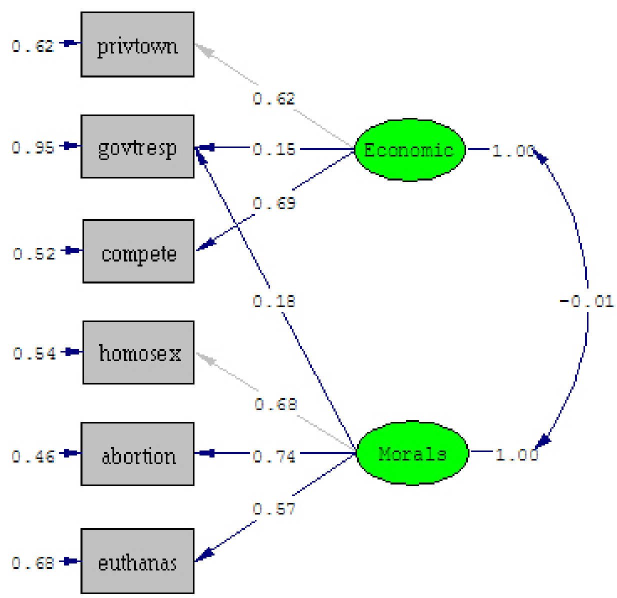
[9] To make sure that PRELIS understands these are missing values, it is necessary to declare them as such. Go to **Data → Define Variables** to open the **Define Variables** dialog box. Highlight each variable name by clicking on PRIVTOWN, holding down the shift key, and clicking on EUTHANAS and others. Click on **Missing Values** to bring up the **Missing Values** box. Click on the **Missing Values** radio button, enter -999999.0 in the first empty field, and check the **Apply to all** option. Click **OK**, then **OK** again.

In 3.2, we drew a path diagram for two factor CFA and then generated a corresponding SIMPLIS or LISREL syntax. This approach is intuitive but wordy to explain each step all the way to the final. Here let us directly write a SIMPLIS program in the text editor.

Go to **File** → **New**, choose **SIMPLIS Project** in the **New** dialog box, and click **OK**. When the **Save As** dialog box opens, type in *values_full*, and then click **Save**. An empty text window opens for you to type in the commands shown in the screenshot in the previous page. Notice that *values_full.psf* is used as input data and the metric of the two common factors is set by constraining these factor loadings on PRIVTOWN and HOMOSEX to equal one.

Click on the **Run LISREL** button to fit the model. The unstandardized estimates are immediately displayed in the path diagram along with two measures of overall fit: χ^2 and RMSEA. To view the standardized results, choose **Standardized Solution** from the **Estimates** pull-down menu as we did in 3.2. The path diagrams for unstandardized (left diagram below) and standardized (right diagram) estimates will look like these:





More detailed information can be obtained by looking at the output text file generated after estimation. This file is given the same name as the path diagram plus an .out extension and stored in the working directory. The file *values_full.out* looks like the following:

```
                        DATE:  1/17/2009
                          TIME: 19:09


                        L I S R E L  8.80

                              BY

                  Karl G. Jöreskog & Dag Sörbom


             This program is published exclusively by
             Scientific Software International, Inc.
```

```
                    7383 N. Lincoln Avenue, Suite 100
                     Lincolnwood, IL 60712, U.S.A.
          Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
        Copyright by Scientific Software International, Inc., 1981-2006
         Use of this program is subject to the terms specified in the
                     Universal Copyright Convention.
                       Website: www.ssicentral.com

 The following lines were read from file c:\Temp\CFA\values_full.spj:

 Two factor model with missing data
 Raw Data from file 'c:\temp\cfa\values_full.psf'

              --------------------------------
               EM Algorithm for missing Data:
              --------------------------------

          Number of different missing-value patterns=      12
          Convergence of EM-algorithm in    3 iterations
          -2 Ln(L) =   33282.55078
          Percentage missing values=   0.71

     Note:
       The Covariances and/or Means to be analyzed are estimated
       by the EM procedure and are only used to obtain starting
       values for the FIML procedure

 Latent Variables Economic Morals
 Relationships
 privtown = 1.00*Economic
 govtresp = Economic Morals
 compete = Economic
 homosex = 1.00*Morals
 abortion = Morals
 euthanas = Morals
 Path Diagram
 End of Problem

 Sample Size =  1200

 Two factor model with missing values

        Covariance Matrix

           privtown   govtresp    compete    homosex   abortion   euthanas
           --------   --------   --------   --------   --------   --------
 privtown      5.02
 govtresp      0.55       7.27
  compete      2.28       0.70       5.74
  homosex     -0.27       1.26      -0.08      10.59
 abortion     -0.15       0.89       0.00       4.81       8.81
 euthanas      0.11       0.64       0.33       3.03       3.08       6.06


 Two factor model with missing values

 Number of Iterations =  5

 LISREL Estimates (Maximum Likelihood)

        Measurement Equations


 privtown = 1.00*Economic, Errorvar.= 3.12 , R² = 0.38
                              (0.60)
                               5.23

 govtresp = 0.30*Economic + 0.21*Morals, Errorvar.= 6.87 , R² = 0.054
          (0.082)        (0.042)              (0.29)
           3.70           5.07                 23.74
```

```
 compete = 1.20*Economic, Errorvar.= 3.00 , R² = 0.48
        (0.37)                (0.85)
         3.26                  3.54

 homosex = 1.00*Morals, Errorvar.= 5.73 , R² = 0.46
                             (0.40)
                              14.48

abortion = 0.99*Morals, Errorvar.= 4.02 , R² = 0.54
        (0.071)              (0.35)
         13.94                11.41

euthanas = 0.63*Morals, Errorvar.= 4.11 , R² = 0.32
        (0.045)              (0.21)
         14.01                19.36


      Covariance Matrix of Independent Variables

         Economic    Morals
         --------   --------
Economic      1.90
           (0.60)
             3.15
  Morals     -0.02      4.85
           (0.13)    (0.48)
            -0.12     10.00


           Global Goodness of Fit Statistics, Missing Data Case


             -2ln(L) for the saturated model =      33282.551
             -2ln(L) for the fitted model   =      33292.471

                        Degrees of Freedom = 7
             Full Information ML Chi-Square  = 9.92 (P = 0.19)
          Root Mean Square Error of Approximation (RMSEA) = 0.019
          90 Percent Confidence Interval for RMSEA = (0.0 ; 0.043)
            P-Value for Test of Close Fit (RMSEA < 0.05) = 0.99
```

The result is quite similar to that of Amos. The χ^2 test yields a value of 9.92 (df=7) and RMSEA is only .019, indicating a good fit. CFI is not, however, reported here. Unconstrained estimates are shown under the `Measurement Equations` heading and standardized solutions on the path diagram. The GOVTRESP has low standardized loadings from both factors (.15 for ECONOMIC and .18 for MORALS) and a very low R^2 of .054, suggesting that it is a weak indicator of both economic and moral values. However, the other indicators have moderate to strong standardized loadings with higher squared multiple correlations ranging from .32 for EUTHANAS to .54 for ABORTION. Finally, the covariance and correlation between the two common factors are -.02 and -.01, respectively, which are not statistically distinguishable from zero.

4.4 CFA with Missing Data using Mplus

Unlike Amos and LISREL, Mplus cannot directly read an SPSS data file. Instead raw data must be saved as an ASCII file in free or fixed format. Since Mplus does not treat blanks as missing values, missing values must be coded as a number in the raw data and explicitly specified in the **VARIABLE** command of the Mplus syntax.

In order to recode missing values, open SPSS, choose **File →New → Syntax** to open a SPSS Syntax Editor window, and then enter the following commands. These two commands recode missing values as -1 and write a tab-delimited text file *values_full.dat* in the C:\temp\CFA directory.

```
RECODE privtown govtresp compete homosex abortion euthanas (SYSMIS=-1).
SAVE TRANSLATE OUTFILE='c:\temp\CFA\values_full.dat' /TYPE=TAB/MAP.
```

Now launch Mplus and type in the following syntax for the two factor model with missing data.[10]

```
TITLE:    Two Factor Model with Missing Data;
DATA:     FILE IS c:\temp\cfa\values_full.dat;
VARIABLE: NAMES ARE privtown govtresp compete
          homosex abortion euthanas;
          MISSING ARE ALL(-1);
MODEL:    economic BY privtown govtresp compete;
          morals BY homosex abortion euthanas govtresp;
OUTPUT:   STANDARDIZED;
          MODINDICES;
```

The TITLE line provides a short description for the analysis. The DATA statement specifies the path name for the tab-delimited raw data file to be analyzed. The VARIABLE statement lists the names of the variables in the order they appear in the data file. The MISSING ARE option tells Mplus to interpret the numeric value -1 as missing for all variables. The MODEL statement tells Mplus that there are two latent variables and six outcome (manifest) variables. The OUTPUT statement here requests that standardized estimates and modification indices appear in the output file.

To begin the estimation, click on **Run** . A text output file appears and is saved in the working directory. The output file looks like the following:

```
INPUT INSTRUCTIONS

  TITLE:    Two Factor Model with Missing Data;
  DATA:     FILE IS c:\temp\cfa\values_full.dat;
  VARIABLE: NAMES ARE privtown govtresp compete
            homosex abortion euthanas;
            MISSING ARE ALL(-1);
  MODEL:    economic BY privtown govtresp compete;
            morals BY homosex abortion euthanas govtresp;
  OUTPUT:   STANDARDIZED;
            MODINDICES;

INPUT READING TERMINATED NORMALLY

Two Factor Model with Missing Values;

SUMMARY OF ANALYSIS

Number of groups                                     1
Number of observations                            1200
```

[10] In previous edition, you need to add "ANALYSIS: TYPE=MISSING H1;" which becomes a default setting in version 5. The TYPE = MISSING option in the ANALYSIS statement tells Mplus to use an estimator appropriate for the presence of missing data, and the *h1* requests a chi-square statistic for model fit.

```
Number of dependent variables                              6
Number of independent variables                            0
Number of continuous latent variables                      2

Observed dependent variables

  Continuous
   PRIVTOWN   GOVTRESP   COMPETE    HOMOSEX    ABORTION   EUTHANAS

Continuous latent variables
   ECONOMIC   MORALS


Estimator                                                 ML
Information matrix                                  OBSERVED
Maximum number of iterations                            1000
Convergence criterion                              0.500D-04
Maximum number of steepest descent iterations             20
Maximum number of iterations for H1                     2000
Convergence criterion for H1                       0.100D-03

Input data file(s)
  c:\temp\cfa\values_full.dat

Input data format  FREE


SUMMARY OF DATA

     Number of missing data patterns          12


COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value   0.100


     PROPORTION OF DATA PRESENT

           Covariance Coverage
              PRIVTOWN      GOVTRESP      COMPETE       HOMOSEX       ABORTION
              ________      ________      ________      ________      ________
 PRIVTOWN        0.994
 GOVTRESP        0.994         0.999
 COMPETE         0.994         0.999         0.999
 HOMOSEX         0.977         0.981         0.981         0.982
 ABORTION        0.988         0.992         0.992         0.976         0.993
 EUTHANAS        0.985         0.990         0.990         0.975         0.985


           Covariance Coverage
              EUTHANAS
              ________
 EUTHANAS        0.991


THE MODEL ESTIMATION TERMINATED NORMALLY


TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                              9.920
          Degrees of Freedom                     7
          P-Value                           0.1932

Chi-Square Test of Model Fit for the Baseline Model

          Value                            910.923
```

```
        Degrees of Freedom                    15
        P-Value                           0.0000

CFI/TLI

        CFI                                0.997
        TLI                                0.993

Loglikelihood

        H0 Value                      -16646.235
        H1 Value                      -16641.275

Information Criteria

        Number of Free Parameters             20
        Akaike (AIC)                   33332.470
        Bayesian (BIC)                 33434.272
        Sample-Size Adjusted BIC       33370.744
          (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                           0.019
        90 Percent C.I.                    0.000  0.043
        Probability RMSEA <= .05           0.987

SRMR (Standardized Root Mean Square Residual)

        Value                              0.016


MODEL RESULTS
                                                    Two-Tailed
                 Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           1.000      0.000    999.000    999.000
    GOVTRESP           0.302      0.082      3.683      0.000
    COMPETE            1.201      0.384      3.124      0.002

 MORALS   BY
    HOMOSEX            1.000      0.000    999.000    999.000
    ABORTION           0.992      0.072     13.864      0.000
    EUTHANAS           0.633      0.046     13.902      0.000
    GOVTRESP           0.215      0.042      5.105      0.000

 MORALS   WITH
    ECONOMIC          -0.016      0.140     -0.112      0.911

 Intercepts
    PRIVTOWN           3.541      0.065     54.608      0.000
    GOVTRESP           4.312      0.078     55.390      0.000
    COMPETE            3.442      0.069     49.763      0.000
    HOMOSEX            4.774      0.095     50.494      0.000
    ABORTION           4.360      0.086     50.780      0.000
    EUTHANAS           2.645      0.071     37.091      0.000

 Variances
    ECONOMIC           1.900      0.627      3.029      0.002
    MORALS             4.850      0.486      9.974      0.000

 Residual Variances
    PRIVTOWN           3.120      0.619      5.041      0.000
    GOVTRESP           6.872      0.289     23.738      0.000
    COMPETE            2.995      0.882      3.395      0.001
    HOMOSEX            5.731      0.398     14.409      0.000
    ABORTION           4.022      0.352     11.416      0.000
    EUTHANAS           4.115      0.212     19.419      0.000
```

```
STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           0.615      0.100      6.167      0.000
    GOVTRESP           0.154      0.037      4.202      0.000
    COMPETE            0.691      0.111      6.217      0.000

 MORALS   BY
    HOMOSEX            0.677      0.027     24.755      0.000
    ABORTION           0.737      0.027     26.842      0.000
    EUTHANAS           0.566      0.027     20.849      0.000
    GOVTRESP           0.176      0.034      5.241      0.000

 MORALS   WITH
    ECONOMIC          -0.005      0.046     -0.112      0.911

 Intercepts
    PRIVTOWN           1.580      0.043     36.432      0.000
    GOVTRESP           1.600      0.044     36.690      0.000
    COMPETE            1.437      0.041     34.904      0.000
    HOMOSEX            1.468      0.042     35.047      0.000
    ABORTION           1.470      0.042     35.166      0.000
    EUTHANAS           1.075      0.036     29.531      0.000

 Variances
    ECONOMIC           1.000      0.000    999.000    999.000
    MORALS             1.000      0.000    999.000    999.000

 Residual Variances
    PRIVTOWN           0.622      0.123      5.065      0.000
    GOVTRESP           0.946      0.016     58.909      0.000
    COMPETE            0.522      0.154      3.397      0.001
    HOMOSEX            0.542      0.037     14.626      0.000
    ABORTION           0.457      0.040     11.307      0.000
    EUTHANAS           0.679      0.031     22.080      0.000


STDY Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           0.615      0.100      6.167      0.000
    GOVTRESP           0.154      0.037      4.202      0.000
    COMPETE            0.691      0.111      6.217      0.000

 MORALS   BY
    HOMOSEX            0.677      0.027     24.755      0.000
    ABORTION           0.737      0.027     26.842      0.000
    EUTHANAS           0.566      0.027     20.849      0.000
    GOVTRESP           0.176      0.034      5.241      0.000

 MORALS   WITH
    ECONOMIC          -0.005      0.046     -0.112      0.911

 Intercepts
    PRIVTOWN           1.580      0.043     36.432      0.000
    GOVTRESP           1.600      0.044     36.690      0.000
    COMPETE            1.437      0.041     34.904      0.000
    HOMOSEX            1.468      0.042     35.047      0.000
    ABORTION           1.470      0.042     35.166      0.000
    EUTHANAS           1.075      0.036     29.531      0.000

 Variances
    ECONOMIC           1.000      0.000    999.000    999.000
```

```
   MORALS           1.000      0.000    999.000    999.000

 Residual Variances
   PRIVTOWN          0.622      0.123      5.065      0.000
   GOVTRESP          0.946      0.016     58.909      0.000
   COMPETE           0.522      0.154      3.397      0.001
   HOMOSEX           0.542      0.037     14.626      0.000
   ABORTION          0.457      0.040     11.307      0.000
   EUTHANAS          0.679      0.031     22.080      0.000


STD Standardization

                                                    Two-Tailed
                Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
   PRIVTOWN          1.378      0.228      6.057      0.000
   GOVTRESP          0.416      0.100      4.159      0.000
   COMPETE           1.656      0.271      6.117      0.000

 MORALS   BY
   HOMOSEX           2.202      0.110     19.949      0.000
   ABORTION          2.185      0.102     21.355      0.000
   EUTHANAS          1.394      0.080     17.423      0.000
   GOVTRESP          0.474      0.092      5.158      0.000

 MORALS   WITH
   ECONOMIC         -0.005      0.046     -0.112      0.911

 Intercepts
   PRIVTOWN          3.541      0.065     54.608      0.000
   GOVTRESP          4.312      0.078     55.390      0.000
   COMPETE           3.442      0.069     49.763      0.000
   HOMOSEX           4.774      0.095     50.494      0.000
   ABORTION          4.360      0.086     50.780      0.000
   EUTHANAS          2.645      0.071     37.091      0.000

 Variances
   ECONOMIC          1.000      0.000    999.000    999.000
   MORALS            1.000      0.000    999.000    999.000

 Residual Variances
   PRIVTOWN          3.120      0.619      5.041      0.000
   GOVTRESP          6.872      0.289     23.738      0.000
   COMPETE           2.995      0.882      3.395      0.001
   HOMOSEX           5.731      0.398     14.409      0.000
   ABORTION          4.022      0.352     11.416      0.000
   EUTHANAS          4.115      0.212     19.419      0.000


R-SQUARE

   Observed                                         Two-Tailed
   Variable       Estimate       S.E.  Est./S.E.    P-Value

   PRIVTOWN          0.378      0.123      3.083      0.002
   GOVTRESP          0.054      0.016      3.388      0.001
   COMPETE           0.478      0.154      3.108      0.002
   HOMOSEX           0.458      0.037     12.378      0.000
   ABORTION          0.543      0.040     13.421      0.000
   EUTHANAS          0.321      0.031     10.425      0.000


QUALITY OF NUMERICAL RESULTS

    Condition Number for the Information Matrix           0.778E-03
      (ratio of smallest to largest eigenvalue)


MODEL MODIFICATION INDICES
```

```
Minimum M.I. value for printing the modification index   10.000

                       M.I.    E.P.C.  Std E.P.C.  StdYX E.P.C.

No modification indices above the minimum value.
```

The results are quite similar to those of Amos and LISREL. Under the `TESTS OF MODEL FIT` heading, the χ^2 test yields a large value of 9.920 with 7 degrees of freedom. The p-value of .1932 is too high to reject the null hypothesis of a good fit. The RMSEA and CFI are .019 and .997, indicating that the model fits the data well.

Unstandardized estimates are listed under the `MODEL RESULTS` heading and standardized estimates under `STDYX Standardization` of the `STANDARDIZED MODEL RESULTS` heading. All parameters are statistically discernable from zero; test statistics under the `Est./S.E.` column are large and corresponding p-values are small enough to reject the null hypothesis at the .05 level. Standardized estimates can be interpreted as the correlation between the latent and observed variables. GOVTRESP has low standardized loadings of .154 and .176 on ECONOMIC and MORALS, respectively; only 5.4 percent of variance in GOVTRESP is accounted for by the two latent variables. Other factor loadings range from .566 for EUTHANAS (R^2 = .321) and .737 for ABORTION (R^2 = .543). The covariance and correlation between the two common factors are negligible -.016 and -.005, respectively.

4.5 Summary

Table 6 compares the unstandardized and standardized estimates that each statistical software package produced. Standard errors appear in parentheses. The final row lists goodness of fit measures of χ^2, RMSEA, and CFI. The results are essentially identical across programs.

Table 6: Two Factor Model with Missing Data (N=1,200, DF=7)

	Amos		LISREL		Mplus	
	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard
F1 → PRIVTOWN	1.00	.615	1.00	.62	1.000	.615 (.100)
F1 → GOVTRESP	.302 (.082)	.154	.30 (.082)	.15	.302 (.082)	.154 (.037)
F1 → COMPETE	1.200 (.369)	.691	1.20 (.37)	.69	1.201 (.384)	.691 (.111)
F2 → HOMOSEX	1.00	.677	1.00	.68	1.000	.677 (.027)
F2 → ABORTION	.992 (.071)	.737	.99 (.071)	.74	.992 (.072)	.737 .027)
F2 → EUTHANAS	.633 (.045)	.566	.63 (.045)	.57	.633 (.046)	.566 (.027)
F2 → GOVTRESP	.215 (.042)	.176	.21 (.042)	.18	.215 (.042)	.176 (.034)
F1 ↔ F2	-.012[a] (.102)	-.005[b]	-.02[a] (.13)	-.01[b]	-.016[a] (.140)	-.005[b] (.046)
χ^2, RMSEA, CFI	9.911 .019 .997			9.92 .019		9.920 .019 .997

Standard errors appear in parentheses
[a] covariance; [b] correlation between F1 (economic values) and F2 (moral values)

5. CFA with Categorical indicators

This section estimates a confirmatory factor model using the polychoric correlation matrix in the presence of categorical manifest variables. Latent variables are assumed to represent continuous (not categorical) constructs. For illustration, original responses ranging from 1 to 3 were recoded as 1; those ranging from 4 to 7 were recoded as 2; and those ranging from 8 to 10 were recoded as 3. Forty observations with missing data are dropped in order to focus only on the problem of categorical outcome variables in the confirmatory factor model. The recoded data are stored into the SPSS file *values_ord.sav* (N=1,160) in the C:\temp\CFA directory. *Values_full.sav* (N=1,200) is also used to show how pairwise deletion usesas much information in the raw data file as possible.

5.1 Background

The maximum likelihood estimation (MLE) approach used in the previous sections relied on the strong assumption of multivariate normality. In practice, a substantial amount of social science data is non-normal. Survey responses are often coded as yes/no or as scores on an ordered scale (e.g. strongly disagree, disagree, neutral, agree, strongly agree). In the presence of categorical or ordinal data, MLE may not work properly, calling for alternative estimation methods.

Mplus, LISREL, and SAS/STAT CALIS employ a multi-step method for ordinal outcome variables that analyzes a matrix of polychoric correlations rather than covariances. This approach works as follows: 1) thresholds are estimated by maximum likelihood, 2) these estimates are used to estimate a polychoric correlation matrix, which in turn is used to 3) estimate parameters through (diagonally) weighted least squares using the inverse of the asymptotic covariance matrix as the weight matrix (Muthén, 1984; Jöreskog, 1990).

In LISREL, the diagonally weighted least squares (DWLS) method needs to be specified. Alternatively, the polychoric correlation matrix and asymptotic covariance matrix is estimated and saved into a LISREL system file (.dsf) using PRELIS before fitting the model. Mplus automatically follows above steps when the syntax includes a line identifying observed variables as categorical. In SAS/STAT CALIS, the `METHOD=WLS` or `METHOD=DWLS` option of the PROC CALIS statement tells SAS to fit the model using the WLS or DWLS method instead of MLE.

5.2 CFA with Categorical Indicators Using LISREL

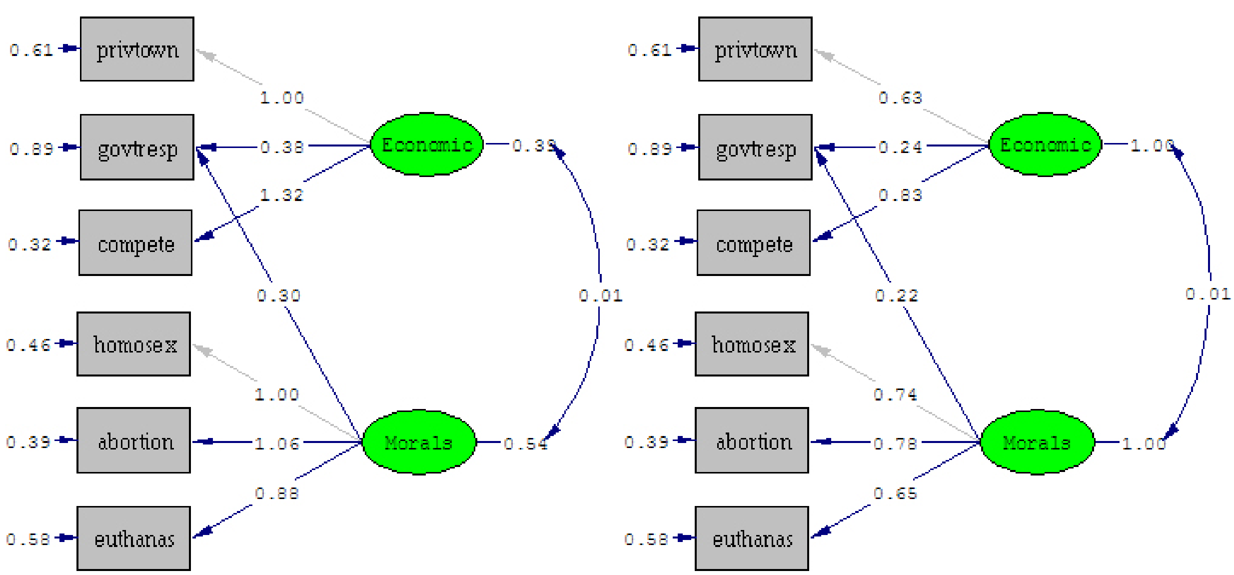
Launch LISREL and load the data by choosing **File → Import Data**. Change the file type to SPSS Data File (*.sav) in the **Open** dialog box, navigate to the C:\temp\CFA folder, choose the file *values_ord.sav*, and click **Open**. When prompted from the **Save As** dialog box, enter the name *values_ord* in the **File Name** field and click **Save**. The data will then be displayed in spreadsheet form.[11]

[11] You may explicitly define the variables as ordinal by going to **Data → Define Variables**, highlighting all the variables, and opening the **Variable Type** menu.

To open an empty text window, go to **File** → **New**, choose **SIMPLIS Project**, and click **OK**. When prompted in the **Save As** dialog box, type in *values_ord*, and then click **Save**. Enter the following commands and save the file. Notice that `Method of Estimation: Diagonally Weighted Least Squares` tells LISREL to fit WLS for categorical observed variables.[12]

```
Two factor model CFA for Ordinal Indicators
SYSTEM FILE from file 'c:\Temp\CFA\values_ord.dsf'
Latent Variables Economic Morals
Relationships
privtown = 1.00*Economic
govtresp = Economic Morals
compete = Economic
homosex = 1.00*Morals
abortion = Morals
euthanas = Morals
Path Diagram
Method of Estimation: Diagonally Weighted Least Squares
End of Problem
```

Click **F5** or the **Run LISREL** button to fit the model (alternatively click on **File** → **Run**). The unstandardized solution appears in the path diagram along with χ^2 and RMSEA statistics. To view the standardized estimates, choose **Standardized Solution** from the **Estimates** drop-down menu. The following display path diagrams for unstandardized (left) and standardized (right) estimates.



Detailed information can be read from the text output file *values_ordinal.out*. The following is selected from the output file.

```
Sample Size =  1160

Two factor model CFA for Ordinal Indicators
```

[12] To request the weighted least squares estimator on menu, go to **Output** → **SIMPLIS Outputs**. Choose **Diagonally Weighted Least Squares**.

```
       Correlation Matrix

          privtown   govtresp    compete    homosex   abortion   euthanas
          --------   --------   --------   --------   --------   --------
privtown      1.00
govtresp      0.15       1.00
 compete      0.52       0.20       1.00
 homosex     -0.05       0.18      -0.03       1.00
abortion     -0.02       0.16      -0.01       0.58       1.00
euthanas      0.08       0.15       0.11       0.47       0.51       1.00

Two factor model CFA for Ordinal Indicators

Number of Iterations =  6

LISREL Estimates (Robust Diagonally Weighted Least Squares)


       Measurement Equations


privtown = 1.00*Economic, Errorvar.= 0.61 , R² = 0.39
                                (0.13)
                                 4.61

govtresp = 0.38*Economic + 0.30*Morals, Errorvar.= 0.89  , R² = 0.11
         (0.079)        (0.055)               (0.065)
          4.77           5.51                 13.78

 compete = 1.32*Economic, Errorvar.= 0.32 , R² = 0.68
         (0.37)                 (0.20)
          3.56                   1.60

 homosex = 1.00*Morals, Errorvar.= 0.46  , R² = 0.54
                              (0.078)
                               5.87

abortion = 1.06*Morals, Errorvar.= 0.39  , R² = 0.61
         (0.078)              (0.080)
          13.56                4.91

euthanas = 0.88*Morals, Errorvar.= 0.58  , R² = 0.42
         (0.068)              (0.077)
          12.93                7.56


       Covariance Matrix of Independent Variables

          Economic    Morals
          --------   --------
Economic      0.39
            (0.12)
              3.34

  Morals      0.01       0.54
            (0.02)     (0.05)
              0.30      10.68


                       Goodness of Fit Statistics

                          Degrees of Freedom = 7
      Normal Theory Weighted Least Squares Chi-Square = 33.75 (P = 0.00)
           Satorra-Bentler Scaled Chi-Square = 13.43 (P = 0.062)
         Chi-Square Corrected for Non-Normality = 13.93 (P = 0.052)
             Estimated Non-centrality Parameter (NCP) = 6.43
           90 Percent Confidence Interval for NCP = (0.0 ; 20.89)

                    Minimum Fit Function Value = 0.010
```

```
           Population Discrepancy Function Value (F0) = 0.0055
          90 Percent Confidence Interval for F0 = (0.0 ; 0.018)
         Root Mean Square Error of Approximation (RMSEA) = 0.028
         90 Percent Confidence Interval for RMSEA = (0.0 ; 0.051)
           P-Value for Test of Close Fit (RMSEA < 0.05) = 0.94

              Expected Cross-Validation Index (ECVI) = 0.036
        90 Percent Confidence Interval for ECVI = (0.030 ; 0.048)
                     ECVI for Saturated Model = 0.036
                    ECVI for Independence Model = 1.25

  Chi-Square for Independence Model with 15 Degrees of Freedom = 1440.45
                        Independence AIC = 1452.45
                            Model AIC = 41.43
                          Saturated AIC = 42.00
                       Independence CAIC = 1488.79
                           Model CAIC = 126.22
                         Saturated CAIC = 169.18

                      Normed Fit Index (NFI) = 0.99
                    Non-Normed Fit Index (NNFI) = 0.99
                 Parsimony Normed Fit Index (PNFI) = 0.46
                    Comparative Fit Index (CFI) = 1.00
                    Incremental Fit Index (IFI) = 1.00
                     Relative Fit Index (RFI) = 0.98

                        Critical N (CN) = 1595.43


                 Root Mean Square Residual (RMR) = 0.032
                         Standardized RMR = 0.032
                    Goodness of Fit Index (GFI) = 1.00
               Adjusted Goodness of Fit Index (AGFI) = 0.99
              Parsimony Goodness of Fit Index (PGFI) = 0.33

      The Modification Indices Suggest to Add the
 Path to  from     Decrease in Chi-Square   New Estimate
homosex   Economic         81.6             -3.40
euthanas  Economic         21.3              0.51

The Modification Indices Suggest to Add an Error Covariance
 Between   and    Decrease in Chi-Square   New Estimate
euthanas  compete          8.2              0.16
```

This model appears to fit the data moderately. The Satorra-Bentler Scaled χ^2 statistic has a large value of 13.43 with seven degrees of freedom. But its corresponding p-value is .062; we do not reject the null hypothesis of a good fit marginally at the .05 level. However, RMSEA and CFI are respectively .028 and 1.00, indicating a very good model fit.

Under the `Measurement Equations` heading appear the unstandardized estimates, standard errors, t-values, and R^2 statistics. All of the unconstrained path coefficients are large enough to be statistically significant. Despite receiving a path from both ECONOMIC and MORALS, GOVTRESP has the smallest R^2 (.11). The other observed variables have moderate to high R^2 statistics, ranging from .39 (PRIVTOWN) to .68 (COMPETE).

The standardized estimates of the loadings displayed in the path diagram range from .22 (MORALS) and .24 (ECONOMIC) for GOVTRESP to .63 for PRIVTOWN and .83 for COMPETE. This result a bit differs from one in previous models. Finally, the covariance and correlation between the two latent variables are .01, which is not statistically discernable from zero.

5.3 CFA with Categorical Indicators using Mplus

Mplus cannot directly read data from an SPSS system file. Raw data must come from a free or fixed format text file. See sections 4.4 for syntax to translate an SPSS file to an ASCII file. For this example the data is saved as the tab-delimited file *values_ord.dat* in the C:\temp\CFA folder.

After launching Mplus, the syntax editor appears. The following commands are used to estimate the confirmatory factor model with ordinal observed variables.

```
TITLE:     Two factor model CFA with ordinal indicators (listwise);
DATA:      FILE IS c:\Temp\CFA\values_ord.dat;
VARIABLE:  NAMES ARE privtown govtresp compete
           homosex abortion euthanas;
           CATEGORICAL ARE privtown govtresp compete
           homosex abortion euthanas;
MODEL:     economic BY privtown govtresp compete;
           morals BY homosex abortion euthanas govtresp;
OUTPUT:    Standardized;
```

The syntax is similar to previous sections except that CATEGORICAL ARE is added to the VARIABLE statement to define the observed variables as categorical. Mplus by default assumes all observed variables are continuous unless otherwise specified. When categorical indicators are declared, Mplus employs by default a robust weighted least squares estimator similar to the Diagonally Weighted Least Squares estimator in LISREL. Click on **Run** to carry out the estimation. A selection of the output file is the following:

```
SUMMARY OF ANALYSIS

Number of groups                                         1
Number of observations                                1160

Number of dependent variables                              6
Number of independent variables                            0
Number of continuous latent variables                      2

Observed dependent variables

  Binary and ordered categorical (ordinal)
   PRIVTOWN    GOVTRESP    COMPETE     HOMOSEX     ABORTION    EUTHANAS

Continuous latent variables
   ECONOMIC    MORALS


Estimator                                              WLSMV
Maximum number of iterations                               1000
Convergence criterion                                 0.500D-04
Maximum number of steepest descent iterations                20
Parameterization                                        DELTA

Input data file(s)
  c:\Temp\CFA\values_ord.dat

Input data format  FREE


SUMMARY OF CATEGORICAL DATA PROPORTIONS
    …
```

```
THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

        Value                           15.516*
        Degrees of Freedom                    6**
        P-Value                          0.0166

*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
    for chi-square difference tests.  MLM, MLR and WLSM chi-square difference
    testing is described in the Mplus Technical Appendices at www.statmodel.com.
    See chi-square difference testing in the index of the Mplus User's Guide.

**  The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to
    a formula given in the Mplus Technical Appendices at www.statmodel.com.
    See degrees of freedom in the index of the Mplus User's Guide.

Chi-Square Test of Model Fit for the Baseline Model

        Value                         1054.754
        Degrees of Freedom                  11
        P-Value                         0.0000

CFI/TLI

        CFI                              0.991
        TLI                              0.983

Number of Free Parameters                     20

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                         0.037

WRMR (Weighted Root Mean Square Residual)

        Value                            0.671


MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           1.000      0.000    999.000    999.000
    GOVTRESP           0.375      0.073      5.166      0.000
    COMPETE            1.313      0.332      3.949      0.000

 MORALS   BY
    HOMOSEX            1.000      0.000    999.000    999.000
    ABORTION           1.059      0.071     14.866      0.000
    EUTHANAS           0.878      0.058     15.040      0.000
    GOVTRESP           0.304      0.055      5.534      0.000

 MORALS   WITH
    ECONOMIC           0.006      0.022      0.254      0.800

 Thresholds
    …

 Variances
    ECONOMIC           0.394      0.105      3.765      0.000
    MORALS             0.543      0.045     12.113      0.000


STANDARDIZED MODEL RESULTS
```

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ECONOMIC BY				
PRIVTOWN	0.628	0.083	7.530	0.000
GOVTRESP	0.236	0.041	5.788	0.000
COMPETE	0.825	0.104	7.936	0.000
MORALS BY				
HOMOSEX	0.737	0.030	24.226	0.000
ABORTION	0.780	0.030	25.647	0.000
EUTHANAS	0.647	0.035	18.679	0.000
GOVTRESP	0.224	0.040	5.654	0.000
MORALS WITH				
ECONOMIC	0.012	0.048	0.254	0.799
Thresholds				
…				
Variances				
ECONOMIC	1.000	0.000	999.000	999.000
MORALS	1.000	0.000	999.000	999.000

STDY Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ECONOMIC BY				
PRIVTOWN	0.628	0.083	7.530	0.000
GOVTRESP	0.236	0.041	5.788	0.000
COMPETE	0.825	0.104	7.936	0.000
MORALS BY				
HOMOSEX	0.737	0.030	24.226	0.000
ABORTION	0.780	0.030	25.647	0.000
EUTHANAS	0.647	0.035	18.679	0.000
GOVTRESP	0.224	0.040	5.654	0.000
MORALS WITH				
ECONOMIC	0.012	0.048	0.254	0.799
Thresholds				
…				
Variances				
ECONOMIC	1.000	0.000	999.000	999.000
MORALS	1.000	0.000	999.000	999.000

STD Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ECONOMIC BY				
PRIVTOWN	0.628	0.083	7.530	0.000
GOVTRESP	0.236	0.041	5.788	0.000
COMPETE	0.825	0.104	7.936	0.000
MORALS BY				
HOMOSEX	0.737	0.030	24.226	0.000
ABORTION	0.780	0.030	25.647	0.000
EUTHANAS	0.647	0.035	18.679	0.000
GOVTRESP	0.224	0.040	5.654	0.000
MORALS WITH				

```
   ECONOMIC          0.012      0.048      0.254      0.799

 Thresholds
   …

 Variances
   ECONOMIC          1.000      0.000    999.000    999.000
   MORALS            1.000      0.000    999.000    999.000


R-SQUARE

   Observed                                        Two-Tailed   Residual
   Variable       Estimate       S.E.  Est./S.E.    P-Value    Variance

   PRIVTOWN          0.394      0.105      3.765      0.000      0.606
   GOVTRESP          0.107      0.025      4.200      0.000      0.893
   COMPETE           0.680      0.171      3.968      0.000      0.320
   HOMOSEX           0.543      0.045     12.113      0.000      0.457
   ABORTION          0.608      0.047     12.824      0.000      0.392
   EUTHANAS          0.418      0.045      9.339      0.000      0.582


QUALITY OF NUMERICAL RESULTS

    Condition Number for the Information Matrix           0.103E-02
      (ratio of smallest to largest eigenvalue)
```

Both LISREL and Mplus (listwise deletion) produce virtually identical results. The χ^2 test yields a value of 15.516, but a corresponding p-value is .0166, which is small enough to reject the null hypothesis of a good fit at the .05 significance level. As explained in the output, Mplus estimates 6 degrees of freedom, which is one smaller than those reported by LISREL. However, a small RMSEA of .037 and a large CFI of .991 suggest that the model fits the data well as a whole.

Under the `MODEL RESULTS` heading, the unstandardized loadings appear along with standard errors, the ratio of the estimates to their standard errors, and p-values. Since all absolute values of the numbers in the `Est./S.E.` column are greater than 1.96, all of the unconstrained loadings estimates are significant at the .05 level.

Standardized estimates are provided under the `STDYX Standardization` heading. GOVTRESP has relatively low standardized loadings on both factors (.236 for ECONOMIC and .224 for MORALS). For PRIVTOWN the loading is .628, for COMPETE it is .825, for HOMOSEX it is .737, for ABORTION it is .780, and for EUTHANAS it is .647. GOVTRESP has a low squared multiple correlation of .107. The remaining R^2 statistics are, in order of increasing magnitude, .394 (PRIVTOWN), .418 (EUTHANAS), .543 (HOMOSEX), .608 (ABORTION), and .680 (COMPETE). Finally, the covariance and correlation between the two common factors are .006 and .012, respectively.

5.4 CFA with categorical indicators and missing data

The previous two subsections explained how to estimate the confirmatory factor model when the observed variables represent ordered categories. A total of 40 observations with missing values on at least one indicator were dropped in the listwise deletion, reducing the original sample of 1,200 to 1,160. It is possible to maximize the information available in the raw data file using pairwise rather than listwise deletion for missing data. In pairwise deletion, correlations will be estimated using all observations with complete data available on both variables. The following Mplus syntax use *values_full.dat* instead of *values_ord.dat*.

```
TITLE:     Two factor model CFA with ordinal indicators (pairwise);
DATA:      FILE IS c:\Temp\CFA\values_full.dat;
VARIABLE:   NAMES ARE privtown govtresp compete
          homosex abortion euthanas;
          CATEGORICAL ARE privtown govtresp compete
          homosex abortion euthanas;
          MISSING ARE all (-1);
MODEL:     economic BY privtown govtresp compete;
          morals BY homosex abortion euthanas govtresp;
OUTPUT:    Standardized;
```

Notice that `MISSING ARE all (-1)` is added to the `VARIABLE` command in order to use pairwise deletion. The following is a selection of Mplus output.

```
SUMMARY OF ANALYSIS

Number of groups                                          1
Number of observations                                 1200

Number of dependent variables                               6
Number of independent variables                             0
Number of continuous latent variables                       2

Observed dependent variables

  Binary and ordered categorical (ordinal)
   PRIVTOWN    GOVTRESP    COMPETE     HOMOSEX     ABORTION    EUTHANAS

Continuous latent variables
   ECONOMIC    MORALS


Estimator                                              WLSMV
Maximum number of iterations                               1000
Convergence criterion                                 0.500D-04
Maximum number of steepest descent iterations                20
Maximum number of iterations for H1                        2000
Convergence criterion for H1                          0.100D-03
Parameterization                                        DELTA

Input data file(s)
  c:\Temp\CFA\values_full.dat

Input data format  FREE



SUMMARY OF DATA

     Number of missing data patterns          12


COVARIANCE COVERAGE OF DATA

Minimum covariance coverage value   0.100

     PROPORTION OF DATA PRESENT

          Covariance Coverage
            PRIVTOWN      GOVTRESP      COMPETE       HOMOSEX       ABORTION
```

```
              ________      ________      ________      ________      ________
 PRIVTOWN      0.994
 GOVTRESP      0.994         0.999
 COMPETE       0.994         0.999         0.999
 HOMOSEX       0.977         0.981         0.981         0.982
 ABORTION      0.988         0.992         0.992         0.976         0.993
 EUTHANAS      0.985         0.990         0.990         0.975         0.985


           Covariance Coverage
              EUTHANAS
              ________
 EUTHANAS      0.991


THE MODEL ESTIMATION TERMINATED NORMALLY


TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                             10.417*
          Degrees of Freedom                     6**
          P-Value                           0.1081

*   The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used
    for chi-square difference tests.  MLM, MLR and WLSM chi-square difference
    testing is described in the Mplus Technical Appendices at www.statmodel.com.
    See chi-square difference testing in the index of the Mplus User's Guide.

**  The degrees of freedom for MLMV, ULSMV and WLSMV are estimated according to
    a formula given in the Mplus Technical Appendices at www.statmodel.com.
    See degrees of freedom in the index of the Mplus User's Guide.

Chi-Square Test of Model Fit for the Baseline Model

          Value                           1528.615
          Degrees of Freedom                    10
          P-Value                           0.0000

CFI/TLI

          CFI                                0.997
          TLI                                0.995

Number of Free Parameters                       62

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.025

WRMR (Weighted Root Mean Square Residual)

          Value                              0.336



MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 ECONOMIC BY
    PRIVTOWN           1.000      0.000    999.000    999.000
    GOVTRESP           0.380      0.051      7.416      0.000
    COMPETE            1.352      0.258      5.245      0.000

 MORALS   BY
    HOMOSEX            1.000      0.000    999.000    999.000
    ABORTION           1.066      0.053     20.239      0.000
    EUTHANAS           0.930      0.045     20.684      0.000
```

```
   GOVTRESP         0.272     0.044     6.207     0.000

 MORALS   WITH
   ECONOMIC         0.014     0.018     0.801     0.423

 Thresholds
   …

 Variances
   ECONOMIC         0.354     0.070     5.060     0.000
   MORALS           0.520     0.033    15.803     0.000


STANDARDIZED MODEL RESULTS

STDYX Standardization

                                                Two-Tailed
                Estimate      S.E.  Est./S.E.    P-Value

 ECONOMIC BY
   PRIVTOWN         0.595     0.059    10.120     0.000
   GOVTRESP         0.226     0.029     7.818     0.000
   COMPETE          0.805     0.077    10.394     0.000

 MORALS   BY
   HOMOSEX          0.721     0.023    31.607     0.000
   ABORTION         0.769     0.023    33.214     0.000
   EUTHANAS         0.670     0.027    25.218     0.000
   GOVTRESP         0.196     0.031     6.236     0.000

 MORALS   WITH
   ECONOMIC         0.033     0.040     0.811     0.417

 Thresholds
   …

 Variances
   ECONOMIC         1.000     0.000   999.000   999.000
   MORALS           1.000     0.000   999.000   999.000


R-SQUARE

   Observed                                     Two-Tailed   Residual
   Variable      Estimate      S.E.  Est./S.E.    P-Value    Variance

   PRIVTOWN         0.354     0.070     5.060     0.000      0.646
   GOVTRESP         0.093     0.018     5.241     0.000      0.907
   COMPETE          0.648     0.125     5.197     0.000      0.352
   HOMOSEX          0.520     0.033    15.803     0.000      0.480
   ABORTION         0.591     0.036    16.607     0.000      0.409
   EUTHANAS         0.449     0.036    12.609     0.000      0.551
```

This CFA model fits the data quite well. The χ^2 of 10.417 with 6 degrees of freedom and its p-value of .1081 do not reject the null hypothesis of a good fit. A small RMSEA of .025 and a large CFI of .997 also suggest that the model fits the data well.

All unstandardized and standardized estimates have large test statistics and appear to be statistically significant. GOVTRESP has lowest factor loadings of .226 from economic values and .196 for moral values. Other factor loadings range from .595 (PRIVTOWN) to .805 (COMPETE). Accordingly, R^2 ranges from .093 for GOVTRESP to .354 for PRIVTOWN and .648 for COMPETE. The covariance and correlation between two latent variables are .014 and .033, respectively.

5.5 CFA with categorical indicators Using SAS/STAT CALIS

SAS/STAT CALIS can also fit a confirmatory factor model with categorical indicators using (diagonally) weighted least squares estimation. In SPSS, you may convert the data set *values_ord.sav* into the SAS data set by choosing **Save As** and select **SAS v7+Windows long extension (*.sas7bdat)** as a file type.

The following is the SAS script for this two factor CFA model with ordinal data. Notice that `METHOD=WLS` in the PROC CALIS statement tells SAS to employ the WLS method instead of the default MLE.[13] Other statements for specification remain unchanged.

```
ODS HTML;

PROC CALIS DATA=sas.values_ord METHOD=WLS PALL;
VAR privtown govtresp compete homosex abortion euthanas;
LINEQS      privtown = 1.0 f1 + e1,
            govtresp = lambda2 f1 + lambda7 f2 + e2,
            compete = lambda3 f1 + e3,
            homosex = 1.0 f2 + e4,
            abortion = lambda5 f2 + e5,
            euthanas = lambda6 f2 + e6;
STD         f1-f2 = phi1-phi2,
            e1-e6 = theta1-theta6;
COV         f1-f2 = phi3;
RUN;

ODS HTML CLOSE;
```

The following is a selection of the SAS output. SAS reports 21 knowns provided by the input covariance matrix and 14 unknown parameters to be estimated; degrees of freedom are 7 = 21 - 14. Basic statistics and correlation matrix display first.

The CALIS Procedure
Covariance Structure Analysis: Weighted Least-Squares Estimation

Observations	1160	**Model Terms**	1
Variables	6	**Model Matrices**	4
Informations	21	**Parameters**	14

Variable	Mean	Std Dev	Skewness	Kurtosis
privtown	1.52672	0.60852	0.70316	-0.47100
govtresp	1.68103	0.71347	0.55257	-0.89216

[13] `METHOD=DWLS` uses diagonally weighted least squares method but, despite successful convergence, failed to report goodness of fit measures in this example.

Variable	Mean	Std Dev	Skewness	Kurtosis
compete	1.49138	0.63267	0.92321	-0.21119
homosex	1.85345	0.78108	0.26214	-1.31825
abortion	1.74052	0.74246	0.45744	-1.07136
euthanas	1.33276	0.60023	1.62442	1.49187

Correlations						
	privtown	govtresp	compete	homosex	abortion	euthanas
privtown	1.0000	0.1051	0.3917	-0.0390	-0.0123	0.0489
govtresp	0.1051	1.0000	0.1315	0.1421	0.1205	0.0909
compete	0.3917	0.1315	1.0000	-0.0218	-0.0057	0.0689
homosex	-0.0390	0.1421	-0.0218	1.0000	0.4670	0.3305
abortion	-0.0123	0.1205	-0.0057	0.4670	1.0000	0.3604
euthanas	0.0489	0.0909	0.0689	0.3305	0.3604	1.0000

Fit Function	0.0114
Goodness of Fit Index (GFI)	1.0000
GFI Adjusted for Degrees of Freedom (AGFI)	0.9999
Root Mean Square Residual (RMR)	0.0220
Parsimonious GFI (Mulaik, 1989)	0.4667
Chi-Square	13.2625
Chi-Square DF	7
Pr > Chi-Square	0.0660
Independence Model Chi-Square	674.00
Independence Model Chi-Square DF	15
RMSEA Estimate	0.0278
RMSEA 90% Lower Confidence Limit	.
RMSEA 90% Upper Confidence Limit	0.0504
ECVI Estimate	0.0356
ECVI 90% Lower Confidence Limit	.
ECVI 90% Upper Confidence Limit	0.0480
Probability of Close Fit	0.9462
Bentler's Comparative Fit Index	0.9905
Akaike's Information Criterion	-0.7375

Bozdogan's (1987) CAIC	-43.1308
Schwarz's Bayesian Criterion	-36.1308
McDonald's (1989) Centrality	0.9973
Bentler & Bonett's (1980) Non-normed Index	0.9796
Bentler & Bonett's (1980) NFI	0.9803
James, Mulaik, & Brett (1982) Parsimonious NFI	0.4575
Z-Test of Wilson & Hilferty (1931)	1.5106
Bollen (1986) Normed Index Rho1	0.9578
Bollen (1988) Non-normed Index Delta2	0.9906
Hoelter's (1983) Critical N	1231

Manifest Variable Equations with Estimates

privtown	=	1.0000		f1	+	1.0000		e1		
govtresp	=	0.3609	*	f1	+	0.2768	*	f2	+ 1.0000	e2
Std Err		0.0831		lambda2		0.0546		lambda7		
t Value		4.3441				5.0658				
compete	=	1.2040	*	f1	+	1.0000		e3		
Std Err		0.3369		lambda3						
t Value		3.5742								
homosex	=	1.0000		f2	+	1.0000		e4		
abortion	=	1.0367	*	f2	+	1.0000		e5		
Std Err		0.0811		lambda5						
t Value		12.7824								
euthanas	=	0.7616	*	f2	+	1.0000		e6		
Std Err		0.0633		lambda6						
t Value		12.0318								

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
f1	**phi1**	0.31984	0.09545	3.35
f2	**phi2**	0.45389	0.04364	10.40
e1	**theta1**	0.68016	0.09550	7.12
e2	**theta2**	0.92369	0.02047	45.12
e3	**theta3**	0.53632	0.13081	4.10
e4	**theta4**	0.54611	0.04374	12.48
e5	**theta5**	0.51219	0.04580	11.18
e6	**theta6**	0.73673	0.03423	21.52

Covariances Among Exogenous Variables

Var1	Var2	Parameter	Estimate	Standard Error	t Value
f1	f2	phi3	-0.0006067	0.01846	-0.03

Manifest Variable Equations with Standardized Estimates

```
privtown = 0.5655    f1       + 0.8247    e1
govtresp = 0.2041 *  f1       + 0.1865 *  f2       + 0.9611    e2
                     lambda2              lambda7
compete  = 0.6809 *  f1       + 0.7323    e3
                     lambda3
homosex  = 0.6737    f2       + 0.7390    e4
abortion = 0.6984 *  f2       + 0.7157    e5
                     lambda5
euthanas = 0.5131 *  f2       + 0.8583    e6
                     lambda6
```

Squared Multiple Correlations

	Variable	Error Variance	Total Variance	R-Square
1	privtown	0.68016	1.00000	0.3198
2	govtresp	0.92369	1.00000	0.0763
3	compete	0.53632	1.00000	0.4637
4	homosex	0.54611	1.00000	0.4539
5	abortion	0.51219	1.00000	0.4878
6	euthanas	0.73673	1.00000	0.2633

Correlations Among Exogenous Variables

Var1	Var2	Parameter	Estimate
f1	f2	phi3	-0.00159

The results are a bit different from those of LISREL and Mplus. The χ^2 is 13.2625 with 7 degrees of freedom and a corresponding p-value is .0660 that indicates a moderate fit. The RMSEA of .0278 and CFI of .9905, however, suggest that this confirmatory factor model fits the date well. This result is similar to what LISREL produced in 5.2.

Unstandardized estimates under the `Manifest Variable Equations with Estimates` heading have large t statistics and appear to be statistically discernable from zero. Standardized estimates display under `Manifest Variable Equations with Standardize d Estimates` Factor loadings for GOVTRESP are .2041 for economic values and .1865 for moral values. Other factor loadings range from .5131 for EUTHANAS to .6984 for ABORTION. Squared multiple correlations range from .0763 for GOVTRESP to .2633 for EUTHANAS and .4878 for ABORTION. The covariance and correlation between two latent variables are -.0006 and

-.0016, respectively. Unlike LISREL and Mplus, SAS/STAT CALIS reports negative covariance and correlation although their magnitudes are virtually zero.

5.6 Summary

Table 7 summarizes results from the LISREL, Mplus, and SAS/STAT CALIS. LISREL and Mplus with listwise deletion produces equal parameter estimates but Mplus reports smaller standard errors. Mplus also supports pairwise deletion analyses using listwise deletion plus the Mplus results using pairwise deletion that produces slightly different parameter estimates and standard errors. The pairwise column should be considered more accurate because it is able to incorporate the most information from the raw data; notice that pairwise deletion uses 1,200 observations without dropping 40 incomplete observations. SAS/STAT CALIS reports goodness of fit measures similar to those LISREL produces, but unstandardized and standardized estimates of two software packages are slightly different.

LISREL and SAS/STAT CALIS reports a χ^2 of about 13, which is different from 10 and 16 in Mplus. LISREL and SAS/STAT CALIS have a marginal p-value of about .06 and thus do not reject the null hypothesis of a good fit at the .05 significance level. In Mplus, CFA model with listwise deletion rejects the null hypothesis at the .05 level, while CFA with pairwise deletion does not. RMSEA and CFI, however, are relatively consistent regardless of models estimated by three software packages, suggesting that all models fit the data well. Notice that Mplus uses a different formula to calculate the degrees of freedom; LISREL and SAS/STAT CALIS report 7 degrees of freedom whereas Mplus returns 6.

Table 7: Two Factor Model with Ordinal Indicators

	LISREL (listwise)		Mplus (listwise)		Mplus (pairwise)		SAS CALIS	
	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard	Unstd.	Standard
F1 → PRIVTOWN	1.00	.63	1.00	.628 (.083)	1.000	.595 (.059)	1.0000	.5655
F1 → GOVTRESP	.38 (.079)	.24	.375 (.073)	.236 (.041)	.380 (.051)	.226 (.029)	.3609 (.0831)	.2041
F1 → COMPETE	1.32 (.37)	.83	1.313 (.332)	.825 (.104)	1.352 (.258)	.805 (.077)	1.2040 (.3369)	.6809
F2 → HOMOSEX	1.00	.74	1.00	.737 (.030)	1.000	.721 (.023)	1.0000	.6737
F2 → ABORTION	1.06 (.078)	.78	1.059 (.071)	.780 (.030)	1.066 (.053)	.769 (.023)	1.0367 (.0811)	.6984
F2 → EUTHANAS	.88 (.068)	.65	.878 (.058)	.647 (.035)	.930 (.045)	.670 (.027)	.7616 (.0633)	.5131
F2 → GOVTRESP	.30 (.055)	.22	.304 (.055)	.224 (.040)	.272 (.044)	.196 (.031)	.2768 (.0546)	.1865
F1 ←→ F2	.01[a] (.02)	.01[b]	.006[a] (.022)	.012[b] (.048)	.014[a] (.018)	.033[b] (.040)	-.0006[a] (.0185)	-.0016[b]
Degrees of freedom	(N=1,160)	7	(N=1,160)	6	(N=1,200)	6	(N=1,160)	7
χ^2, RMSEA, CFI		13.93 .028 1.00		15.516 .037 .991		10.417 .025 .997		13.2625 .0278 .9905

Standard errors appear in parentheses
[a] covariance; [b] correlation between F1 (economic values) and F2 (moral values)

6. Conclusion

Factor analysis is a widely used method for situations in which a small set of unobserved (latent) variables is believed to underlie a larger set of observed (manifest) variables. Exploratory factor analysis (EFA), available in most general statistics packages, is a technique to identify structure in data and generating hypotheses without imposing any restrictions. EFA differ from confirmatory factor analysis (CFA) in that CFA is much more theory driven (as opposed to data driven) and is generally used to test explicit hypotheses.

CFA is the basis of the measurement model in full structural equation modeling (SEM) and can be estimated using SEM software. Amos and LISREL are the most user-friendly, although Mplus syntax is not at all difficult to learn. Amos and LISREL can read data files generated by a variety of different software packages and can estimate models by simply drawing a path diagram. SAS/STAT CALIS has the systematic grammar structure in the integrated environment, but it has messy output and does not support advanced models with missing values. LISREL, Mplus, and SAS/STAT CALIS can handle ordinal observed variables using (diagonally) weighted least squares estimation methods. Amos cannot accurately estimate models when the observed variables are categorical. All four software packages handle models assuming the latent variable to be continuous, although Mplus can also estimate models in which the latent variables are assumed to be categorical.

Mplus is recommended for its simplicity of syntax, informative output, and support for various advanced models although it has poor data management capability and does not support a path diagram. LISREL is generally suggested to take advantage of switching back and forth from a path diagram to SIMPLIS and LISREL. By contrast, Amos Graphics and Program Editor work independently and the output, although rich enough, is not easy to navigate.

Consult the documentation for the respective package for additional information on Amos, LISREL, and Mplus. Additionally IU students, staff, and faculty may schedule an appointment with a consultant at the UITS Stat/Math Center by calling 5-4724 or emailing statmath@indiana.edu.

Acknowledgements

Amy Drayton and Takuya Noguchi offered helpful suggestions to make this document more informative and readable, though they are not responsible for any errors that remain.

References

Akaike, H. (1987), "Factor analysis and AIC," *Psychometrika*, 52, 317-332.
Bollen, K.A. (1980), "Issues in the comparative measurement of political democracy" *American Sociological Review*, 45, 370-390.
Bollen, K.A. (1989), *Structural Equation Models with Latent Variables*, New York: Wiley & Sons.
Bollen, K.A. and Long, J.S., eds. (1993), *Testing Structural Equation Models*, Newbury Park, CA: Sage.
Brown, T. A. (2006), *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Press.
Carmines, E.G. and Zeller, R.A. (1979), *Reliability and Validity Assessment*, Beverly Hills, CA: Sage.
Enders, C.K. (2001), "A primer on maximum likelihood algorithms for use with missing data. *Structural Equation Modeling*, 8, 128-141.
European Values Study Group and World Values Survey Association. (2005), European and World Values Surveys Integrated Data File, 1999-2002 Release I, ICPSR3975 [computer file], Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-06-15.
Gabel, M.J. and Huber, J.D. (2000), "Putting parties in their place: Inferring party left- right ideological positions from party manifestos data," *American Journal of Political Science*, 44, 94-103.
Hatcher, Larry. (1994), *Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling*. Cary, NC: SAS Institute.
Hu, L. and Bentler, P.M. (1999), "Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives," *Structural Equation Modeling*, 6, 1-55.
Jöreskog, K.G. (1969), "A general approach to confirmatory factor analysis," *Psychometrika*, 34, 183-202.
Jöreskog, K.G. (1990), "New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares," *Quality and Quantity* 24, 387-404.
McDonald, R.P. (1978), "A simple comprehensive model for the analysis of covariance structures," *British Journal of Mathematical and Statistical Psychology*, 37, 234-251.
Muthén, B.O. (1984), "A general structural equation model with dichotomous, ordered, categorical, and continuous latent variable indicators," *Pychometrika*, 49, 115132.
Muthén, L.K. and Muthén, B.O. (2006), *Mplus*, Los Angeles: Muthén and Muthén.
Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, 6, 461-464.
Spearman, C. (1904). "General intelligence, objectively determined and measured," *American Journal of Psychology*, 15, 201-293.
Steiger, J. H. and Lind, J. (1980), "Statistically-based tests for the number of common factors," Paper presented at the Annual Spring Meeting of the Psychometric Society, Iowa City.
Jöreskog, K.G. and Sörbom, D. (2004), *LISREL 8.7*. Scientific Software International, Inc.
Arbuckle, J.L. (2007), *Amos16.0 User's Guide*, Chicago, IL: SPSS Inc.