Confirmatory Factor Analysis using Amos, LISREL, and Mplus

This document summarizes confirmatory factor analysis and illustrates how to estimate individual models using Amos 7.0, LISREL 8.8, and Mplus 5.1.

Jeremy J. Albright

1. Introduction

1.1. Factor Analysis and Latent Variables

Factor analysis is a common 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). For example, 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).

1.2. Exploratory versus Confirmatory Factor Analysis

It is possible to distinguish between two categories of factor analysis depending on whether the investigator wishes to explore patterns in the data or to test explicitly stated hypotheses. Exploratory factor analysis (EFA), corresponding to the former task, is available in general purpose statistical software such as SPSS, SAS, and Stata. When carrying out an EFA no substantive constraints are imposed on the data. Instead it is assumed that each common factor affects every observed variable and that the common factors are either all correlated or uncorrelated. Confirmatory factor analysis (CFA), on the other hand, is theory-driven. With CFA it is possible to place substantively meaningful constraints on the factor model, such as setting the effect of one latent variable to equal zero on a subset of the observed variables. The advantage of CFA is that it allows for testing hypotheses about a particular factor structure.

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 and as such is estimated using SEM software. 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.

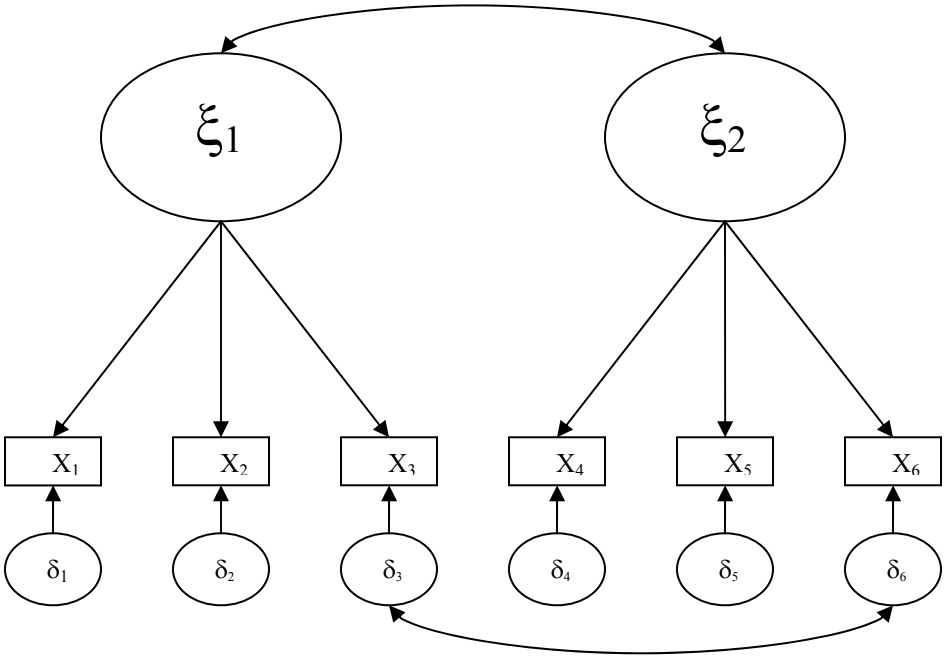
1.3. Model Specification and Identification

It is common to display confirmatory factor models as path diagrams in which squares represent observed variables and circles represent the latent concepts. Additionally single-headed arrows are used to imply a direction of assumed causal influence, and double-headed arrows are used to represent covariance between two latent variables. Figure 1 below provides an example.

In factor analysis the researcher almost always assumes that the latent variables "cause" the observed variables, as shown by the single-headed arrows pointing away from the circles and towards the manifest variables. The two ξ (ksi) latent variables represent *common factors*, with paths pointing to more than one observed variable. 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 not captured by the common factors, such as measurement error. In this model the two ξ_i are expected to covary, as

represented by the two-headed arrow. Additionally error in the measurement of x_3 is expected to correlate to some extent with measurement error for x_6. 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.

Figure 1: Path Diagram of a Confirmatory Factor Model



To facilitate formal presentation it is conventional to treat the observed and latent variables as deviations from their means. The confirmatory factor model can be summarized by the equation

$$\mathbf{x} = \Lambda\xi + \delta \tag{1}$$

in which x is the vector of observed variables, Λ (lambda) is the matrix of 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:

$$\begin{aligned} x_1 &= \lambda_{11}\xi_1 + \delta_1 & x_2 &= \lambda_{21}\xi_1 + \delta_2 & x_3 &= \lambda_{31}\xi_1 + \delta_3 \\ x_4 &= \lambda_{42}\xi_2 + \delta_4 & x_5 &= \lambda_{52}\xi_2 + \delta_5 & x_6 &= \lambda_{62}\xi_2 + \delta_6 \end{aligned} \tag{2}$$

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

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

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 Σ (sigma), can be decomposed as follows:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{3}$$

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, pg. 236). Estimation proceeds by finding the parameters $\hat{\boldsymbol{\Lambda}}$, $\hat{\boldsymbol{\Phi}}$, and $\hat{\boldsymbol{\Theta}}$ whose predicted $\hat{\boldsymbol{\Sigma}}$ is as close to the sample *x* covariance matrix 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, including Full Information Maximum Likelihood (FIML) for situations in which there are missing values in the raw data file. It will also describe a weighted least squares (WLS) approach suitable for situations in which the *x* variables are categorical.

One essential step in CFA is determining whether the specified model is identified. An unidentified model is one for which it is impossible to come up with unique parameter estimates. As a simple example, the equation 10 = 2x + 3y is not identified because an infinite number of values for *x* and *y* could make the equation true. 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. 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. When the sufficient rules are not met, however, identification can only be definitively proven by carrying out the algebraic manipulations needed to rewrite the unknown parameters in terms of the observed variances and covariances.

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.

1.5. Goodness of Fit

After estimating a CFA the next step is to assess how well the model matches the observed data. A large class of omnibus tests exists for determining overall model fit. It is conventional to report at least one of these as well as the individual regression weights (factor loadings) and some indication of their significance. The example used in this document reports a χ^2 statistic, available in all packages, for which the null hypothesis is

that the implied covariance matrix is equivalent to the observed covariance matrix. Failure to reject the null is therefore 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 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.

Because of these drawbacks many alternative fit statistics have been developed, though each has its own advantages and disadvantages. Another commonly reported statistic is the Root Mean Square Error of Approximation (RMSEA), a measure of fit introduced by Steiger and Lind (1980). The Amos 6 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, pg. 496). Additional statistics, such as the Akaike Information Criterion (Akaike, 1987) and Schwarz's Bayesian Information Criterion (Schwartz, 1978), can be used to compare to differently specified models.

Appendix C of the Amos 7.0 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. Hu and Bentler (1999) provide rules of thumb for deciding which statistics to report and choosing cut-off values for declaring significance.

This document provides step-by-step examples for conducting a CFA with three commonly used packages: Amos, LISREL, and Mplus. The next section provides an example of a one- and two-factor CFA with six observed indicators. Section 3 extends Section 2 to cover cases involving missing data. Section 4 discusses the commonly encountered situation in which the observed variables are categorical rather than continuous. Section 5 provides a brief summary.

2. Confirmatory Factor Analysis

2.1 Preliminaries

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?

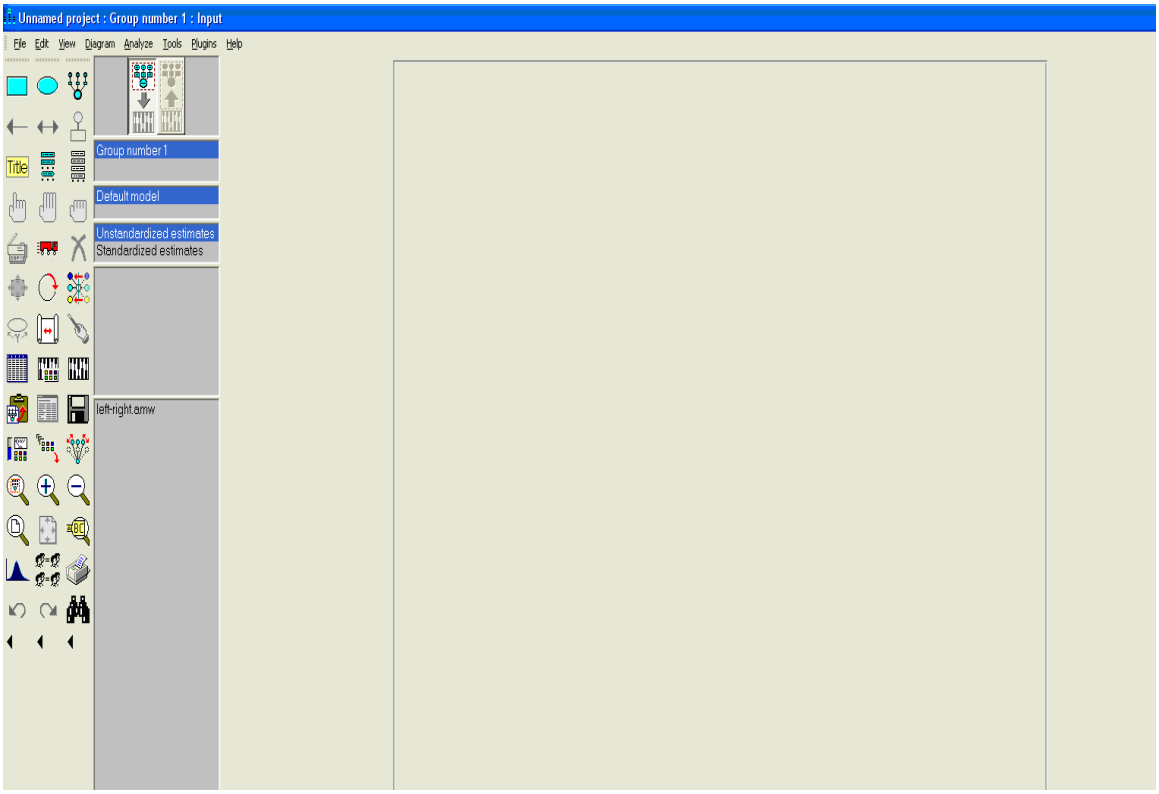
This example uses data from the American sample of the European Values Survey (European Values Group and World Values Survey Association, 2005) to determine whether a model with one or two common latent factors adequately describes attitudes on economic and moral issues. The survey queried a random sample of 1200 respondents about their economic, political, and moral values. Three questions summarizing economic attitudes and three questions summarizing moral attitudes, all measured on 10-point 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).

In order to provide an example of the most basic and common approach to confirmatory factor analysis, most of this document will demonstrate examples that rely on maximum likelihood estimation (MLE). However, MLE assumes multivariate normality among the observed variables, and preliminary univariate diagnostics 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. Section 4, however, does consider factor analysis with categorical indicators.

For this section missing data is handled by listwise deletion (all cases with missing observations on any indicator are removed). Section 3 will show an alternative approach for estimating a confirmatory factor model in the presence of missing data. 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.

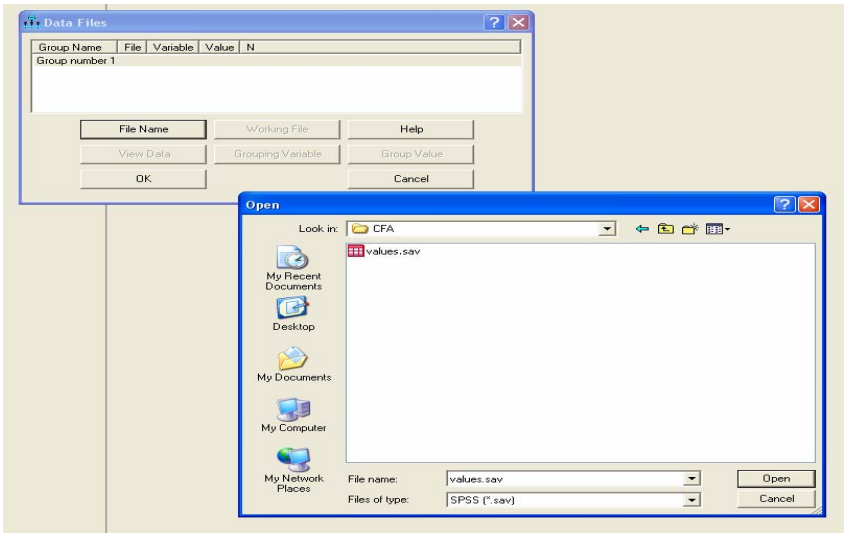
2.2 CFA using Amos

Amos can be launched from any computer running Windows in the UITS Student Technology Centers by going to **Start → Programs → Statistical Software → Amos 6 → Amos Graphics**. The following screen will display:



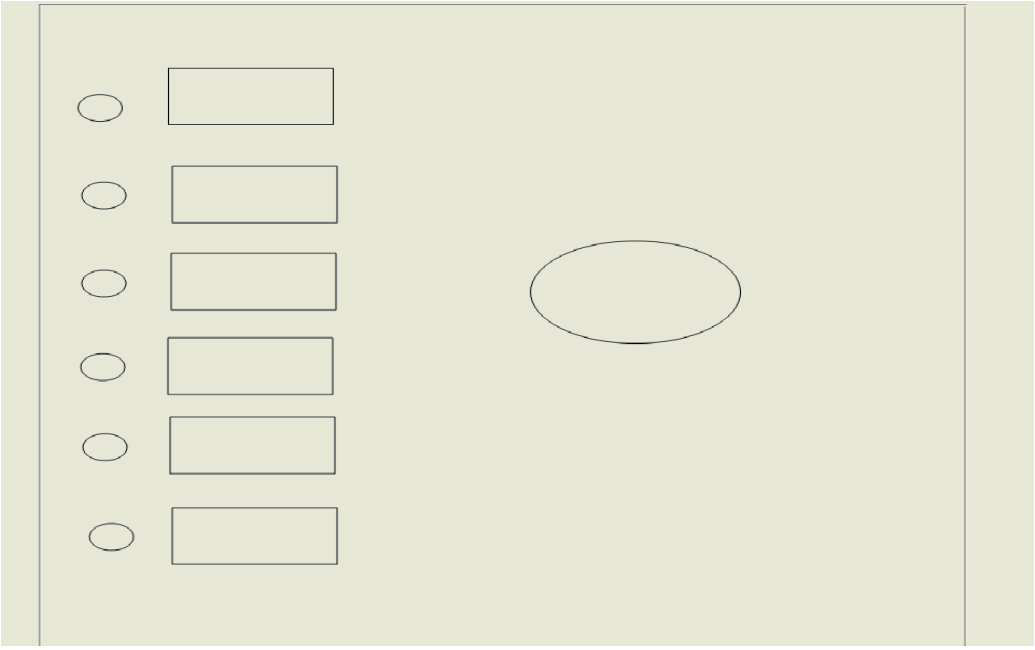
On the far left appear the different tools that can be used to create path diagrams. Just to the right of the toolbar buttons is a column that will display information about the model after estimates have been calculated. The remainder of the screen contains the area where the path diagram will be drawn.

To load the data go to **File → Data Files**. The **Data Files** dialog box then opens. Click on **File Name** and navigate to the location where the data file is stored. By default, Amos looks for an SPSS file.

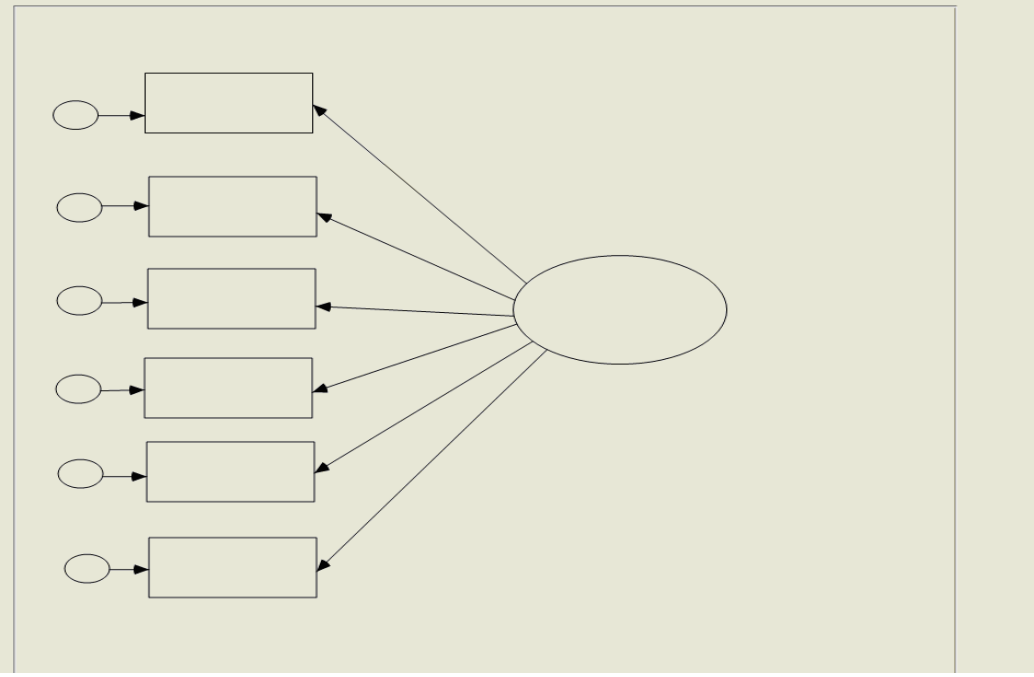


Choose the data file you wish to open, click **Open**, then **OK**.

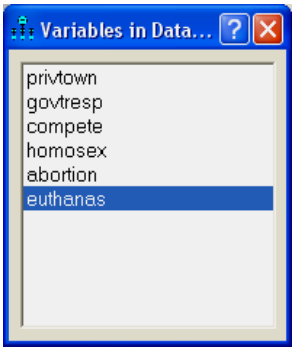
The next step is to draw the path diagram. A model where a single latent variable is assumed to underlie the six observed variables will be drawn first. To add the observed variables to the diagram, first click on the blue rectangle in the upper left corner of the tool bar  (alternatively click on **Diagram → Draw Observed**). Then in the empty drawing area hold down the left mouse button to draw a rectangle. Left-click five more times to create a total of six equally sized boxes. To add the latent variables, click on the blue oval in the tool box  (alternatively click on **Diagram → Draw Unobserved**). There will be a total of seven latent variables in the diagram: one representing the common factor and six additional variables representing measurement error specific to each of the observed indicators. After drawing one large oval on the right draw another smaller one on the left and single click five times. You can move the objects you have just drawn by clicking on the **Move Objects** button  located on the tool bar and dragging the pieces of the diagram to where you want them. The path diagram will look something like this:



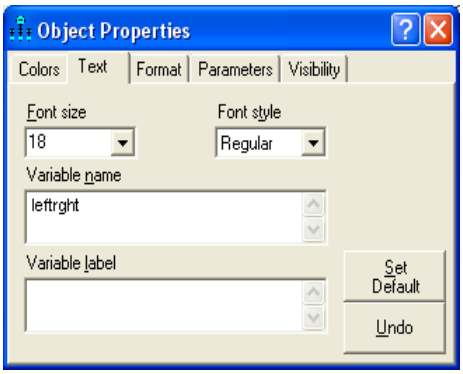
Next, click on the **Draw Paths** button  and click and drag from the common and unique factors to the appropriate observed variable. For SEMs latent variables are almost always assumed to "cause" the manifest variables, thus the arrows all point towards the rectangles.



The next step is to name each of the variables. The easiest way to name the observed variables is to bring up a list of variable names in the loaded data file. Go to **View →Variables in Dataset**. The **Variables in Dataset** window then opens.

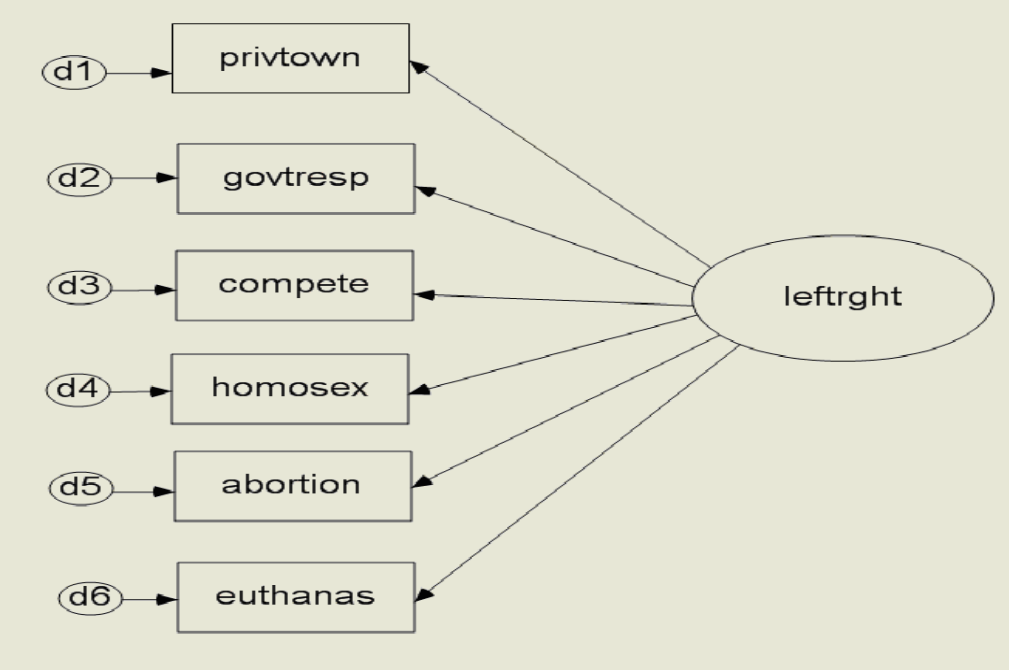


It is now possible to click-and-drag each variable to its corresponding rectangle in the path diagram.

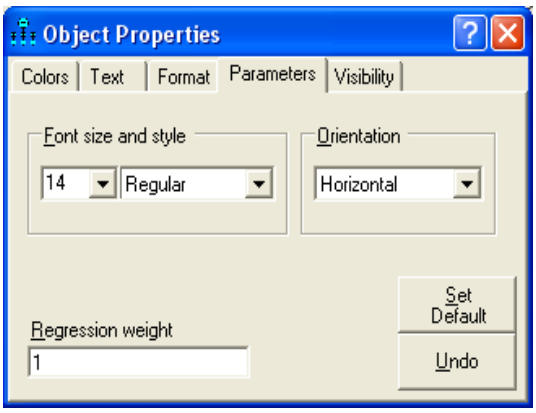


To name the common latent variable right click inside it and choose **Object Properties**. Then click on the Text tab and enter LEFTRGHT in the **Variable name** box. Amos applies the change immediately to the path diagram. It is also possible, if desired, to add a label describing the variable.

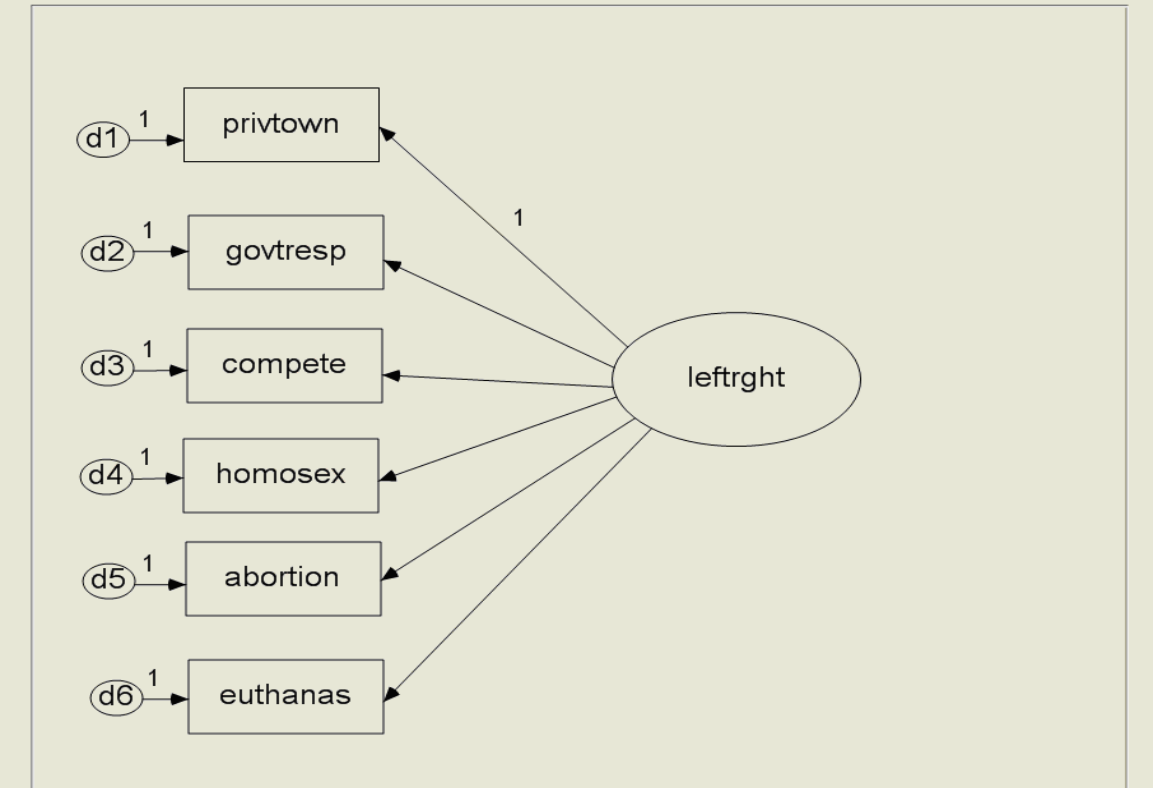
To name the unique factors representing measurement error follow the same process. Name these d_1 through d_6, yielding the following path diagram:



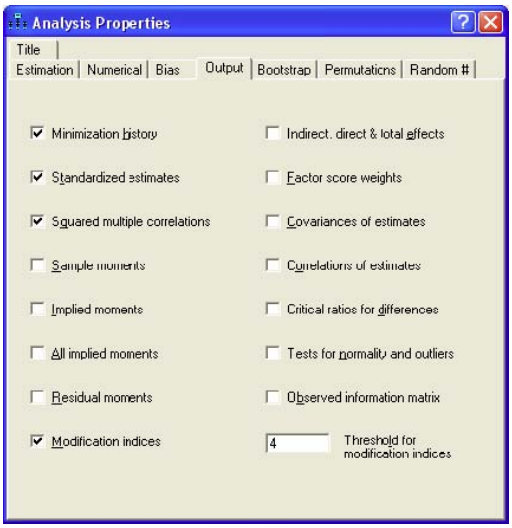
Without introducing some additional constraints the scales of the latent variables are meaningless and the model is not identified. A common procedure for setting the latent variable scale is to constrain a factor loading to equal one. Do this for the arrow pointing to PRIVTOWN by left-clicking directly on the arrow and choosing **Object Properties**. Then click on the **Paramaters** tab and enter 1 in the field labeled **Regression weight**.



Close the box by clicking on the **X** in the upper right hand corner. Follow the same steps for the arrows connecting each of the error terms with their respective indicators. When finished, the path diagram should look like this:

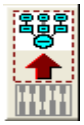


Note that there are several ways to draw a path diagram in Amos. A more efficient means of creating the same diagram may have been to draw the oval representing the common latent variable, clicking the **Draw a latent variable or add an indicator to a latent variable** button  , placing the curser inside the oval, and clicking six times. Amos adds six rectangles representing observed indicators along with ovals representing measurement error. The scales of the unique factors are automatically set by constraining the regression weights to equal one. The variables could then be named as described above.

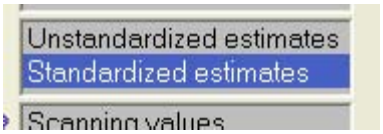


Before estimating the model it is possible to choose the information that will be provided in the output by going to **View → Analysis Properties**. Click on the **Output** tab and choose the following options: **Minimization history, Standardized estimates, Squared multiple correlations**, and **Modification indices**.

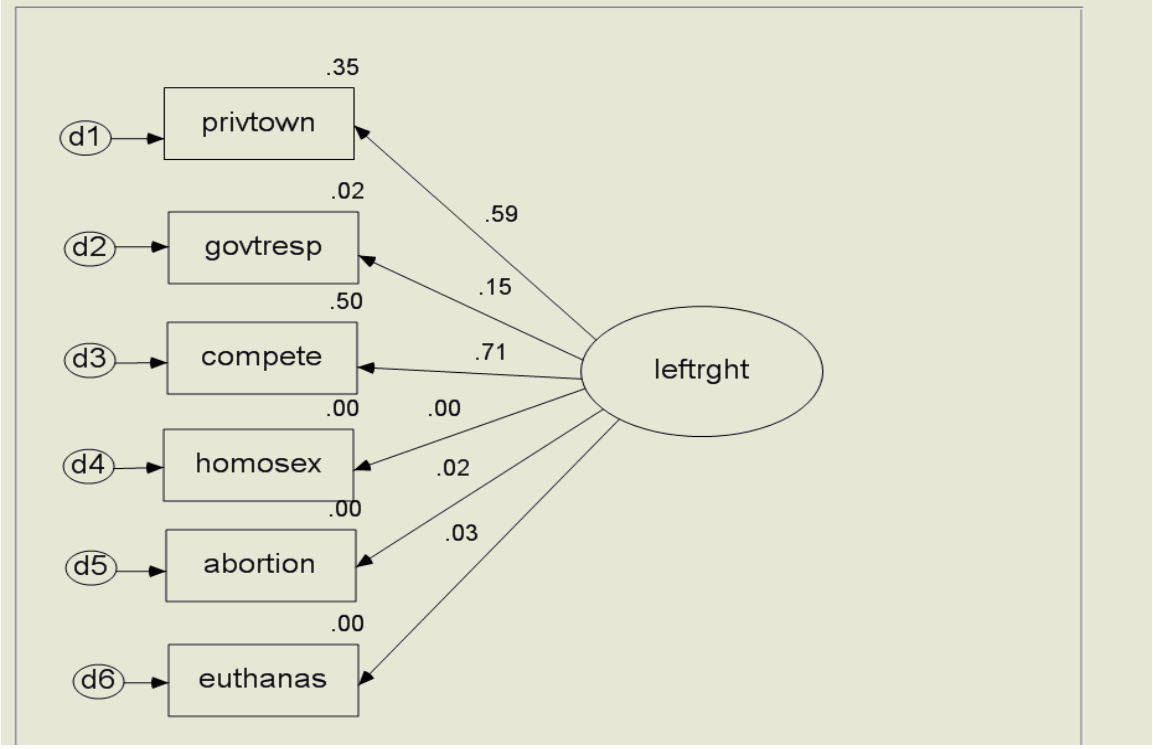
Close the box by clicking on the **X** in the upper right hand corner. To start the estimation choose **Analyze → Calculate Estimates**. After the estimation is completed it is possible to view the parameter estimates in the path diagram by clicking the **View the output path diagram** button.



By default the unstandardized estimates will display. To bring up the standardized estimates, click on the **Standardized estimates** option in the column between the tools and the drawing space.



The screen should look like this:



The path diagram now displays the standardized regression weights (factor loadings) for the common factor and each of the indicators. The squared multiple correlation coefficients (R^2), describing the amount of variance the common factor accounts for in the observed variables, are also displayed. Additionally, a χ^2 (chi-square) statistic is listed in the column between the tools and the path diagram.

It is evident that the three items related to economic values load on the common factor while the standardized regression weights for the three morality items are near zero. The PRIVTOWN and COMPETE variables appear to be the best indicators of LEFTRGHT. Their standardized regression weights are, respectively, .59 and .15. Additionally left-right explains about 35% of the variance in PRIVTOWN and 50% of the variance in COMPETE. GOVTRESP is the poorest among the economic indicators of LEFTRGHT, with an R^2 of .02 and a standardized regression weight of .15. Meanwhile the three moral values do not appear to have any relationship with the hypothesized LEFTRGHT factor. Each morality item has a corresponding R^2 of 0, meaning that LEFTRGHT explains practically no variance in these items. The χ^2 statistic of 798.6 (df=9) is very large. The null hypothesis that the model is a good fit to the data is easily rejected.

It is possible to get more information about the model than what appears in the path diagram by going to **View → Text** output. This opens an output window giving information about the raw data, the model, estimation, model fit, and any additional information requested with the **Analysis Properties** box utilized earlier. For now consider only the parameter estimates, which are displayed in the output window as follows:

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
privtown <--- leftrght	1.000				
govtresp <--- leftrght	.304	.085	3.591	***	
compete <--- leftrght	1.270	.411	3.087	.002	
homosex <--- leftrght	.003	.093	.035	.972	
abortion <--- leftrght	.039	.085	.458	.647	
euthanas <--- leftrght	.063	.088	.716	.474	

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

	Estimate
privtown <--- leftrght	.591
govtresp <--- leftrght	.150
compete <--- leftrght	.707
homosex <--- leftrght	.001
abortion <--- leftrght	.017
euthanas <--- leftrght	.027

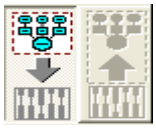
Variances: (Group number 1 - Default model)

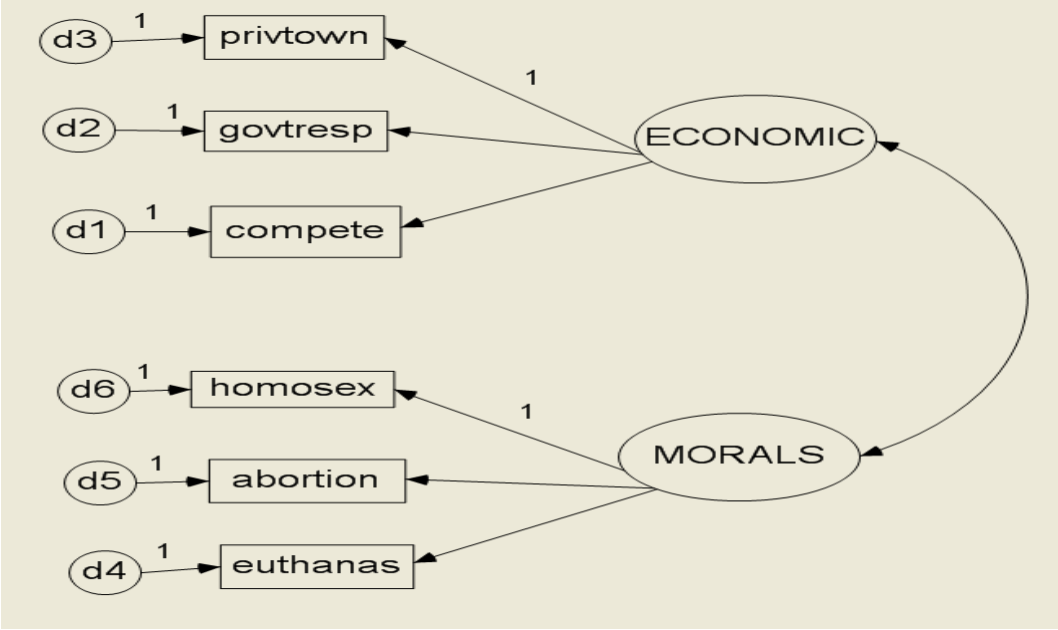
	Estimate	S.E.	C.R.	P	Label
leftrght	1.754	.588	2.985	.003	
d6	9.571	.398	24.061	***	
d5	8.897	.370	24.068	***	
d4	10.604	.441	24.073	***	
d3	2.839	.920	3.085	.002	
d2	7.024	.297	23.681	***	
d1	3.263	.582	5.606	***	

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

	Estimate
euthanas	.001
abortion	.000
homosex	.000
compete	.499
govtresp	.023
privtown	.350

In addition to the information available in the path diagram the output also displays standard errors, critical ratios (estimate/standard error), and p-values for the regression weights. No p-value is listed for the PRIVTOWN variable because it was constrained to one. Three stars (***) mean that the p-value is less than .001. GOVTRESP and COMPETE have p-values smaller than the conventional .05 significance level while the three p-values for the moral indicators are quite large. In other words, the regression weights associated with the moral values indicators are not significantly different from zero.

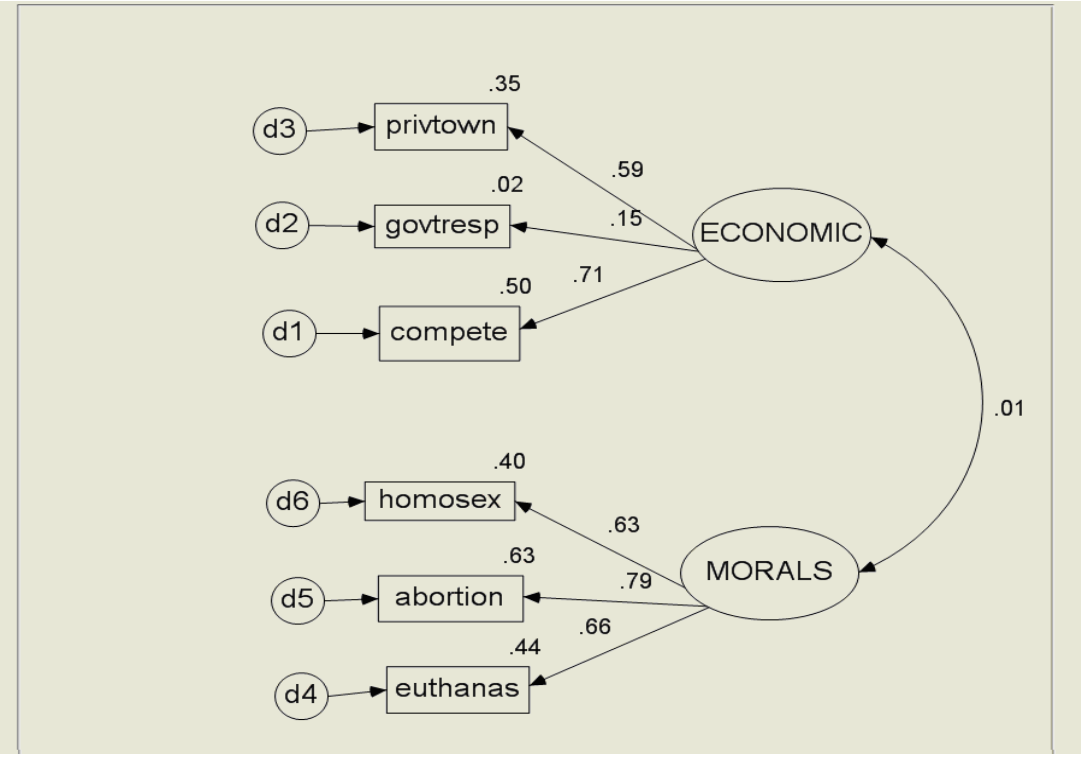
It is likely that a two factor model is more appropriate to describe the economic and moral values of Americans. To test this possibility, bring the path diagram back up without the estimates by clicking on **View the path diagram**: 



Click on the **Erase objects** button  and then on the three lines connecting the common factor to the moral values indicators. Next, add a new latent variable as before and name it MORALS. Change the name LEFTRGHT to ECONOMIC to keep the theoretical meaning of the latent variables clear. Draw three single-headed arrows to connect each of the observed moral values items to the MORALS factor and set the scale

of the new latent variable by constraining the regression weight for HOMOSEX to equal 1. Finally, allow ECONOMIC and MORALS to covary by connecting them with a two-headed arrow. The new diagram should look like the following:

To estimate this model, go to **Analyze → Calculate Estimates**. The standardized output can be viewed by clicking on the **View the output path diagram** button and then on **Standardized estimates**. This produces:



To get more detailed information about the results, go to **View → Text output**. Here is a sampling of the output:

Notes for Model (Default model)

Computation of degrees of freedom (Default model)

```
            Number of distinct sample moments:   21
Number of distinct parameters to be estimated:   13
                 Degrees of freedom (21 - 13):    8
```

Result (Default model)

```
Minimum was achieved
Chi-square = 35.299
Degrees of freedom = 8
Probability level = .000
```

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
privtown <--- economic	1.000				
compete <--- economic	1.268	.426	2.978	.003	
homosex <--- morals	1.000				
abortion <--- morals	1.148	.074	15.619	***	
euthanas <--- morals	1.000	.062	16.204	***	
govtresp <--- economic	.298	.085	3.518	***	

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

	Estimate
privtown <--- economic	.592
compete <--- economic	.706
homosex <--- morals	.632
abortion <--- morals	.792
euthanas <--- morals	.665
govtresp <--- economic	.148

Covariances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
economic <-→ morals	.024	.118	.206	.837	

Correlations: (Group number 1 - Default model)

	Estimate
economic <-→ morals	.009

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
economic	1.760	.610	2.887	.004	
morals	4.232	.423	10.009	***	
D6	5.343	.323	16.556	***	
D5	3.320	.338	9.820	***	
D4	6.373	.353	18.034	***	
D3	2.840	.954	2.978	.003	
D1	3.257	.604	5.391	***	
D2	7.030	.297	23.694	***	

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

	Estimate
govtresp	.022
homosex	.399
abortion	.627
euthanas	.442
compete	.499
privtown	.351

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	13	35.299	8	.000	4.412

Model	NPAR	CMIN	DF	P	CMIN/DF
Saturated model	21	.000	0		
Independence model	6	1037.631	15	.000	69.175

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.054	.037	.073	.319
Independence model	.243	.230	.255	.000

The χ^2 statistic for model fit is still significant, meaning that the null hypothesis of a good fit to the data can be rejected. The RMSEA likewise suggests that the fit of the model is questionable. The value of .054 exceeds the .05 suggested as a cut-off for accepting the model fit.

Under the `Regression Weights` heading the unstandardized loadings appear along with standard errors, a critical ratio, and p-values. All of the unconstrained estimates are significant. The `Standardized Regression Weights` can be interpreted as the correlation between the observed variable and the corresponding common factor. For this two-factor model the regression weights are all significant. In addition, the R^2 corresponding to five of the six observed variables indicate that the respective factor explains a respectable portion of the variance (between 35.1% and 62.7%). Only GOVTRESP has a negligible R^2 (.022), raising the possibility that this item does not tap the same values dimension as the other two economics questions.

While this model appears to fit the data substantially better than the single factor model, it may still be possible to improve the fit further. Previously when the **Analysis Properties** dialog box was opened we requested that Amos report modification indices. 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. Amos made the following suggestions:

Modification Indices (Group number 1 - Default model)

Covariances: (Group number 1 - Default model)

	M.I.	Par Change
d2 <-→ morals	26.732	.955
d4 <-→ d2	12.286	.755

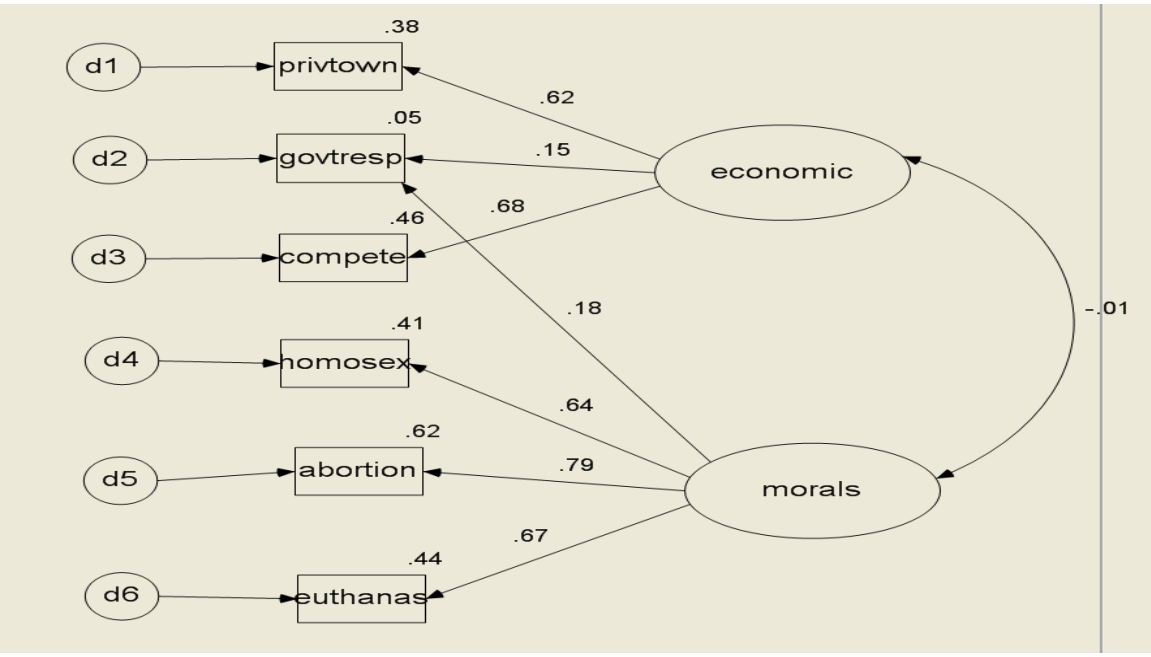
Regression Weights: (Group number 1 - Default model)

	M.I.	Par Change
govtresp <--- morals	26.728	.226
govtresp <--- homosex	28.448	.128
govtresp <--- abortion	16.866	.108
govtresp <--- euthanas	13.033	.091
homosex <--- govtresp	11.399	.102

The first box suggests adding either a covariance between the d_2 error term and MORALS or a covariance between d_4 and d_2. The first option violates the assumption

that the common and the unique factors are uncorrelated, and the second does not make much sense theoretically. The `Regression Weights` box, meanwhile, suggests adding single-headed arrows between GOVTRESP and any of several other variables. The first suggestion, to add a path from MORALS to GOVTRESP, seems most plausible. The GOVTRESP variable 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. According to the modification index, freeing this loading will reduce the χ^2 value by about 26.728.

A final model was thus estimated after adding a path from MORALS to GOVTRESP. The standardized output, as displayed in the path diagram, is the following:



Selected output is the following:

Result (Default model)

```
Minimum was achieved
Chi-square = 7.927
Degrees of freedom = 7
Probability level = .339
```

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
privtown <--- economic	1.000				
compete <--- economic	1.176	.382	3.078	.002	

	Estimate	S.E.	C.R.	P	Label
homosex <--- morals	1.000				
abortion <--- morals	1.127	.071	15.940	***	
euthanas <--- morals	.991	.061	16.304	***	
govtresp <--- economic	.291	.084	3.480	***	
govtresp <--- Morals	.227	.044	5.152	***	

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

	Estimate
privtown <--- economic	.615
compete <--- economic	.680
homosex <--- morals	.638
abortion <--- morals	.785
euthanas <--- morals	.666
govtresp <--- economic	.150
govtresp <--- Morals	.176

Covariances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
economic <-→ morals	-.030	.126	-.243	.808	

Correlations: (Group number 1 - Default model)

	Estimate
economic <-→ morals	-.011

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
economic	1.899	.636	2.986	.003	
morals	4.322	.425	10.163	***	
d6	5.333	.319	16.717	***	
d5	3.411	.327	10.430	***	
d4	6.282	.351	17.910	***	
d3	3.044	.859	3.544	***	
d1	3.118	.628	4.965	***	
d2	6.807	.291	23.433	***	

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

	Estimate
govtresp	.053
homosex	.408
abortion	.617
euthanas	.443
compete	.463
privtown	.378

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	14	7.927	7	.339	1.132
Saturated model	21	.000	0		
Independence model	6	1037.631	15	.000	69.175

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE

```
Model               RMSEA   LO 90   HI 90   PCLOSE
Default model         .011    .000    .039     .994
Independence model    .243    .230    .255     .000
```

The overall model fit appears quite good. The χ^2 test yields a value of 7.927 which, evaluated with 7 degrees of freedom, has a corresponding p-value of .339. This p-value is too high to reject the null of a good fit. Additionally the RMSEA is .011, well below the .05 cut-off. Both tests suggest that the model is a good fit to the data.

GOVTRESP has low standardized loadings on both factors (.150 for ECONOMIC and .176 for MORALS), suggesting that it is a rather unreliable indicator of both economic and moral values. However, the other items have moderate to strong standardized loadings. For PRIVTOWN the loading is .615; for COMPETE it is .680; for HOMOSEX it is .638; for ABORTION it is .785, and for EUTHANAS it is .666. The squared multiple correlations provide information on how much variance the common factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a low R^2 of only .053. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.378), HOMOSEX (.408), EUTHANAS (.443), COMPETE (.463), and ABORTION (.617). 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 ($p<.808$).

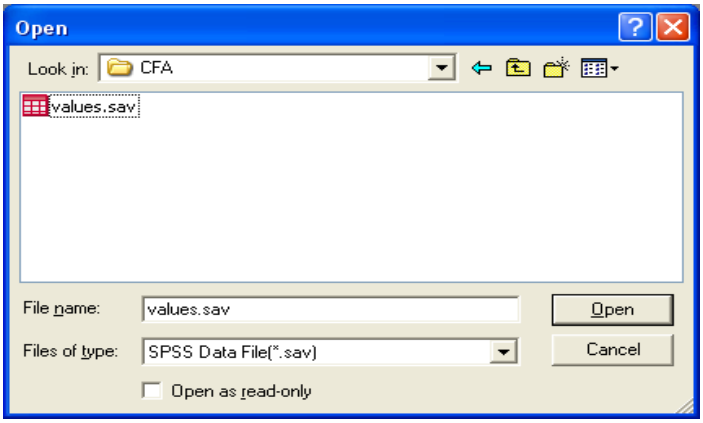
In summary, there appear to be two orthogonal dimensions which underlie American attitudes on a number of different issues: one representing economic values and the other representing moral values. It is unclear which dimension the GOVTRESP item was tapping, however, and future surveys should employ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

2.3 CFA USING LISREL

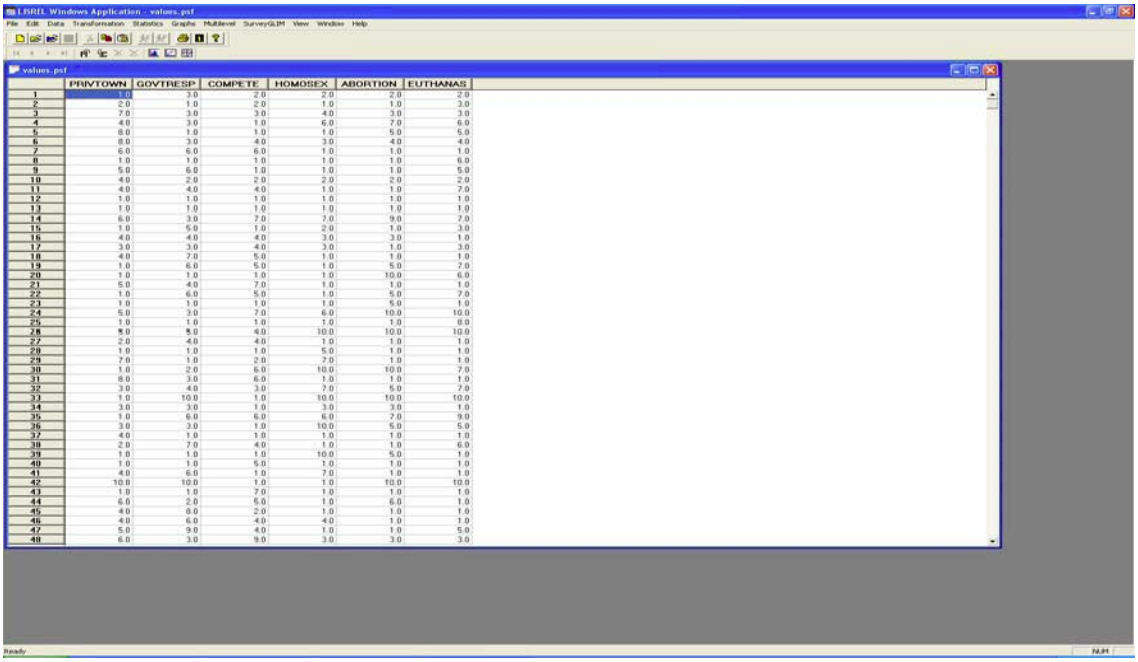
This section demonstrates how to estimate a confirmatory factor model using LISREL 8.7. Because the previous subsection revealed that the single common factor model was a poor fit to the data, this section will begin with the two factor model.

LISREL can be launched from any computer running Windows in the UITS Student Technology Centers by going to **Start → All Programs → Statistical Software →LISREL 8.80 → LISREL 8.80**.

The easiest way to get LISREL to analyze raw data is to import the data file 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 SPSS. To open the SPSS file *values.sav* saved in the C:\temp\CFA folder, go to **File → Import External Data in Other Formats**. The **Open** dialog box opens. Change **Files of Type** to *SPSS Data File(*.sav)*, navigate to the correct folder, and click on *values.sav*.



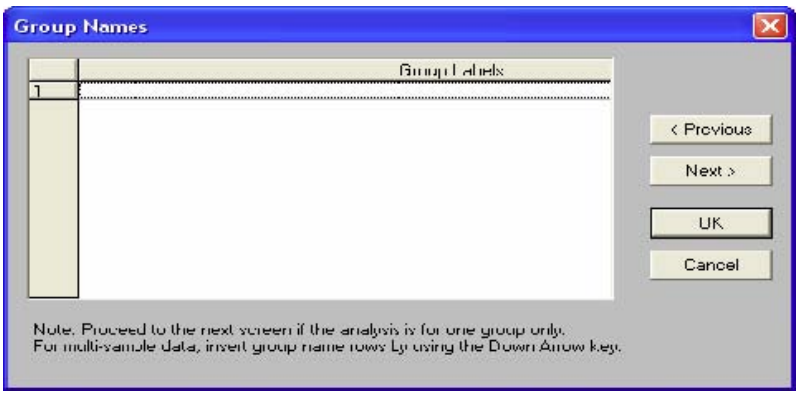
Click **Open.** You will then be prompted to save the data as a .psf file. Name the file *values* and click **Save** to store it in the working directory (currently C:/temp/CFA). A spreadsheet with the raw data will display.

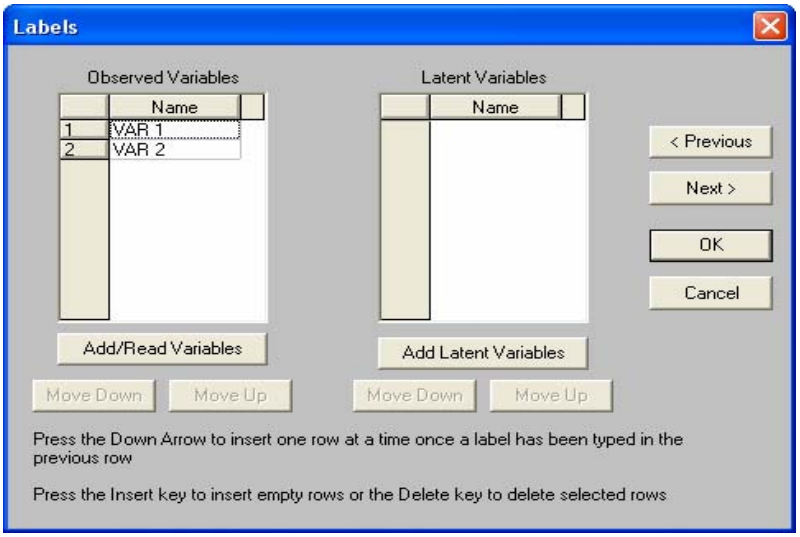


Like Amos, LISREL allows you to construct a path diagram of the model you wish to estimate. Go to **File → New** and choose **Path Diagram**. You will be immediately prompted to save the path diagram. Name the file *values* and click **Save** (the .pth extension will differentiate this file from the 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 *Two Factor Model* in the **Title** field and click **Next**.

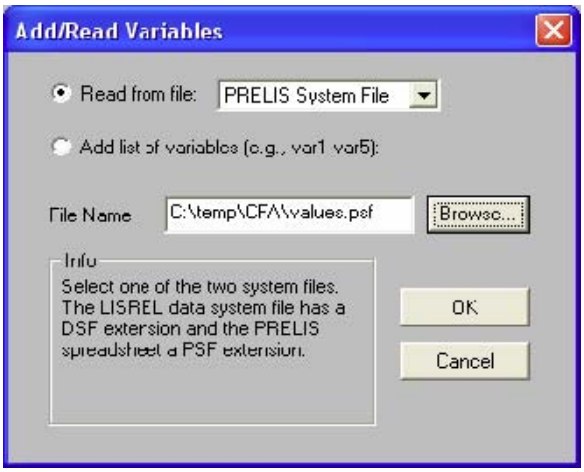
The **Group Names** box opens, which is used to label different groups when comparing models for multiple independent samples. Because we are interested only in the single sample of American respondents we can skip this box by clicking **Next**.

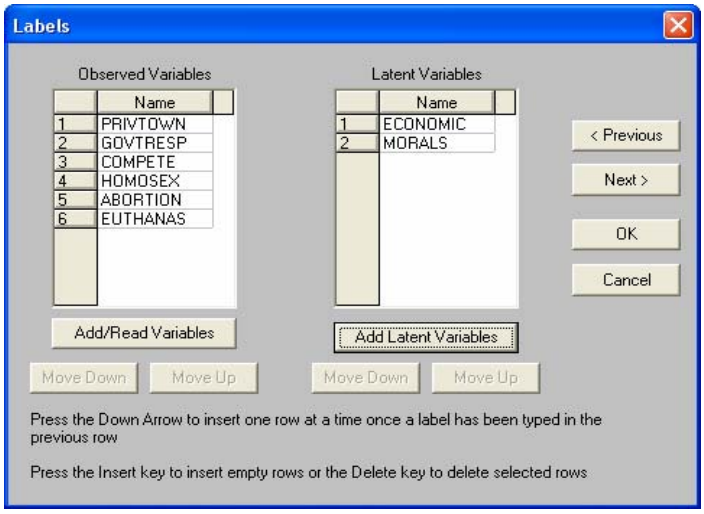




The **Labels** dialog box then opens, which is used to identify the latent and observed variables to be analyzed.

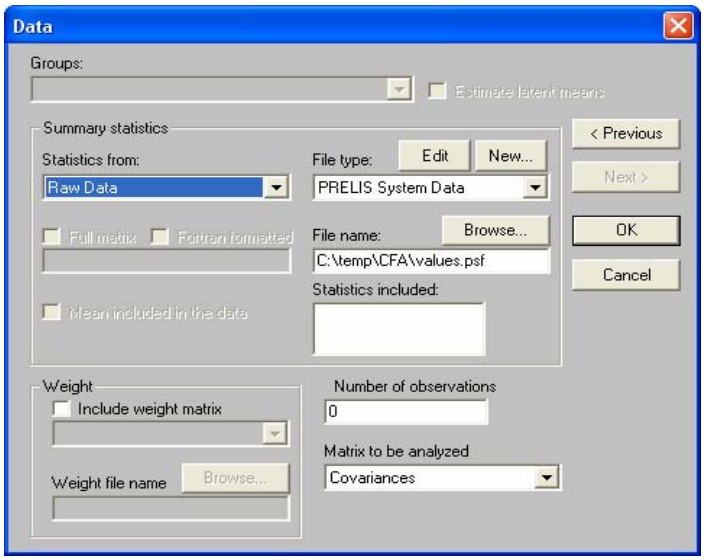
Currently no variables have been selected. To choose variable names click on **Add/Read Variables**. This opens a new dialog box used to locate the PRELIS system file. Verify that the **Read from file** radio button is chosen and pick **PRELIS System File** from the drop-down menu. Then click **Browse** to choose the PRELIS system file created earlier. Click **OK**.





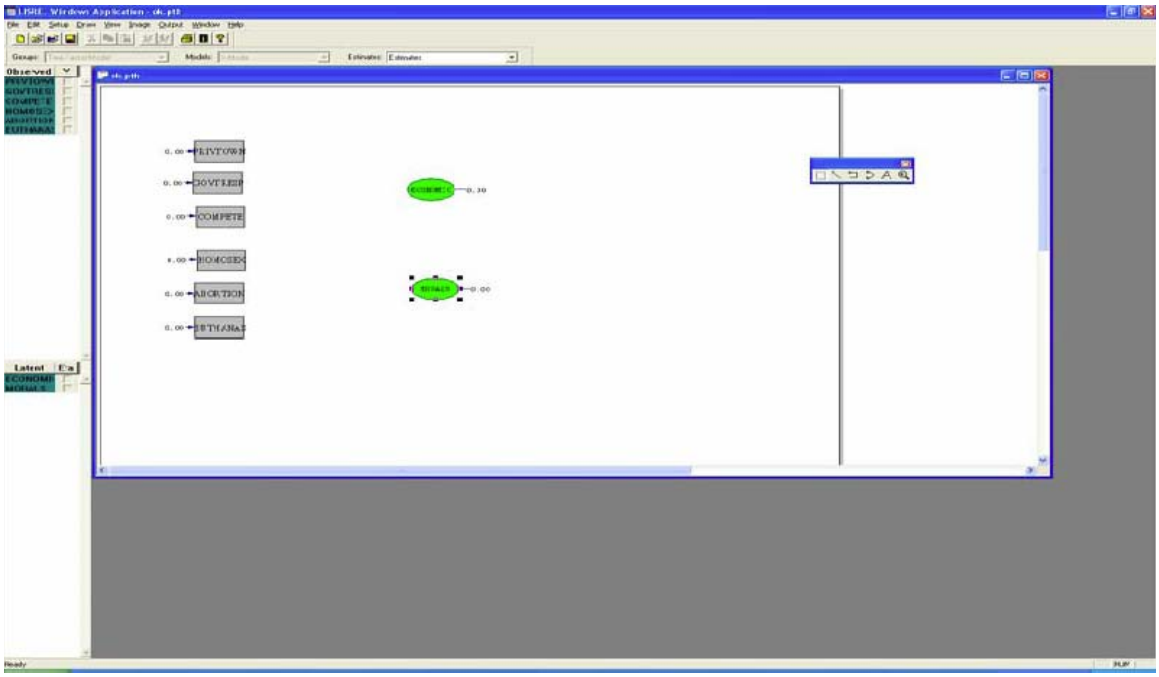
The names of the observed variables are now listed in the **Labels** box. To add the names of the latent variables click **Add Latent Variables**. Enter ECONOMIC in the box that opens. Repeat to enter the name of the second common factor MORALS. Click **OK**.

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

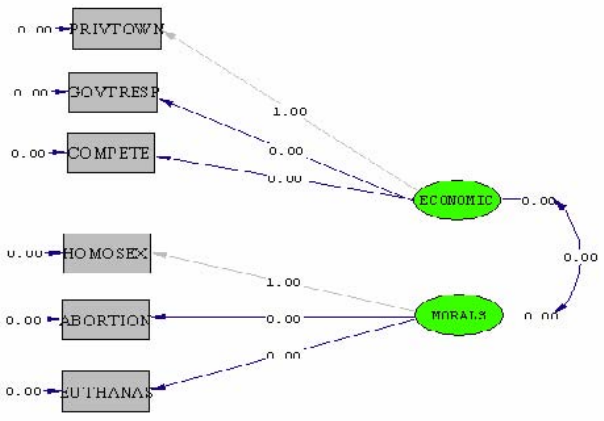


Raw data from a PRELIS system file will be analyzed. If desired, the data can be viewed and edited by clicking on the **Edit** button. Because this system file already contains information about the sample size it is not necessary to make further changes. Click **OK**.

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.

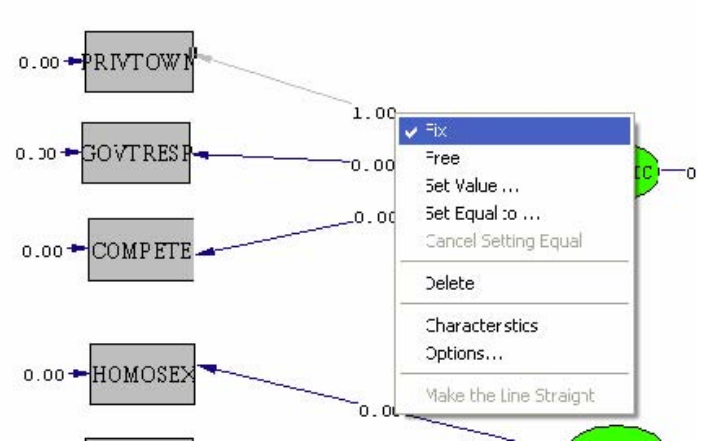


Next click on the single-headed arrow on the tool bar and connect the ECONOMIC factor to PRIVTOWN, GOVTRESP, and COMPETE. Also draw arrows from MORALS to HOMOSEX, ABORTION, and EUTHANAS. Because the usual assumption is that the latent variables "cause" the observed variables, the arrows should point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



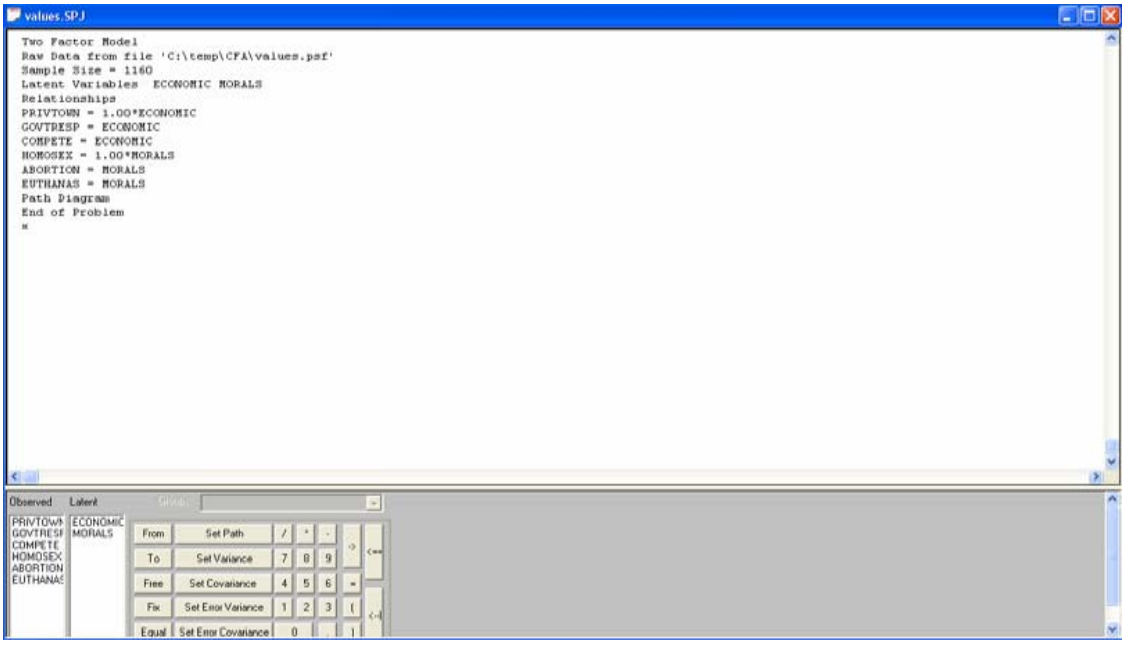
Unlike Amos, it is not necessary to draw the unique factors representing measurement error for each of the observed variables. LISREL includes these by default and

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**.

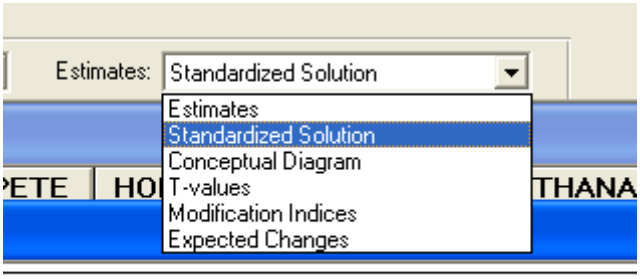


Do the same for the path connecting MORALS to HOMOSEX to set the metric for the second common factor.

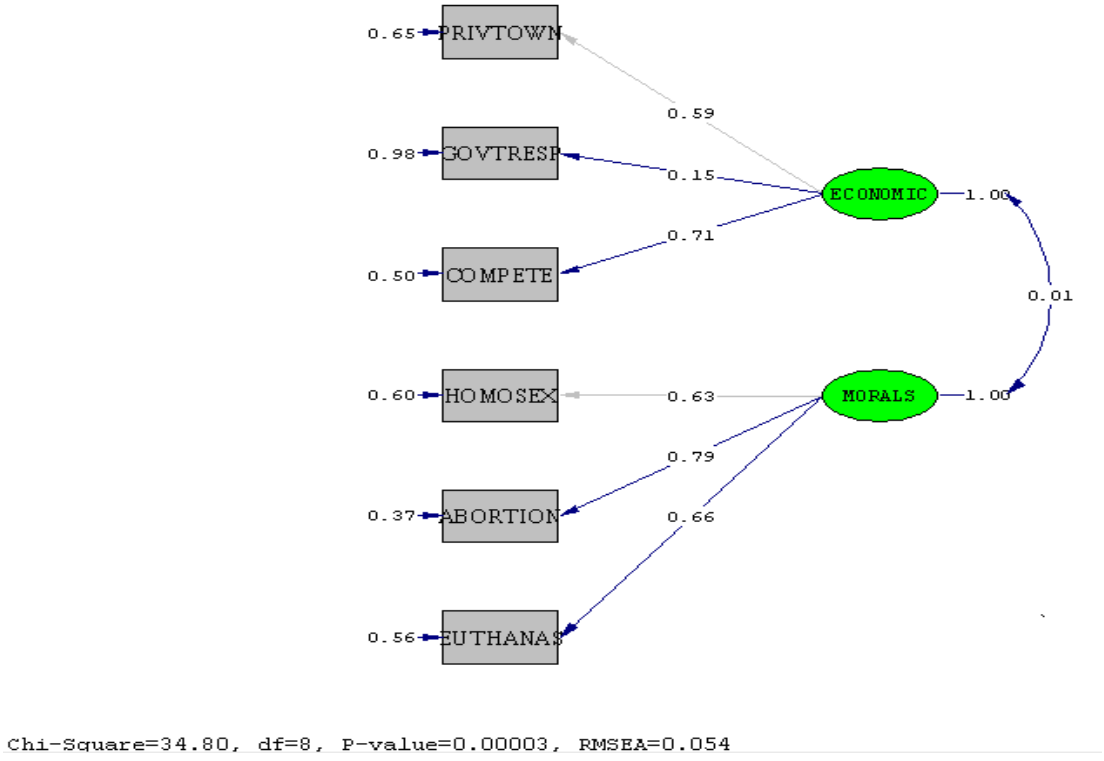
The final step before estimation is to build from the path diagram the corresponding syntax LISREL uses for estimation. There are actually two languages that LISREL understands: LISREL syntax and SIMPLIS syntax. As its name suggests, SIMPLIS is more straightforward and easy to read than LISREL syntax. 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.



To begin estimation click the **Run LISREL** button . The unstandardized estimates will then appear in the path diagram by default. To view the standardized estimates choose **Standardized Solution** from the **Estimates** drop-down menu:



The path diagram then will look like the following:



The unstandardized estimates, standardized estimates, t-values, and modification index information can all be obtained by choosing the appropriate option from the **Estimates** drop-down menu. Alternatively, each time the **Run LISREL** button is clicked a text output file is written to the working directory (extension .out) which contains additional information. 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:  6/ 7/2006
                            TIME: 14:46


                          L I S R E L  8.70

                                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., 1981-2004
        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:

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

Two Factor Model

       Covariance Matrix

          PRIVTOWN   GOVTRESP    COMPETE    HOMOSEX   ABORTION   EUTHANAS
          --------   --------   --------   --------   --------   --------
PRIVTOWN      5.02
GOVTRESP      0.53       7.19
 COMPETE      2.23       0.66       5.67
 HOMOSEX     -0.28       1.34      -0.04      10.61
ABORTION     -0.13       0.96       0.06       4.86       8.91
EUTHANAS      0.01       0.89       0.08       4.24       4.87       9.59



Two Factor Model

Number of Iterations =  6

LISREL Estimates (Maximum Likelihood)

       Measurement Equations


PRIVTOWN = 1.00*economic, Errorvar.= 3.26 , R² = 0.35
                                (0.60)
                                 5.39

GOVTRESP = 0.30*economic, Errorvar.= 7.04 , R² = 0.022
         (0.085)                 (0.30)
          3.52                    23.69
```

```
 COMPETE = 1.27*economic, Errorvar.= 2.84 , R² = 0.50
         (0.43)                (0.95)
          2.98                  2.98

 HOMOSEX = 1.00*morals, Errorvar.= 6.38 , R² = 0.40
                           (0.35)
                            18.03

ABORTION = 1.15*morals, Errorvar.= 3.32 , R² = 0.63
         (0.074)              (0.34)
          15.62                9.82

EUTHANAS = 1.00*morals, Errorvar.= 5.35 , R² = 0.44
         (0.062)              (0.32)
          16.20                16.56
```

Covariance Matrix of Independent Variables

	economic	morals
economic	1.76	
	(0.61)	
	2.89	
morals	0.02	4.24
	(0.12)	(0.42)
	0.21	10.01

Goodness of Fit Statistics

```
                    Degrees of Freedom = 8
          Minimum Fit Function Chi-Square = 35.30 (P = 0.00)
    Normal Theory Weighted Least Squares Chi-Square = 34.80 (P = 0.00)
           Estimated Non-centrality Parameter (NCP) = 26.80
       90 Percent Confidence Interval for NCP = (12.17 ; 48.96)

                Minimum Fit Function Value = 0.030
          Population Discrepancy Function Value (F0) = 0.023
        90 Percent Confidence Interval for F0 = (0.011 ; 0.042)
        Root Mean Square Error of Approximation (RMSEA) = 0.054
      90 Percent Confidence Interval for RMSEA = (0.036 ; 0.073)
        P-Value for Test of Close Fit (RMSEA < 0.05) = 0.34

           Expected Cross-Validation Index (ECVI) = 0.052
      90 Percent Confidence Interval for ECVI = (0.040 ; 0.072)
                 ECVI for Saturated Model = 0.036
               ECVI for Independence Model = 0.96

 Chi-Square for Independence Model with 15 Degrees of Freedom = 1099.69
                   Independence AIC = 1111.69
                      Model AIC = 60.80
                    Saturated AIC = 42.00
                  Independence CAIC = 1148.02
                     Model CAIC = 139.53
                   Saturated CAIC = 169.18

                 Normed Fit Index (NFI) = 0.97
              Non-Normed Fit Index (NNFI) = 0.95
           Parsimony Normed Fit Index (PNFI) = 0.52
              Comparative Fit Index (CFI) = 0.97
              Incremental Fit Index (IFI) = 0.97
               Relative Fit Index (RFI) = 0.94

                   Critical N (CN) = 660.70
```

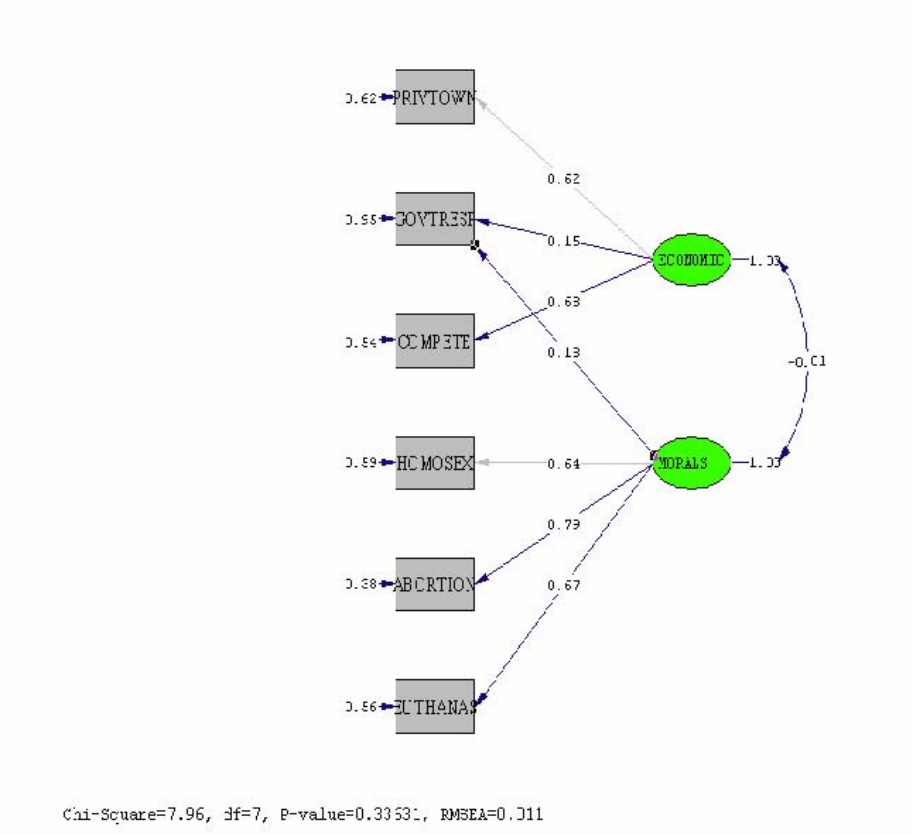
```
                Root Mean Square Residual (RMR) = 0.41
                      Standardized RMR = 0.049
                  Goodness of Fit Index (GFI) = 0.99
              Adjusted Goodness of Fit Index (AGFI) = 0.97
            Parsimony Goodness of Fit Index (PGFI) = 0.38

      The Modification Indices Suggest to Add the
 Path to  from     Decrease in Chi-Square    New Estimate
GOVTRESP  morals           26.9               0.23

The Modification Indices Suggest to Add an Error Covariance
 Between    and    Decrease in Chi-Square    New Estimate
HOMOSEX   GOVTRESP         12.3               0.76

                     Time used:    0.031 Seconds
```

The χ^2 statistic for model fit is 35.3 (df=8), which is large enough to reject the null that the model is a good fit to the data (the path diagram displays the Normal Theory Weighted Least Squares χ^2; to be consistent with the output from Amos this example considers the Minimum Fit Function χ^2 printed in the output). In addition the Root Mean Square Error of Approximation is .054. Using a cut-off rule of .05, the RMSEA is too high to indicate a good fit.

The standardized loadings represent the correlation between each observed variable and the corresponding factor. Considering first the indicators of ECONOMIC, they are .59 for PRIVOWN, .15 for GOVTRESP, and .71 for COMPETE. Considering the indicators of MORALS, the standardized loadings are .63 for HOMOSEX, .79 for ABORTION, and .66 for EUTHANAS. 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.

Additionally a good deal of the variance in each observed variable, with the exception of GOVTRESP, is accounted for. The R^2 for PRIVTOWN is .35; for COMPETE it is .50; for HOMOSEX it is .40; for ABORTION it is .63; and for EUTHANAS it is .44. 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 numbers offer suggestions for improving the overall model fit. Two suggestions are given in the output: add an error covariance between HOMOSEX and GOVTRESP or add a path from GOVTRESP to MORALS. Both of these suggest that 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. One final model is therefore estimated adding a path from MORALS to GOVTRESP. This modification results in the following standardized solution:



The (abbreviated) output is the following:

```
Two Factor Model

        Covariance Matrix

           PRIVTOWN   GOVTRESP    COMPETE    HOMOSEX   ABORTION   EUTHANAS
           --------   --------   --------   --------   --------   --------
 PRIVTOWN      5.02
 GOVTRESP      0.53       7.19
  COMPETE      2.23       0.66       5.67
  HOMOSEX     -0.28       1.34      -0.04      10.61
 ABORTION     -0.13       0.96       0.06       4.86       8.91
 EUTHANAS      0.01       0.89       0.08       4.24       4.87       9.59


 Two Factor Model

 Number of Iterations =  6

 LISREL Estimates (Maximum Likelihood)

        Measurement Equations


 PRIVTOWN = 1.00*economic, Errorvar.= 3.12 , R² = 0.38
                              (0.63)
                               4.97

 GOVTRESP = 0.29*economic + 0.23*morals, Errorvar.= 6.81 , R² = 0.053
          (0.084)         (0.044)              (0.29)
           3.48            5.15                23.43
```

```
 COMPETE = 1.18*economic, Errorvar.= 3.05 , R² = 0.46
         (0.38)                (0.86)
          3.08                  3.54

 HOMOSEX = 1.00*morals, Errorvar.= 6.29 , R² = 0.41
                              (0.35)
                               17.91

ABORTION = 1.13*morals, Errorvar.= 3.41 , R² = 0.62
         (0.071)              (0.33)
          15.94                10.43

EUTHANAS = 0.99*morals, Errorvar.= 5.34 , R² = 0.44
         (0.061)              (0.32)
          16.30                16.72
```

Covariance Matrix of Independent Variables

	economic	morals
economic	1.90	
	(0.64)	
	2.99	
morals	-0.03	4.33
	(0.13)	(0.43)
	-0.24	10.16

Goodness of Fit Statistics

Degrees of Freedom = 7
Minimum Fit Function Chi-Square = 7.93 (P = 0.34)
Normal Theory Weighted Least Squares Chi-Square = 7.96 (P = 0.34)
Chi-Square Difference with 1 Degree of Freedom = 26.84 (P = 0.0)
Estimated Non-centrality Parameter (NCP) = 0.96
90 Percent Confidence Interval for NCP = (0.0 ; 12.23)

Minimum Fit Function Value = 0.0068
Population Discrepancy Function Value (F0) = 0.00083
90 Percent Confidence Interval for F0 = (0.0 ; 0.011)
Root Mean Square Error of Approximation (RMSEA) = 0.011
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.039)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.99

Expected Cross-Validation Index (ECVI) = 0.031
90 Percent Confidence Interval for ECVI = (0.030 ; 0.041)
ECVI for Saturated Model = 0.036
ECVI for Independence Model = 0.96

Chi-Square for Independence Model with 15 Degrees of Freedom = 1099.69
Independence AIC = 1111.69
Model AIC = 35.96
Saturated AIC = 42.00
Independence CAIC = 1148.02
Model CAIC = 120.74
Saturated CAIC = 169.18

Normed Fit Index (NFI) = 0.99
Non-Normed Fit Index (NNFI) = 1.00
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) = 2702.18

```
          Root Mean Square Residual (RMR) = 0.11
                Standardized RMR = 0.014
              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:    0.016 Seconds
```

This model fits the data well. The χ^2 measure of model fit is 7.93 (df=7), which is too small to reject the null of a good fit (p=.34). Additionally the RMSEA has declined to .011, which is small enough (below .05) to indicate a good fit.

The unconstrained loadings are all statistically significant at the .05 level, having estimates that are more than twice the size of their standard errors. GOVTRESP continues to have a low correlation with the ECONOMIC factor (.15) and has a similarly low correlation with MORALS (.18). However, the remaining standardized loadings range from .62 (PRIVTOWN) to .79 (ABORTION). In between are HOMOSEX (.64), EUTHANAS (.67), and COMPETE (.68).

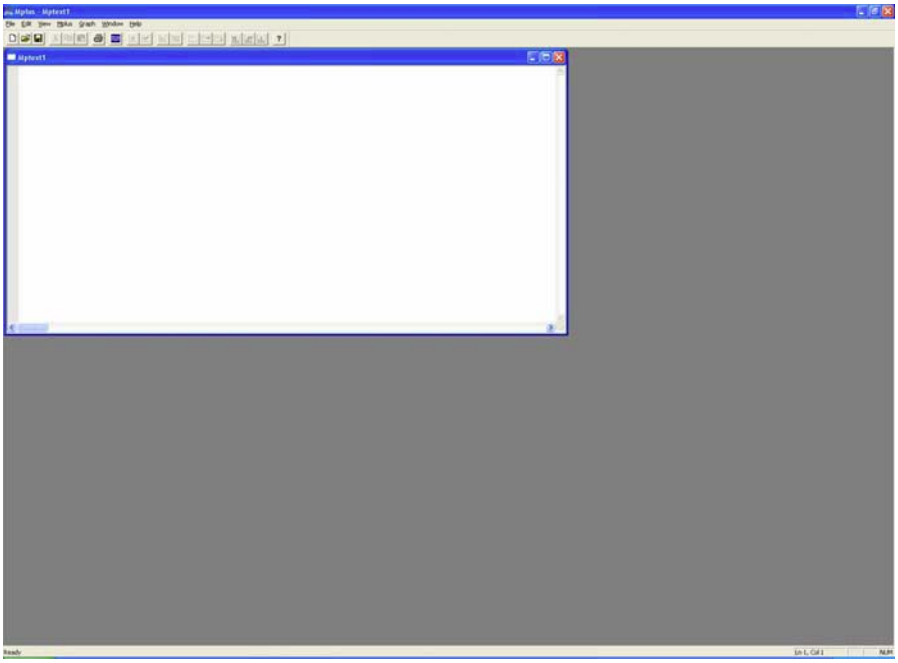
Despite receiving a path from both common factors GOVTRESP continues to have by far the smallest R^2 (.053). The remaining statistics are moderately well accounted for by the corresponding factors. The R^2 values are, in order of increasing magnitude, .38 for PRIVTOWN, .41 for HOMOSEX, .44 for EUTHANAS, .46 for COMPETE, and .62 for ABORTION. Finally, the correlation between ECONOMIC and MORALS is a negligible -.01 and, for this sample, is statistically indistinguishable from zero (covariance = -.03, std. error = .13).

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.

2.4 CFA using Mplus

This section demonstrates how to estimate a confirmatory factor model using Mplus 5.1. 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 → Statistical Software → Mplus → 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). Additionally Mplus cannot directly read a file saved in SPSS format. Instead the data must be saved as a free or fixed format ASCII file. 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.

To convert the SPSS file *values.sav* into tab delimited format open SPSS and load the file. Open the syntax editor in SPSS by going to **File → New → Syntax**. In the editor, type the following:

```
SAVE TRANSLATE OUTFILE='c:\temp\CFA\values.dat' /TYPE=TAB /MAP /REPLACE.
```

This will create a new ASCII file in the C:\temp\CFA folder that contains only values of observations without variable names. The first six rows of the file *values.dat* look like the following:

1	3	2	2	2	2
2	1	2	1	1	3
7	3	3	4	3	3
4	3	1	6	7	6
8	1	1	1	5	5
8	3	4	3	4	4

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:      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;
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 pathnames 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 above). 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 4.0
MUTHEN & MUTHEN
06/08/2006  12:21 PM

INPUT INSTRUCTIONS

  TITLE:     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;
  OUTPUT:    standardized;
           modindices;

INPUT READING TERMINATED NORMALLY

Two Factor Model;

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                                EXPECTED
Maximum number of iterations                          1000
Convergence criterion                            0.500D-04
Maximum number of steepest descent iterations           20

Input data file(s)
  values.dat

Input data format  FREE

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

        Value                             35.329
        Degrees of Freedom                    8
        P-Value                           0.0000

Chi-Square Test of Model Fit for the Baseline Model

        Value                           1038.526
        Degrees of Freedom                   15
        P-Value                           0.0000

CFI/TLI

        CFI                               0.973
        TLI                               0.950

Loglikelihood

        H0 Value                     -16407.791
        H1 Value                     -16390.126

Information Criteria

        Number of Free Parameters           13
        Akaike (AIC)                 32841.581
        Bayesian (BIC)               32907.312
        Sample-Size Adjusted BIC      32866.019
          (n* = (n + 2) / 24)
```

RMSEA (Root Mean Square Error Of Approximation)

Estimate	0.054	
90 Percent C.I.	0.037	0.073
Probability RMSEA <= .05	0.319	

SRMR (Standardized Root Mean Square Residual)

Value	0.049

MODEL RESULTS

	Estimates	S.E.	Est./S.E.	Std	StdYX
ECONOMIC BY					
PRIVTOWN	1.000	0.000	0.000	1.327	0.592
GOVTRESP	0.298	0.085	3.520	0.396	0.148
COMPETE	1.268	0.425	2.980	1.682	0.706
MORALS BY					
HOMOSEX	1.000	0.000	0.000	2.057	0.632
ABORTION	1.148	0.073	15.625	2.362	0.792
EUTHANAS	1.000	0.062	16.211	2.058	0.665
MORALS WITH					
ECONOMIC	0.024	0.118	0.206	0.009	0.009
Variances					
ECONOMIC	1.760	0.609	2.888	1.000	1.000
MORALS	4.232	0.423	10.013	1.000	1.000
Residual Variances					
PRIVTOWN	3.257	0.604	5.393	3.257	0.649
GOVTRESP	7.030	0.297	23.704	7.030	0.978
COMPETE	2.840	0.953	2.979	2.840	0.501
HOMOSEX	6.373	0.353	18.042	6.373	0.601
ABORTION	3.320	0.338	9.823	3.320	0.373
EUTHANAS	5.343	0.323	16.564	5.343	0.558

R-SQUARE

Observed Variable	R-Square
PRIVTOWN	0.351
GOVTRESP	0.022
COMPETE	0.499
HOMOSEX	0.399
ABORTION	0.627
EUTHANAS	0.442

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	26.907	0.227	0.467	0.174
WITH Statements				
HOMOSEX WITH GOVTRESP	12.301	0.756	0.756	0.087

Beginning Time: 12:21:04

```
       Ending Time:  12:21:04
      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) 1998-2006 Muthen & Muthen
```

The overall model fit is not great, with a χ^2 statistic of 35.329 (df=8) large enough to reject the null of a good fit. In addition the RMSEA is .054, which is higher than the cut-off value of .05 chosen to indicate a good fit.

Under the `Model Results` heading the unstandardized loadings appear along with standard errors, the ratio of the estimates to their standard errors, and two standardized estimates. The `Est./S.E.` column can be used to evaluate significance. 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. In this case all of the unconstrained loading estimates are significant. The column `StdYX` can be interpreted as the correlation between the latent and observed variables and is equivalent to the standardized estimates provided by Amos and LISREL. The standardized loadings are of moderate to strong magnitude. The smallest standardized estimate corresponds to the GOVTRESP variable and is .148. The remaining estimates for the standardized regression weights range from .592 (PRIVTOWN) to .792 (ABORTION). Finally, the output reports R^2 statistics describing the amount of variance accounted for by the respective factor. When there is only a single path from a factor to an observed variable these are equivalent to the standardized loadings squared. GOVTRESP has the lowest R^2 with .022. The remaining variables have R^2 statistics that range from .351 (PRIVTOWN) to .627 (ABORTION).

While the loadings are statistically significant, there remain some problems with the model. The overall test of fit still suggests that the model does not adequately fit the data. Also, the low R^2 for GOVTRESP suggests that the variable may not really be tapping the same economic values dimension as PRIVTOWN and COMPETE. Perhaps an alternative model would be more appropriate.

For hints at what such a model may be examine the modification indices reported in the output. Mplus makes two suggestions: 1) add a covariance between the HOMOSEX and GOVTRESP variables, and 2) add a path from the MORALS latent variable to the GOVTRESP variable. Both of these suggest that 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.

Enter the following syntax in the editor and 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;
```

The output is the following:

```
Mplus VERSION 4.0
MUTHEN & MUTHEN
06/20/2006   9:46 AM

INPUT INSTRUCTIONS

  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;

INPUT READING TERMINATED NORMALLY

Second Two Factor Model;

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                                EXPECTED
Maximum number of iterations                          1000
Convergence criterion                            0.500D-04
Maximum number of steepest descent iterations           20

Input data file(s)
  values.dat

Input data format  FREE

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT
```

```
Chi-Square Test of Model Fit

        Value                          7.934
        Degrees of Freedom                 7
        P-Value                       0.3383

Chi-Square Test of Model Fit for the Baseline Model

        Value                       1038.526
        Degrees of Freedom                15
        P-Value                       0.0000

CFI/TLI

        CFI                           0.999
        TLI                           0.998

Loglikelihood

        H0 Value                  -16394.093
        H1 Value                  -16390.126

Information Criteria

        Number of Free Parameters          14
        Akaike (AIC)                32816.187
        Bayesian (BIC)              32886.973
        Sample-Size Adjusted BIC    32842.505
          (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                       0.011
        90 Percent C.I.                0.000  0.039
        Probability RMSEA <= .05       0.994

SRMR (Standardized Root Mean Square Residual)

        Value                          0.014


MODEL RESULTS

                Estimates    S.E.  Est./S.E.    Std    StdYX

 ECONOMIC BY
    PRIVTOWN        1.000   0.000      0.000   1.376   0.614
    GOVTRESP        0.292   0.084      3.485   0.402   0.150
    COMPETE         1.179   0.383      3.078   1.622   0.681

 MORALS   BY
    HOMOSEX         1.000   0.000      0.000   2.079   0.638
    ABORTION        1.127   0.071     15.946   2.343   0.785
    EUTHANAS        0.991   0.061     16.311   2.060   0.666
    GOVTRESP        0.227   0.044      5.155   0.471   0.176

 MORALS   WITH
    ECONOMIC       -0.030   0.125     -0.241  -0.011  -0.011

 Variances
    ECONOMIC        1.894   0.634      2.986   1.000   1.000
    MORALS          4.322   0.425     10.166   1.000   1.000

 Residual Variances
    PRIVTOWN        3.123   0.627      4.985   3.123   0.623
    GOVTRESP        6.807   0.290     23.443   6.807   0.947
    COMPETE         3.037   0.861      3.528   3.037   0.536
    HOMOSEX         6.283   0.351     17.919   6.283   0.592
    ABORTION        3.410   0.327     10.432   3.410   0.383
    EUTHANAS        5.333   0.319     16.725   5.333   0.557
```

```
R-SQUARE

    Observed
    Variable  R-Square

    PRIVTOWN    0.377
    GOVTRESP    0.053
    COMPETE     0.464
    HOMOSEX     0.408
    ABORTION    0.617
    EUTHANAS    0.443


     Beginning Time:  09:46:44
        Ending Time:  09:46:45
       Elapsed Time:  00:00:01



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) 1998-2006 Muthen & Muthen
```

The overall model fit appears quite good. The χ^2 test yields a value of 7.934 (df=7), which has a corresponding p-value of .3383. This p-value is too high to reject the null of a good fit. The RMSEA is .011, sufficiently low to indicate acceptable fit.

The GOVTRESP variable has relatively low standardized loadings on each common factor (.150 for ECONOMIC and .176 for MORALS), suggesting that it is a weak indicator of both economic and moral values. However, the other indicators have moderate to strong standardized loadings. For PRIVTOWN the loading is .614, for COMPETE it is .681, for HOMOSEX it is .638, for ABORTION it is .785, and for EUTHANAS it is .666. The squared multiple correlations provide information on how much variance the factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a low R^2 of only .053. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.377), HOMOSEX (.408), EUTHANAS (.443), COMPETE (.464), and ABORTION (.617). Finally, the correlation between the two common factors is a very small -.011, and the covariance estimate of -.030 is not statistically distinguishable from zero.

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.

2.5 Summary

Tables 1 and 2 compare the unstandardized and standardized results from each package for both two-factor models. Standard errors for the unstandardized estimates are

displayed in parentheses. Additionally the χ^2 and Root Mean Square Error of Approximation (RMSEA) statistics are displayed. Results vary little across packages.

Table 1: Comparison of Estimates: Two Factor Model – A

	Amos		***LISREL***		Mplus	
	Unstand.	Stand	Unstand	Stand.	Unstand.	Stand.
ECONOMIC → PRIVTOWN	1.00	.592	1.00	.59	1.00	.592
ECONOMIC → GOVTRESP	.298 (.085)	.148	.30 (.085)	.15	.298 (.085)	.148
ECONOMIC → COMPETE	1.268 (.426)	.706	1.27 (.43)	.71	1.268 (.425)	.706
MORALS → HOMOSEX	1.00	.632	1.00	.63	1.00	.632
MORALS → ABORTION	1.148 (.074)	.792	1.15 (.074)	.79	1.148 (.073)	.792
MORALS → EUTHANAS	1.000 (.062)	.665	1.00 (.062)	.66	1.00 (.062)	.665
ECONOMIC ←→ MORALS	.024[a] (.118)	.009[b]	.02[a] (.12)	.01[b]	.024[a] (.118)	.009[b]
χ^2 (df=8)	35.299		35.3		35.329	
RMSEA	.054		.054		.054	

Standard errors appear in parentheses
[a] covariance; [b] correlation

Table 2 : Comparison of Unstandardized Estimates: Two Factor Model – B

	Amos		***LISREL***		Mplus	
	Unstand.	Stand	Unstand	Stand.	Unstand.	Stand.
ECONOMIC → PRIVTOWN	1.00	.615	1.00	.62	1.00	.614
ECONOMIC → GOVTRESP	.291 (.084)	.150	.29 (.084)	.15	.292 (.084)	.150
ECONOMIC → COMPETE	1.176 (.382)	.680	1.18 (.38)	.68	1.179 (.383)	.681
MORALS → HOMOSEX	1.00	.638	1.00	.64	1.00	.638
MORALS → ABORTION	1.127 (.071)	.785	1.13 (.071)	.79	1.127 (.071)	.785
MORALS → EUTHANAS	.991 (.061)	.666	.99 (.061)	.67	.991 (.061)	.666
MORALS → GOVTRESP	.227 (.044)	.176	.23 (.044)	.18	.227 (.044)	.176
ECONOMIC ←→ MORALS	-.030[a] (.126)	-.011[b]	-.03[a] (.13)	-.01[b]	-.030[a] (.125)	-.011[b]
χ^2 (df=7)	7.927		7.93		7.934	
RMSEA	.011		.011		.011	

Standard errors appear in parentheses
[a] covariance; [b] correlation

3. Confirmatory Factor Analysis with Missing Data

3.1 Preliminaries

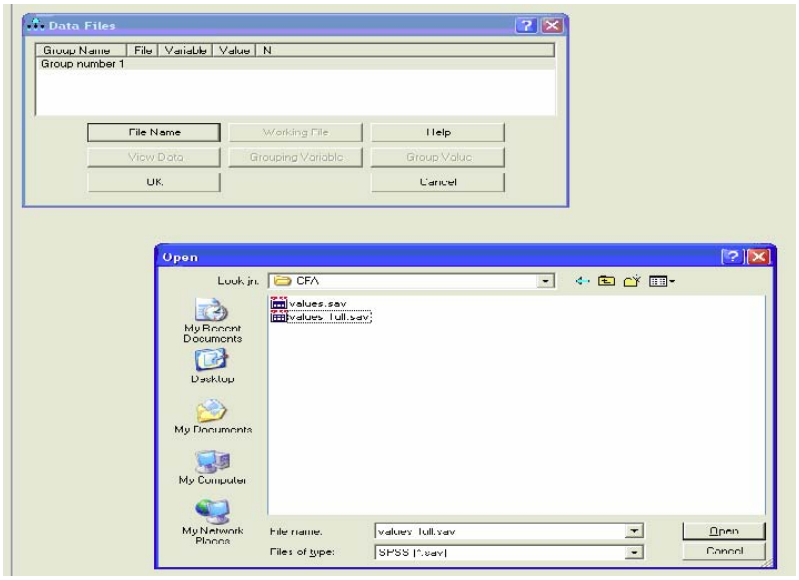
Missing data is a pervasive problem in the social sciences. A subject may fail to complete a test in an experimental setting, refuse to give an answer to a particular survey item, or drop out of a panel. In many cases, including the previous example, researchers choose to drop all observations from subjects 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.

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), an estimator 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 program. 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.

This section shows how to estimate the two-factor model for political values introduced in the previous section 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.

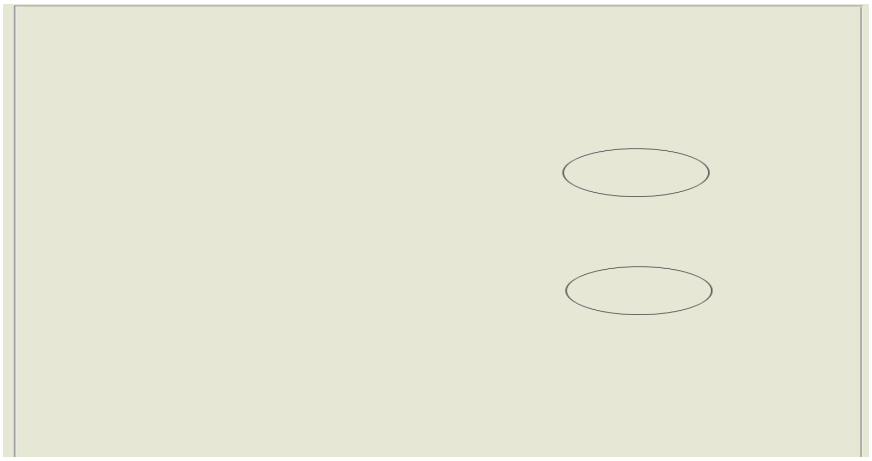
3.2 CFA with Missing Data using Amos

Launch Amos Graphics. The complete data containing missing values is saved as the SPSS file *values_full.sav* in the C:\temp\CFA folder. To load the data, choose **File →Data Files**. After the **Data Files** dialog box opens, click on **File Name**. Navigate to the CFA directory and choose the correct file.

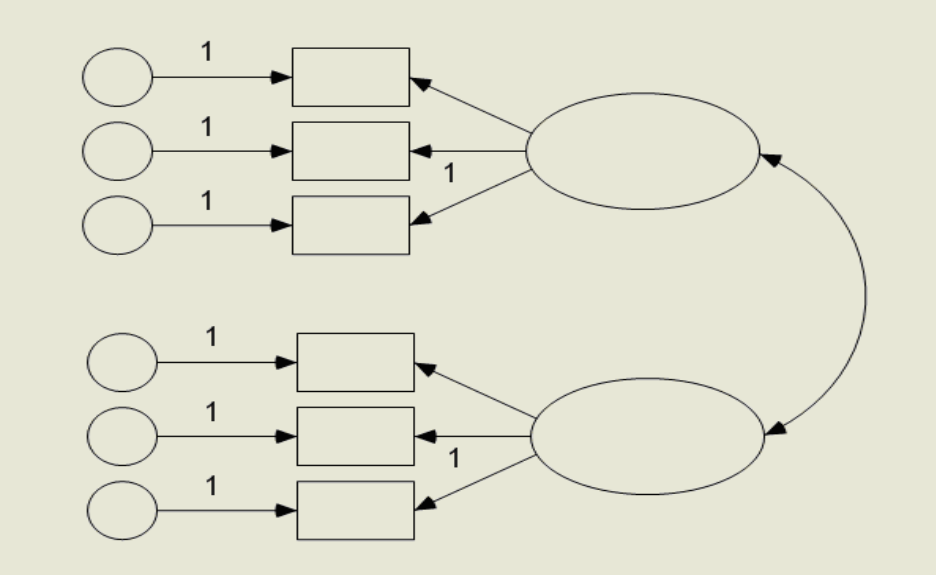


Click **Open**, Then **Okay**.

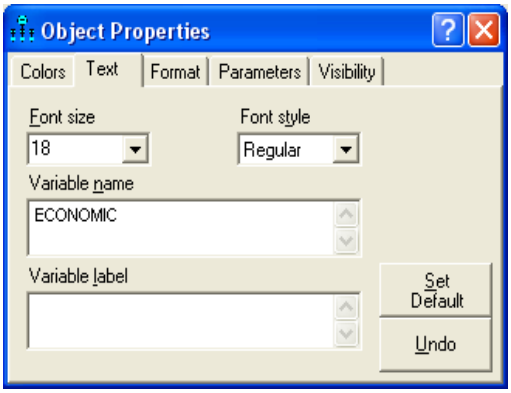
Begin drawing the path diagram by clicking on the **Add Unobserved Variable** button  and drawing an oval to represent a latent variable. Move the cursor just below the oval and click once to create a second oval of the same size.



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. By default Amos sets the metric of each error term by constraining the path parameters to one. The factor loading of the first indicator for each latent variable is also set to one. If you are not happy with where Amos added the indicators it is possible to rotate each latent variable by choosing the **Rotate the Indicators of a Latent Variable** button  and clicking each factor until you are satisfied with the appearance. Finally, add a covariance between the two common factors by choosing the **Draw Covariance** button  and drawing the two-headed arrow. Your screen should now look like the following:

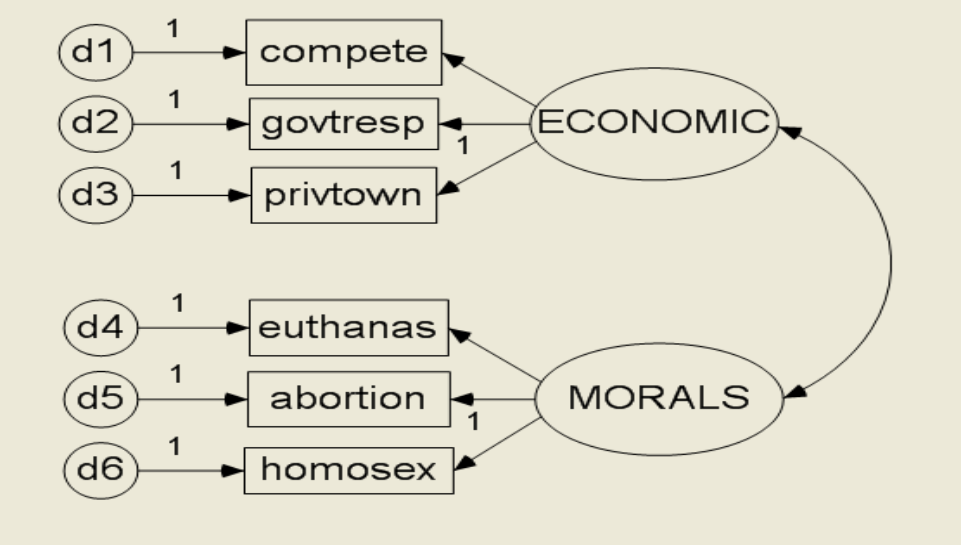


To label the latent variables, right-click in one of the ovals and choose **Object Properties**. When the **Object Properties** dialog box opens choose the **Text** tab. Name one variable ECONOMIC and the other MORALS.

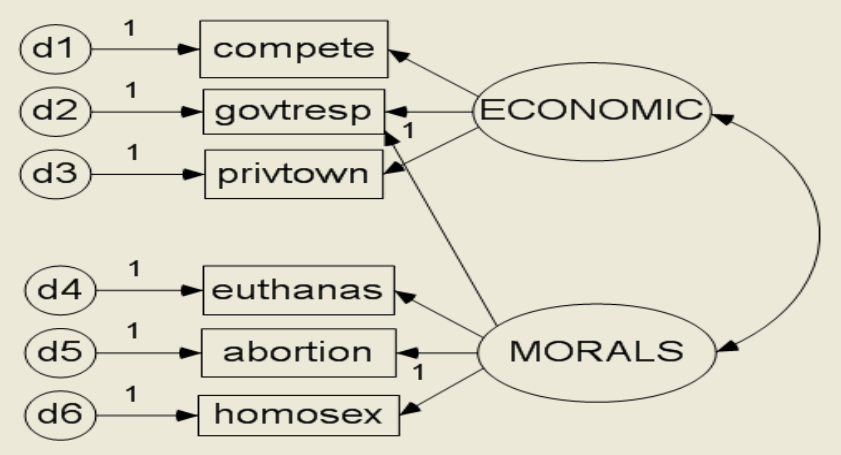


Do the same to name the error terms d_1 through d_6.

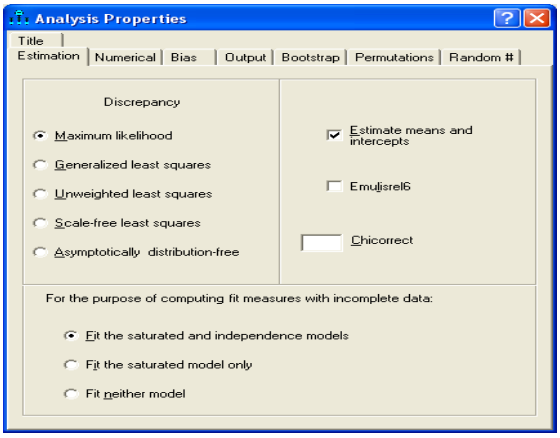
To name the observed variables choose **View → Variables in Dataset**. Click and drag the names of each variable to the appropriate box in the path diagram. If the names do not fit you can resize the box after clicking on the **Change the Shape of Objects** button  . The path diagram should look something like the following:



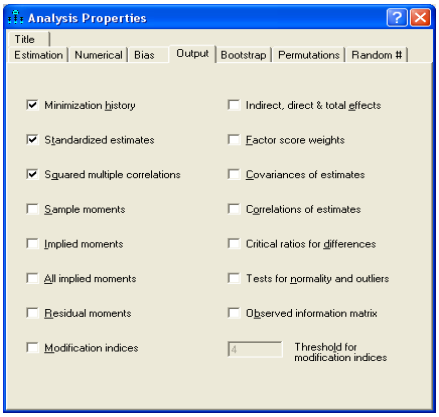
Earlier analysis suggested that GOVTRESP was not strongly tapping purely economic values, and modification indices suggested an improved model fit by adding a path connecting it to the MORALS factor (see the previous section). In the path diagram add an arrow from the MORALS latent variable to the GOVTRESP indicator.

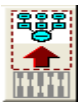


When missing values are present it is necessary to request that Amos estimate means and intercepts (required for FIML estimation), which is not the default. Choose **View →Analysis Properties**, click the **Estimation** tab in the **Analysis Properties** dialog box, and select **Estimate means and intercepts**.



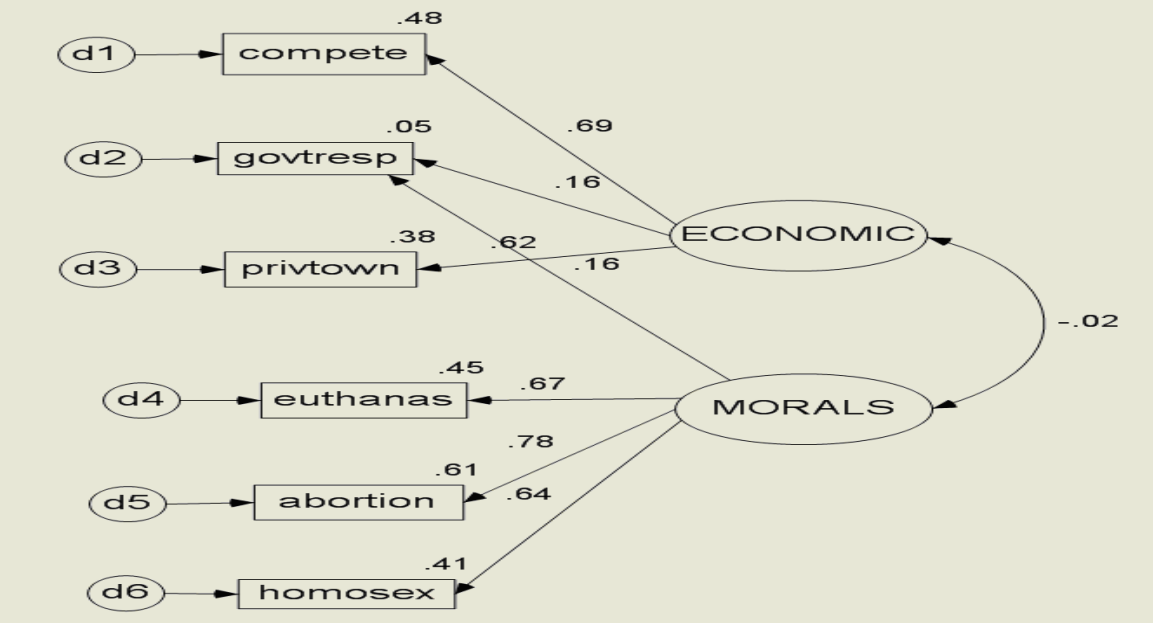
Next click on the **Output** tab. **Minimization History** is checked by default. Also place checks next to **Standardized Estimates** and **Squared Multiple Correlations**.



To estimate the model, go to **Analyze → Calculate Estimates**. To see the results in the path diagram click on the **View the Output Path Diagram** button. 

The unstandardized estimates are displayed by default. Choose instead to display the standardized estimates.

If the results are hard to read it is possible to move elements of the diagram by displaying the input path diagram, choosing the **Move Objects** button  , and changing the position of parts of the drawing. The standardized output can then be viewed again by requesting the output path diagram.



Amos now displays the standardized factor loadings, the squared multiple correlation coefficient for each observed variable, and a χ^2 statistic of model fit. 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. For this simple model, however, there is no problem. To see more detail about the results go to **View → Text Output**. A selected portion of the output is the following:

Result (Default model)

```
Minimum was achieved
Chi-square = 7.402
Degrees of freedom = 7
Probability level = .388
```

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
privtown	<---	ECONOMIC	1.000				
govtresp	<---	ECONOMIC	.306	.082	3.739	***	
compete	<---	ECONOMIC	1.201	.364	3.295	***	
homosex	<---	MORALS	1.000				
abortion	<---	MORALS	1.122	.070	15.959	***	
euthanas	<---	MORALS	.994	.061	16.366	***	
govtresp	<---	MORALS	.210	.044	4.798	***	

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

			Estimate
privtown	<---	ECONOMIC	.615
govtresp	<---	ECONOMIC	.156
compete	<---	ECONOMIC	.691

			Estimate
homosex	<---	MORALS	.637
abortion	<---	MORALS	.783
euthanas	<---	MORALS	.667
govtresp	<---	MORALS	.161

Intercepts: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
privtown	3.541	.065	54.587	***	
govtresp	4.312	.078	55.367	***	
compete	3.442	.069	49.743	***	
homosex	4.773	.095	50.493	***	
abortion	4.356	.086	50.738	***	
euthanas	5.048	.089	56.523	***	

Covariances: (Group number 1 - Default model)

		Estimate	S.E.	C.R.	P	Label
MORALS	<--> ECONOMIC	-.051	.123	-.412	.681	

Correlations: (Group number 1 - Default model)

			Estimate
MORALS	<-->	ECONOMIC	-.018

Variances: (Group number 1 - Default model)

	Estimate	S.E.	C.R.	P	Label
ECONOMIC	1.901	.596	3.187	.001	
MORALS	4.288	.421	10.196	***	
d3	3.119	.588	5.302	***	
d2	6.907	.290	23.834	***	
d1	2.997	.837	3.581	***	
d6	6.282	.348	18.034	***	
d5	3.397	.322	10.537	***	
d4	5.272	.315	16.724	***	

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

	Estimate
euthanas	.445
abortion	.614
homosex	.406
compete	.478
govtresp	.049
privtown	.379

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	20	7.402	7	.388	1.057
Saturated model	27	.000	0		
Independence model	6	1059.583	21	.000	50.456

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.007	.000	.037	.997
Independence model	.203	.193	.214	.000

The overall model fit appears quite good. The χ^2 test yields a statistic of 7.402 (df=7), which has a corresponding p-value of .388. This p-value is too high to reject the null of a good fit. Additionally the RMSEA is only .007, offering further evidence that the model fits the data well.

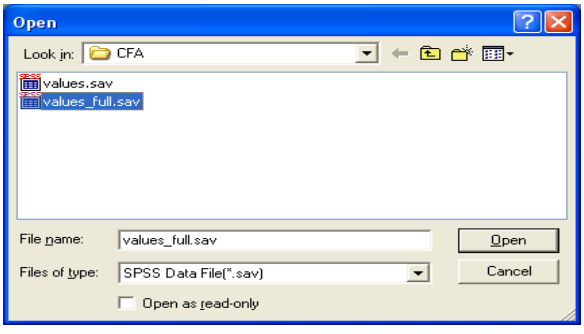
Under the `Regression Weights` heading the unstandardized loadings appear along with standard errors, a critical ratio, and p-values. The critical ratio and p-values can be used to ascertain statistical significance. A critical ratio greater than 1.96 or a p-value smaller than .05 signifies significance. Three asterisks (`***`) indicate that the p-value is smaller than .001. In this case all of the unconstrained estimates are significant.

The unstandardized weights are highly sensitive to model constraints, whereas the standardized regression weights provide more intuitive information about the strength of loadings. The GOVTRESP has low standardized loadings on both factors (.156 for ECONOMIC and .161 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 PRIVTOWN the loading is .615, for COMPETE it is .691, for HOMOSEX it is .637, for ABORTION it is .783, and for EUTHANAS it is .667.

The squared multiple correlations provide information about how much variance the factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a very low R^2 of only .049. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.379), HOMOSEX (.406), EUTHANAS (.445), COMPETE (.478), and ABORTION (.614). Finally, the correlation between the two common factors is -.018, and the covariance estimate of -.051 is not statistically distinguishable from zero (p<.823).

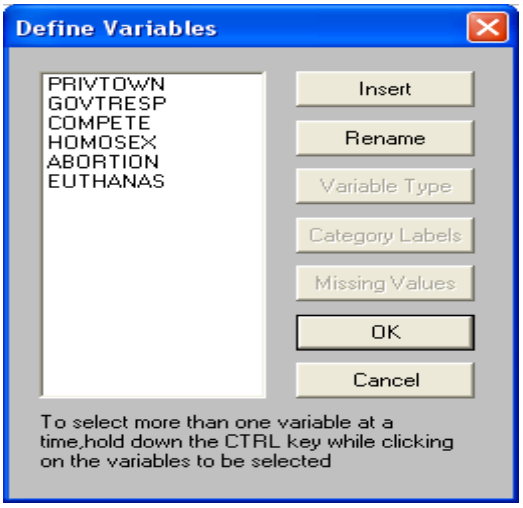
3.3 CFA with Missing Data using LISREL

After launching LISREL, open the file *values_full.sav* by choosing **File → Import External Data in Other Formats**. Change **Files of type** to *SPSS Data File(*.sav)*, navigate to the folder C:\temp\CFA, and choose the correct file.

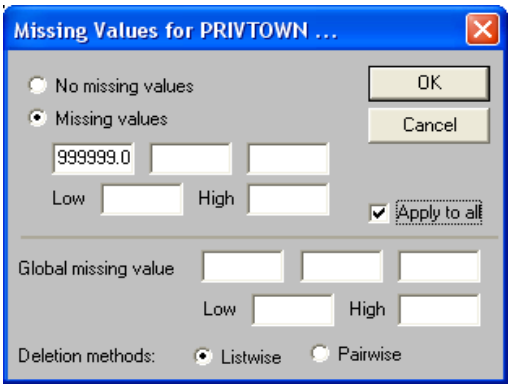


Click **Open**. A prompt appears immediately to save the file as a PRELIS system file (.psf). Enter the name *values_full* and click **Save**. A spreadsheet will open displaying the

data. Notice that missing observations are coded -999999.0. To make sure PRELIS understands these are missing values it is necessary to declare them as such. Go to **Data → Define Variables**. A **Define Variables** dialog box opens.



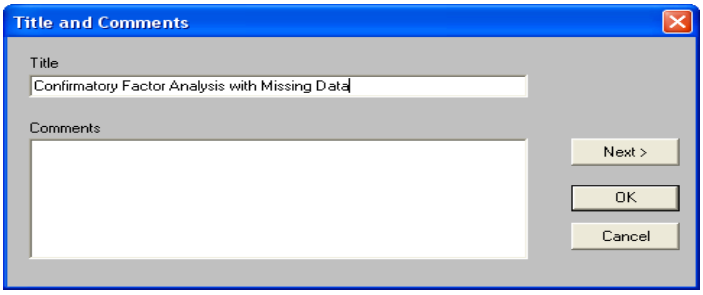
Highlight each variable name by clicking on PRIVTOWN, holding down the shift button, and clicking on EUTHANAS. Click on **Missing Values** to bring up the **Missing Values** box. Click on the **Missing Values** radio button, enter -999999.0 in the first empty field, and check the **Apply to all** option.



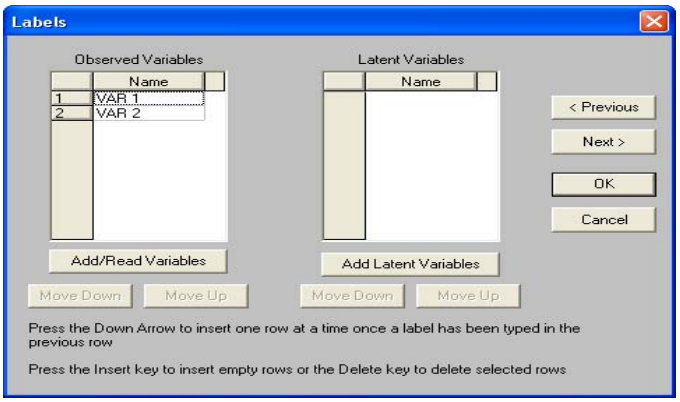
Click **OK**, then **OK** again. Save the data so that the changes take effect.

The next step is to draw the path diagram. Go to **File → New**, choose **Path Diagram**, and click **OK**. You will be prompted to save the path diagram. Name it *values_full* and click **Save**. An empty window opens where the path diagram will be drawn.

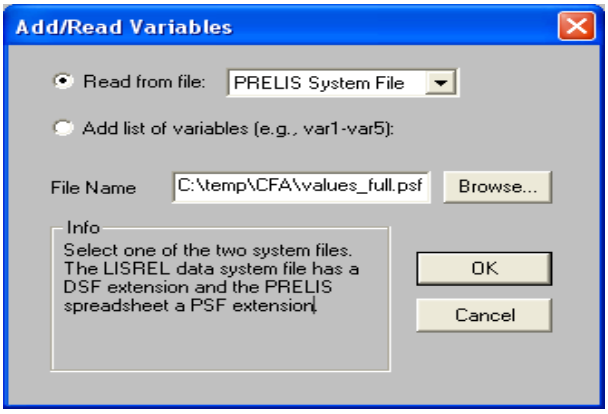
The next step is to name the observed and latent variables and tell LISREL to call the raw data from the PRELIS file. Go to **Setup → Title and Comments** to open a dialog box to label the analysis. Enter "Confirmatory Factor Analysis with Missing Data" in the **Title** field.



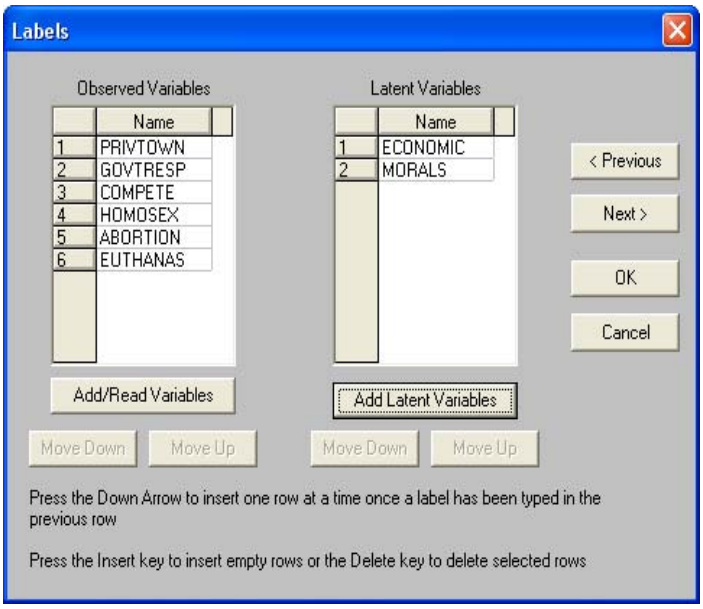
Click **Next**. This opens the **Group Names** dialog box, which is used when comparing models across different clusters of observations. Because this example is concerned with only a single sample, click **Next** to bring up the **Labels** box.



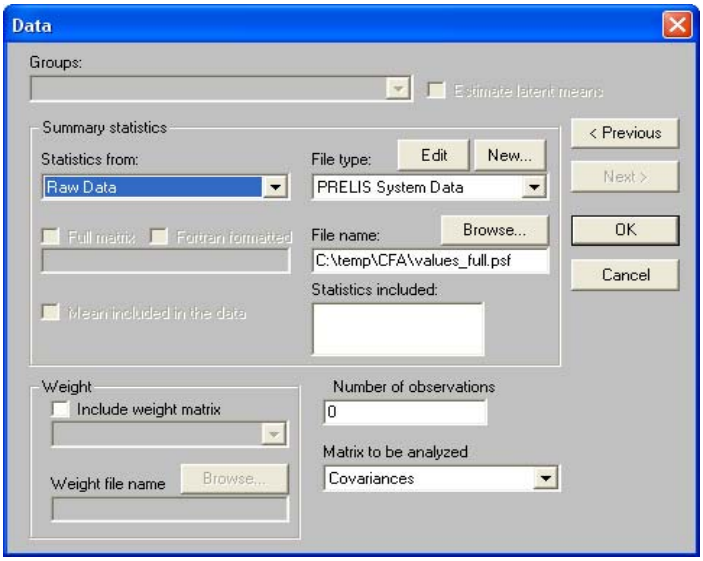
To read in the names of the observed variables, click on the **Add/Read Variables** button. Make sure the **Read from** file radio button is chosen along with the PRELIS System File option. Browse to the C:\temp\CFA directory to choose the *values_full.psf* file.



Click **OK**. The next step is to name the latent variables. Click on the **Add Latent Variables** button and add the name ECONOMIC. Click again to add MORALS.

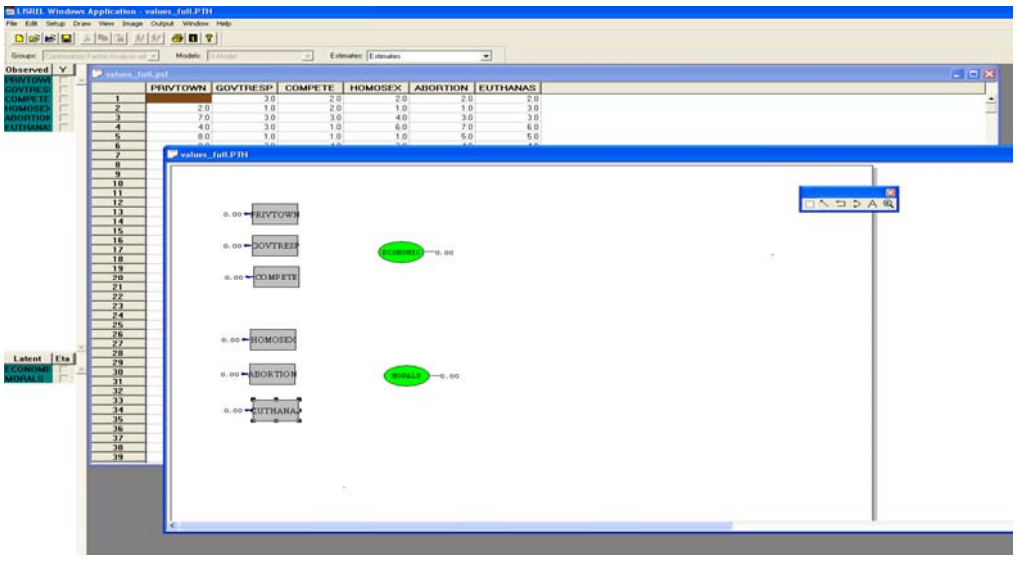


Choose **Next**. A final box appears to specify the location of the data file to be analyzed. Make sure **Raw Data** is entered in the **Statistics from** field and that **PRELIS System Data** is chosen in the **File type** field. The PRELIS system file contains information about the number of observations and missing data codes, so no further changes need to be made.

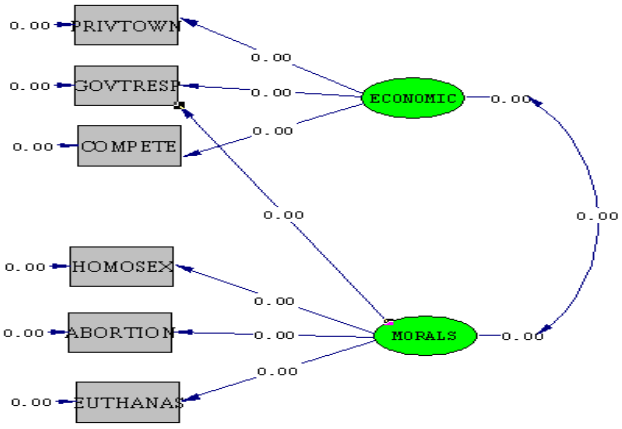


Click **OK**.

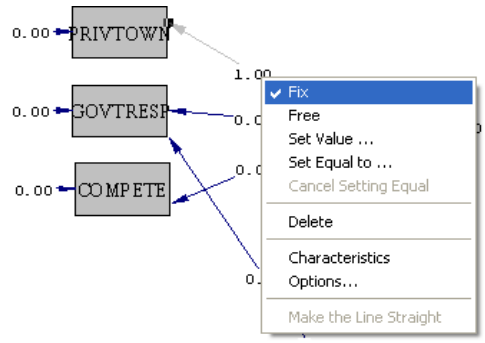
To begin drawing the path diagram, first drag each of the variable names from the left-hand side of the screen to the drawing area.



Next, draw single-headed arrows pointing from ECONOMIC to PRIVTOWN, GOVTRESP, and COMPETE. Draw additional arrows pointing from MORALS to HOMOSEX, ABORTION, EUTHANAS, and GOVTRESP. Unlike with Amos, it is not necessary to draw unique factors corresponding to measurement error in the observed indicators; LISREL includes these by default. Finally, add a two-headed arrow to represent the covariance between ECONOMIC and MORALS. The path diagram should now look like the following:

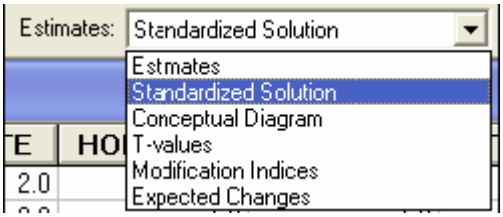


The last step is to set the metric of the two common factors by constraining factor loadings to equal one. Double-click the 0.00 on the path from ECONOMIC to PRIVTOWN and change the loading to 1.00. Right-click and choose **Fix** to constrain the loading to one.



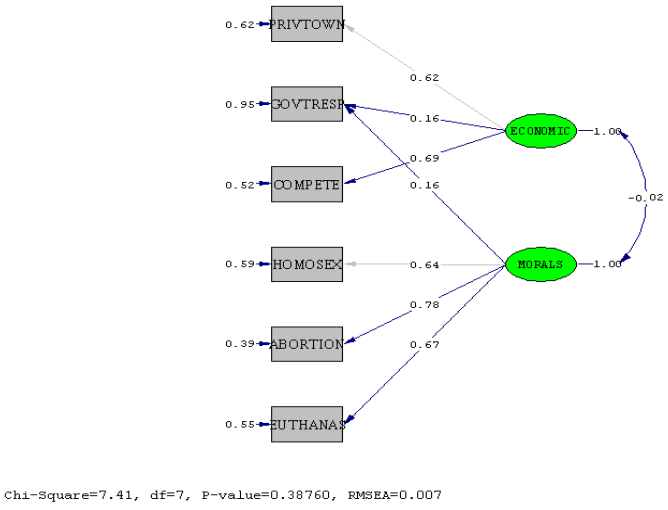
Do the same for the path between MORALS and HOMOSEX.

Before estimating it is necessary to build from the path diagram the syntax LISREL uses to estimate the model. Choose **Setup → Build SIMPLIS syntax**. This opens the syntax editor along with the commands required to estimate the model drawn in the path diagram. If we had not previously told PRELIS which codes were missing it would be necessary to add the line MISSING VALUE CODE = -999999. To begin the estimation click on the **Run LISREL** button  .



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.

The path diagram will now look like this:



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:  6/14/2006
                            TIME: 16:28


                          L I S R E L  8.70

                                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., 1981-2004
        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:

Confirmatory Factor Analysis with Missing Data
Raw Data from file 'C:\temp\CFA\values_full.psf'

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

         Number of different missing-value patterns=      11
         Convergence of EM-algorithm in    3 iterations
         -2 Ln(L) =    33675.45983
         Percentage missing values=   0.69

     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

Confirmatory Factor Analysis with Missing Data

       Covariance Matrix

          PRIVTOWN   GOVTRESP    COMPETE    HOMOSEX   ABORTION   EUTHANAS
          --------   --------   --------   --------   --------   --------
PRIVTOWN      5.02
GOVTRESP      0.55       7.27
 COMPETE      2.28       0.70       5.74
 HOMOSEX     -0.28       1.25      -0.07      10.58
ABORTION     -0.15       0.88       0.01       4.80       8.80
EUTHANAS     -0.01       0.77       0.05       4.23       4.82       9.51
```

```
Confirmatory Factor Analysis with Missing Data

Number of Iterations =  4

LISREL Estimates (Maximum Likelihood)

       Measurement Equations


PRIVTOWN = 1.00*ECONOMIC, Errorvar.= 3.12 , R² = 0.38
                             (0.59)
                              5.30

GOVTRESP = 0.31*ECONOMIC + 0.21*MORALS, Errorvar.= 6.91 , R² = 0.049
         (0.082)        (0.044)              (0.29)
          3.74           4.80                23.84

 COMPETE = 1.20*ECONOMIC, Errorvar.= 3.00 , R² = 0.48
         (0.36)                (0.84)
          3.30                  3.58

 HOMOSEX = 1.00*MORALS, Errorvar.= 6.28 , R² = 0.41
                           (0.35)
                            18.04

ABORTION = 1.12*MORALS, Errorvar.= 3.40 , R² = 0.61
         (0.070)              (0.32)
          15.97                10.54

EUTHANAS = 0.99*MORALS, Errorvar.= 5.27 , R² = 0.45
         (0.061)              (0.32)
          16.37                16.73


       Covariance Matrix of Independent Variables

          ECONOMIC    MORALS
          --------   --------
ECONOMIC      1.90
            (0.60)
              3.19

  MORALS     -0.05      4.29
            (0.12)    (0.42)
             -0.41     10.20


           Global Goodness of Fit Statistics, Missing Data Case


             -2ln(L) for the saturated model =      33675.460
             -2ln(L) for the fitted model   =      33682.869

                         Degrees of Freedom = 7
             Full Information ML Chi-Square  = 7.41 (P = 0.39)
           Root Mean Square Error of Approximation (RMSEA) = 0.0070
           90 Percent Confidence Interval for RMSEA = (0.0 ; 0.037)
            P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00


The Modification Indices Suggest to Add an Error Covariance
 Between    and    Decrease in Chi-Square    New Estimate
ABORTION  HOMOSEX          17.1              -20.72
EUTHANAS  ABORTION         27.2               13.58

                     Time used:    0.063 Seconds
```

The overall model fit appears quite good. The χ^2 test yields a value of 7.402 (df=7), which has a corresponding p-value of .388. This p-value is too high to reject the null of a good fit. Additionally the RMSEA is only .007, offering further evidence that the model fits the data well.

Under the `Measurement Equations` heading the unstandardized loadings appear along with standard errors, t-values, estimates of error variance, and squared multiple correlation coefficients (R^2). A standard error at least twice the size of the estimate can be considered evidence of significance. In this case all of the unconstrained estimates are significant. The unstandardized weights are highly sensitive to model constraints, whereas the standardized regression weights provide more intuitive information about the loadings. Standardized solutions are not printed by default in the output but can be recovered from the path diagram. They are interpreted as the correlation between the observed and unobserved variables. The GOVTRESP has low standardized loadings on both factors (.16 for ECONOMIC and .16 for MORALS), suggesting that it is a weak indicator of both economic and moral values. However, the other indicators have moderate to strong standardized loadings. For PRIVTOWN the loading is .62, for COMPETE it is .69, for HOMOSEX it is .64, for ABORTION it is .78, and for EUTHANAS it is .67.

The squared multiple correlations provide information on how much variance the factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a very low R^2 of only .049. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.38), HOMOSEX (.41), EUTHANAS (.45), COMPETE (.48), and ABORTION (.61). Finally, the correlation between the two common factors is a very small -.02 (according to the path diagram), and the covariance estimate of -.05 is not statistically distinguishable from zero (Standard Error = .12).

3.4 CFA with Missing Data using Mplus

After launching Mplus a screen displaying the Mplus syntax editor appears. Unlike Amos and LISREL, Mplus cannot directly read an SPSS file. Instead raw data must be saved as an ASCII file in free or fixed format. If a file is saved in free format Mplus does not understand blanks to be missing. Instead a number code must be entered in the raw data and specified to mean missing in the Mplus syntax. Open SPSS, choose **File →New → Syntax**, and enter the following in the SPSS Syntax Editor window:

```
RECODE
  privtown govtresp compete homosex abortion euthanas (SYSMIS=-1).

SAVE TRANSLATE OUTFILE='c:\temp\CFA\values_full.dat' /TYPE=TAB/MAP.
```

This syntax recodes missing values as -1 and writes a tab-delimited text file to the C:\temp\CFA directory.

The syntax for the confirmatory factor model to be tested is the following:

```
TITLE:      Two Factor Model with Missing Data;
DATA:       FILE IS values_full.dat;
VARIABLE:   NAMES ARE privtown govtresp compete
            homosex abortion euthanas;
            MISSING ARE all (-1);
ANALYSIS:   TYPE = missing h1
MODEL:      leftrght 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 pathname for the tab-delimited raw data file to be analyzed. The NAMES ARE portion of 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 value -1 as missing for all variables. The TYPE = missing option to 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. Convergence may be very slow for models with many parameters and missing data if a χ^2 test is requested, and in such cases the *h1* may be excluded. The MODEL statement tells Mplus that there are two latent predictor variables and six outcome variables. The OUTPUT statement here requests that standardized estimates and modification indices appear in the output file. See section 2.4 above for additional information on the rules of Mplus syntax.

To begin the estimation click on **Run**  . A text output file appears and is saved in the working directory. For this model, the file is the following:

```
Mplus VERSION 4.0
MUTHEN & MUTHEN
06/15/2006  10:15 AM

INPUT INSTRUCTIONS

  TITLE:     Two Factor Model with Missing Data;
  DATA:      FILE IS values_full.dat;
  VARIABLE:   NAMES ARE privtown govtresp compete
             homosex abortion euthanas;
             MISSING ARE all (-1);
  ANALYSIS:   TYPE = missing h1
  MODEL:     leftrght BY privtown govtresp compete;
             morals BY homosex abortion euthanas govtresp;
  OUTPUT:     standardized;
             modindices;


INPUT READING TERMINATED NORMALLY


Single Factor Model;

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

  Continuous
   PRIVTOWN    GOVTRESP    COMPETE     HOMOSEX     ABORTION    EUTHANAS

Continuous latent variables
   LEFTRGHT    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)
  values_full.dat

Input data format  FREE


SUMMARY OF DATA

     Number of patterns        11


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       0.986        0.991        0.991        0.976        0.985


           Covariance Coverage
              EUTHANAS
              ________
 EUTHANAS       0.992


THE MODEL ESTIMATION TERMINATED NORMALLY


TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                          7.409
          Degrees of Freedom                 7
          P-Value                       0.3875

Chi-Square Test of Model Fit for the Baseline Model
```

```
        Value                          1060.466
        Degrees of Freedom                   15
        P-Value                         0.0000

CFI/TLI

        CFI                             1.000
        TLI                             0.999

Loglikelihood

        H0 Value                   -16841.434
        H1 Value                   -16837.730

Information Criteria

        Number of Free Parameters           20
        Akaike (AIC)                33722.868
        Bayesian (BIC)              33824.670
        Sample-Size Adjusted BIC      33761.142
          (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)

        Estimate                        0.007
        90 Percent C.I.                 0.000  0.037
        Probability RMSEA <= .05          0.997

SRMR (Standardized Root Mean Square Residual)

        Value                           0.012


MODEL RESULTS

                    Estimates    S.E.  Est./S.E.    Std    StdYX

 LEFTRGHT BY
   PRIVTOWN          1.000    0.000     0.000    1.378    0.615
   GOVTRESP          0.306    0.082     3.712    0.421    0.156
   COMPETE           1.201    0.373     3.221    1.655    0.691

 MORALS   BY
   HOMOSEX           1.000    0.000     0.000    2.071    0.637
   ABORTION          1.122    0.070    15.971    2.323    0.783
   EUTHANAS          0.994    0.061    16.386    2.057    0.667
   GOVTRESP          0.210    0.044     4.817    0.434    0.161

 MORALS   WITH
   LEFTRGHT         -0.050    0.127    -0.395   -0.018   -0.018

 Intercepts
   PRIVTOWN          3.541    0.065    54.608    3.541    1.580
   GOVTRESP          4.312    0.078    55.390    4.312    1.600
   COMPETE           3.442    0.069    49.763    3.442    1.437
   HOMOSEX           4.773    0.094    50.514    4.773    1.468
   ABORTION          4.356    0.086    50.759    4.356    1.469
   EUTHANAS          5.048    0.089    56.546    5.048    1.637

 Variances
   LEFTRGHT          1.900    0.610     3.117    1.000    1.000
   MORALS            4.288    0.420    10.213    1.000    1.000

 Residual Variances
   PRIVTOWN          3.120    0.601     5.191    3.120    0.622
   GOVTRESP          6.907    0.290    23.836    6.907    0.950
   COMPETE           2.996    0.856     3.501    2.996    0.522
   HOMOSEX           6.282    0.349    17.998    6.282    0.594
   ABORTION          3.396    0.320    10.609    3.396    0.386
   EUTHANAS          5.272    0.312    16.891    5.272    0.555
```

```
R-SQUARE

    Observed
    Variable  R-Square

    PRIVTOWN    0.378
    GOVTRESP    0.050
    COMPETE     0.478
    HOMOSEX     0.406
    ABORTION    0.614
    EUTHANAS    0.445


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.



     Beginning Time:  10:15:02
        Ending Time:  10:15:02
       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
```

The overall model fit appears quite good. The χ^2 test yields a value of 7.409 (df=7), which has a corresponding p-value of .388. This p-value is too high to reject the null of a good fit. Additionally the RMSEA is only .007, offering further evidence that the model fits the data well.

Under the `Model Results` heading the unstandardized loadings appear along with standard errors, the ratio of the estimates to their standard errors, and two standardized estimates. The `Est./S.E.` column can be used to evaluate significance. 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. In this case all of the unconstrained estimates have `Est/SE` ratios greater than 1.96.

The unstandardized weights are highly sensitive to model constraints, whereas the standardized regression weights provide more intuitive information about the strength of loadings. The column `StdYX` can be interpreted as the correlation between the latent and observed variables and is equivalent to the standardized estimates provided by Amos and LISREL. GOVTRESP has low standardized loadings on both factors (.156 for ECONOMIC and .161 for MORALS), suggesting that it is a weak indicator of both economic and moral values. However, the other indicators have moderate to strong

standardized loadings. For PRIVTOWN the loading is .615, for COMPETE it is .691, for HOMOSEX it is .637, for ABORTION it is .783, and for EUTHANAS it is .667.

The squared multiple correlations provide information on how much variance the factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a very low R^2 of only .050. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.378), HOMOSEX (.406), EUTHANAS (.445), COMPETE (.478), and ABORTION (.614). Finally, the correlation between the two common factors is a very small -.018, and the covariance estimate of -.050 is not statistically distinguishable from zero.

3.5 Summary

Table 3 compares the unstandardized and standardized estimates that each package produced. Standard errors appear in parentheses, and the final two rows list χ^2 and RMSEA estimates. The results are essentially identical across programs.

Table 3: Comparison of Unstandardized Estimates: Two Factor Model with Missing Data

	Amos		***LISREL***		Mplus	
	Unstand.	Stand	Unstand	Stand.	Unstand.	Stand.
ECONOMIC → PRIVTOWN	1.00	.615	1.00	.62	1.00	.615
ECONOMIC → GOVTRESP	.306 (.082)	.156	.31 (.082)	.16	.306 (.082)	.156
ECONOMIC → COMPETE	1.201 (.364)	.691	1.20 (.36)	.69	1.201 (.373)	.691
MORALS → HOMOSEX	1.00	.637	1.00	.64	1.00	.637
MORALS → ABORTION	1.122 (.070)	.783	1.12 (.070)	.78	1.122 (.070)	.783
MORALS → EUTHANAS	.994 (.061)	.667	.99 (.061)	.67	.994 (.061)	.667
MORALS → GOVTRESP	.210 (.044)	.161	.21 (.044)	.16	.210 (.044)	.161
ECONOMIC ←→ MORALS	-.051[a] (.123)	-.018[b]	-.05[a] (.12)	-.02[b]	-.050[a] (.127)	-.018[b]
χ^2 (df=7)	7.402		7.41		7.409	
RMSEA	.007		.007		.007	

Standard errors appear in parentheses
[a] covariance; [b] correlation

4. CFA with Categorical indicators

4.1 Preliminaries

The maximum likelihood estimation approaches used in the previous two sections relied on the rather 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 alternative estimators are more appropriate.

This section continues to use the political values example developed above, but with scores recoded on a three point scale. 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. Initially observations with missing data are dropped (40 of the original 1200) in order to focus only on the problem of categorical outcome variables in the confirmatory factor model. However, the final subsection will show how to request pairwise rather than listwise deletion in Mplus so as to use as much information in the raw data file as possible. This section does not consider Amos, which lacks a means for estimating the polychoric correlation matrix used by the other two programs in the presence of categorical data. The assumption in this section continues to be, as in the previous examples, that the latent variables represent continuous (not categorical) concepts.

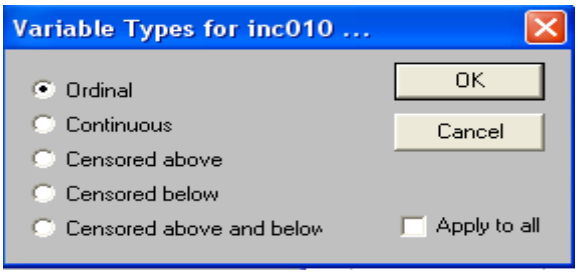
Mplus and LISREL employ a multi-step method for ordinal outcome variables that analyzes a matrix of polychoric correlations rather than covariances. This approach works as follows: 1) thresholds are estimated by maximum likelihood, 2) these estimates are used to estimate a polychoric correlation matrix, which in turn is used to 3) estimate parameters through weighted least squares using the inverse of the asymptotic covariance matrix as the weight matrix (Muthén, 1984; Jöreskog, 1990). In Mplus these steps take place automatically when the syntax includes a line identifying outcomes as categorical. In LISREL the polychoric correlation matrix and asymptotic covariance matrix must first be estimated using PRELIS. This will produce a LISREL system file (.dsf) containing both the polychoric correlations and information about where the covariance matrix is saved. The information in the .dsf file is then used by LISREL when the weighted least squares estimator is requested.

The recoded data file that will be used for these examples is the SPSS file *values_ord.sav* located in the C:\temp\CFA directory.

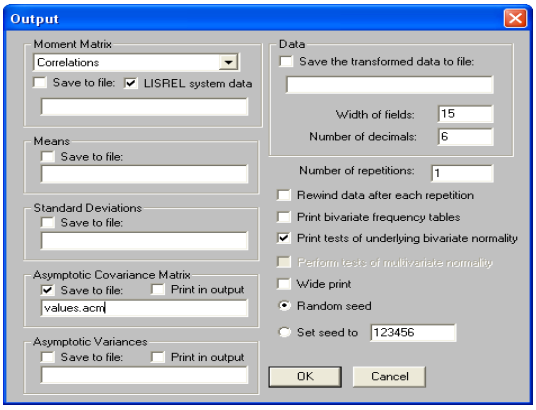
4.2 CFA with Categorical Indicators Using LISREL

Launch LISREL and load the data by choosing **File → Import External Data in Other Formats**. Navigate to the C:\temp\CFA folder, choose the file *values_ord.sav*, and click Open. When prompted, enter the name *values_ord* in the **File Name** field and click **Save**. The data will then be displayed in spreadsheet form. It is possible to define the variables

in the dataset as ordinal by going to **Data → Define Variables**, highlighting all the variables, and opening the **Variable Type** menu.



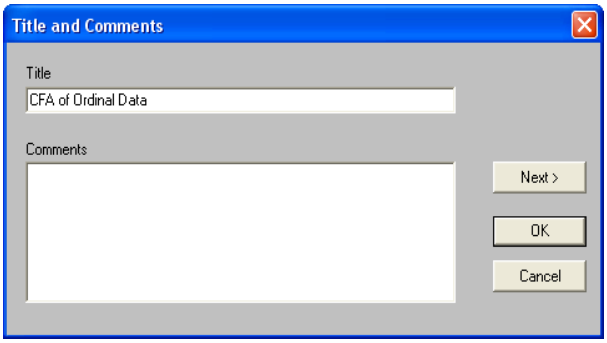
The next step is to estimate the polychoric correlation and asymptotic covariance matrices. Go to **Statistics → Output Options**. In the dialog box that opens choose **Correlations** from the **Moment Matrix** drop-down menu and click in the empty box next to **LISREL system file**. Also place a check in the box next to **Save to file** under **Asymptotic Covariance Matrix** and enter the name *values.acm*.



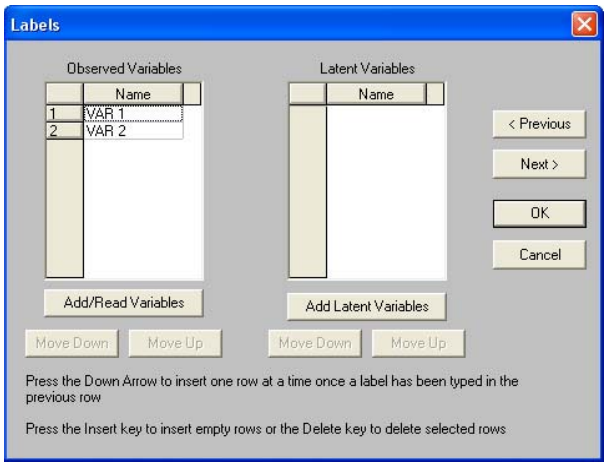
Click OK to save both the LISREL system file and the asymptotic covariance matrix to the working directory, in this case C:\temp\CFA.

The next step is to create the path diagram of the confirmatory factor model. Click on **File → New** and choose **Path diagram**. When prompted, save the diagram. Give it the name *values_ord* and click Save. An empty drawing space will open.

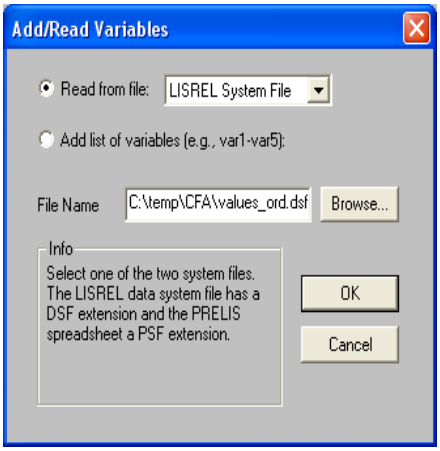
Go to **Setup → Titles and Comments** to give the model a name. In the box that appears, enter "CFA of Ordinal Data" in the **Title** field.



Click **Next**. The **Group Names** box opens but can be ignored because the model will be tested on only a single sample. Click **Next** again. The **Labels** box now displays.

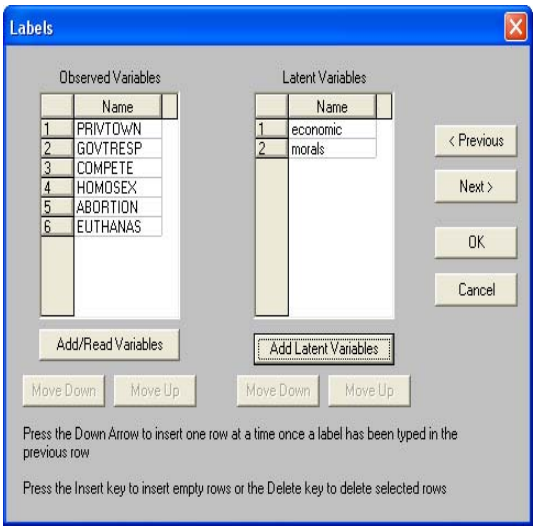


The names of the observed variables are saved in the LISREL system file created previously. Click the **Add/Read Variable** button to access them.

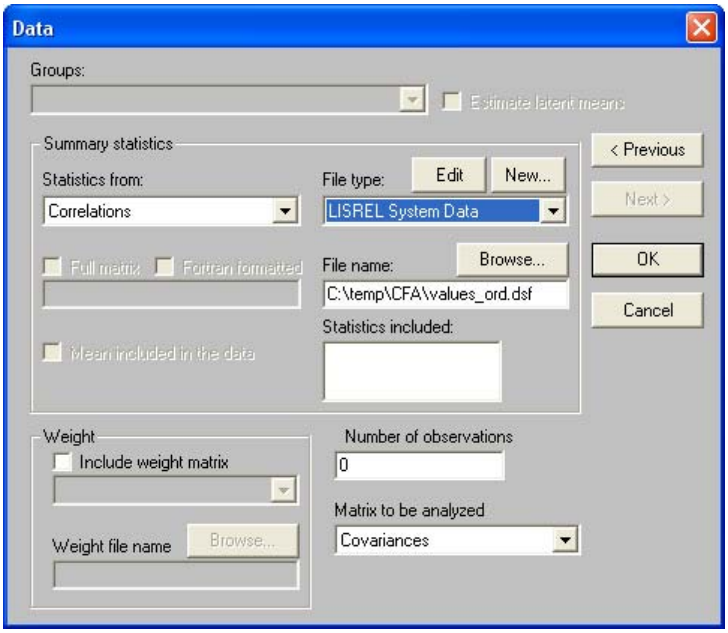


In the Add/Read Variables box make sure the **Read from file** radio button is selected and **LISREL System File** is chosen from the drop-down menu. Browse to the location where the LISREL System File was saved, choose *values_ord.dsf*, and click **OK**. The observed variable names will then appear in the **Labels** box. To add the names of the latent

variables click on **Add Latent Variables**. In the box that appears enter Economic and click **OK**. Do the same to add the label Morals.

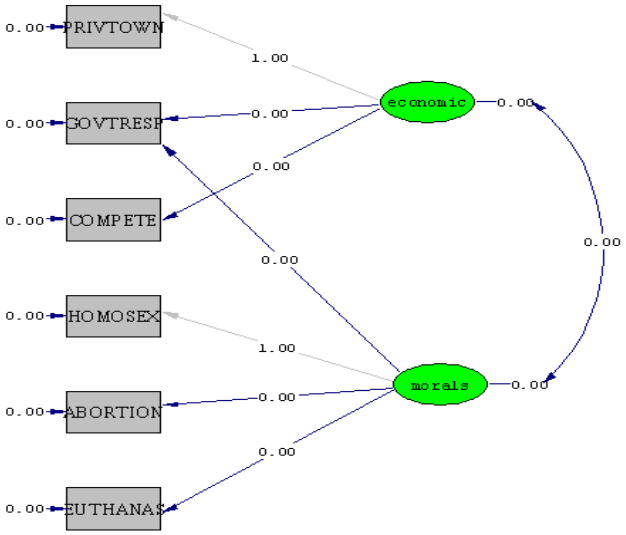


When finished, choose **Next**. The **Data** dialog box then appears. Choose **Correlations** in the **Statistics from** drop-down menu and make sure **LISREL System Data** is chosen under **File type**.

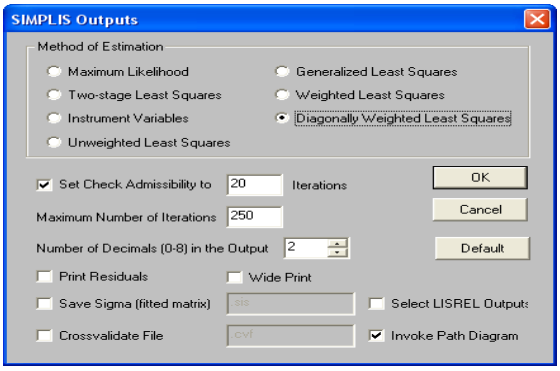


Click **OK**. Next create the path diagram by dragging the names of each of the observed and latent variables to the drawing space. As in the previous examples, add single-headed arrows pointing from the latent variables to the observed variables and a double-headed arrow between ECONOMIC and MORALS. Finally, set the scale of the latent variables by constraining the path from ECONOMIC to PRIVTOWN and from MORALS to HOMOSEX to one. This is done by first double-clicking the weight, changing it to 1.00,

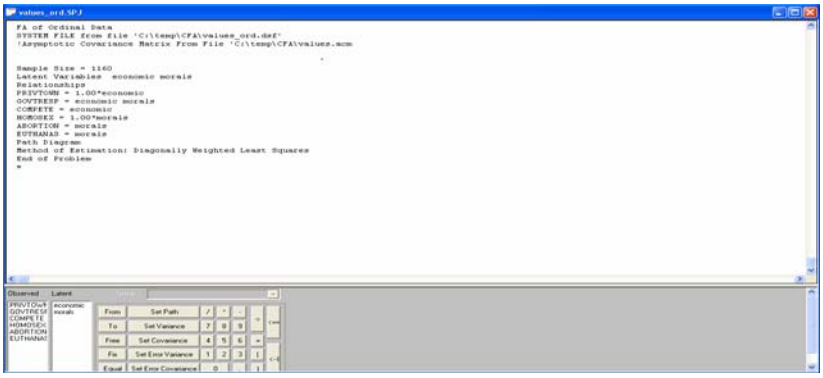
right-clicking, and choosing **Fix**. When finished the path diagram should look like the following:



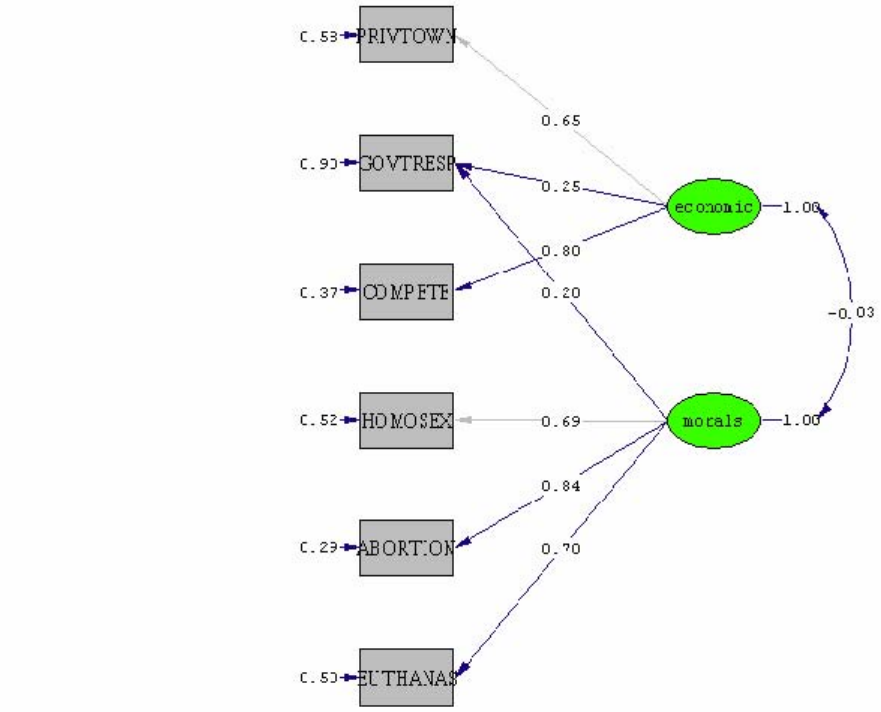
To request the weighted least squares estimator go to Output → **SIMPLIS Outputs**. Choose **Diagonally Weighted Least Squares**.



Click **OK**. To create the SIMPLIS syntax based on the path diagram go to **Setup →Build SIMPLIS syntax**. The SIMPLIS syntax editor will open displaying the commands needed for estimation.



Click the **Run LISREL** button . The unstandardized solution appears in the path diagram along with χ^2 and RMSEA statistics to assess model fit. To view the standardized estimates choose **Standardized Solution** from the **Estimates** drop-down menu.



Chi-Square=6.23, df=7, P-value=0.51363, RMSEA=0.000

Additional information can be read from the text output file *values_ord.out* automatically saved in the working directory.

```
 DATE:  6/19/2006
 TIME: 10:29


 L I S R E L  8.70

      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., 1981-2004
  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_ord.SPJ:

CFA for Ordinal Data
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

Sample Size =  1160

CFA for Ordinal Data

       Correlation Matrix

          PRIVTOWN   GOVTRESP    COMPETE    HOMOSEX   ABORTION   EUTHANAS
          --------   --------   --------   --------   --------   --------
PRIVTOWN      1.00
GOVTRESP      0.16       1.00
 COMPETE      0.51       0.20       1.00
 HOMOSEX     -0.06       0.18      -0.03       1.00
ABORTION     -0.02       0.15      -0.01       0.58       1.00
EUTHANAS      0.00       0.10       0.00       0.48       0.60       1.00


CFA for Ordinal Data

Number of Iterations =  5

LISREL Estimates (Robust Diagonally Weighted Least Squares)

       Measurement Equations


PRIVTOWN = 1.00*economic, Errorvar.= 0.58 , R² = 0.42
                             (0.13)
                              4.53

GOVTRESP = 0.39*economic + 0.29*morals, Errorvar.= 0.90  , R² = 0.10
         (0.084)        (0.055)              (0.064)
          4.65           5.31                14.06

 COMPETE = 1.23*economic, Errorvar.= 0.37 , R² = 0.63
         (0.32)                (0.17)
          3.90                  2.13

 HOMOSEX = 1.00*morals, Errorvar.= 0.52  , R² = 0.48
                            (0.071)
                             7.32

ABORTION = 1.22*morals, Errorvar.= 0.29  , R² = 0.71
         (0.071)              (0.076)
          17.29                3.77

EUTHANAS = 1.02*morals, Errorvar.= 0.50  , R² = 0.50
         (0.057)              (0.072)
          17.96                7.00



       Covariance Matrix of Independent Variables
```

```
          economic    morals
          --------   --------
economic      0.42
            (0.11)
             3.66

  morals     -0.02      0.48
            (0.02)     (0.04)
             -0.68     11.80


                    Goodness of Fit Statistics

                      Degrees of Freedom = 7
    Normal Theory Weighted Least Squares Chi-Square = 14.35 (P = 0.045)
          Satorra-Bentler Scaled Chi-Square = 6.23 (P = 0.51)
       Chi-Square Corrected for Non-Normality = 7.05 (P = 0.42)
           Estimated Non-centrality Parameter (NCP) = 0.0
        90 Percent Confidence Interval for NCP = (0.0 ; 9.18)

                Minimum Fit Function Value = 0.0034
             Population Discrepancy Function Value (F0) = 0.0
        90 Percent Confidence Interval for F0 = (0.0 ; 0.0079)
        Root Mean Square Error of Approximation (RMSEA) = 0.0
       90 Percent Confidence Interval for RMSEA = (0.0 ; 0.034)
         P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00

            Expected Cross-Validation Index (ECVI) = 0.030
       90 Percent Confidence Interval for ECVI = (0.030 ; 0.038)
                 ECVI for Saturated Model = 0.036
               ECVI for Independence Model = 1.34

  Chi-Square for Independence Model with 15 Degrees of Freedom = 1540.28
                   Independence AIC = 1552.28
                       Model AIC = 34.23
                     Saturated AIC = 42.00
                   Independence CAIC = 1588.62
                       Model CAIC = 119.01
                     Saturated CAIC = 169.18

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

                    Critical N (CN) = 3440.37


            Root Mean Square Residual (RMR) = 0.017
                  Standardized RMR = 0.017
                Goodness of Fit Index (GFI) = 1.00
           Adjusted Goodness of Fit Index (AGFI) = 1.00
          Parsimony Goodness of Fit Index (PGFI) = 0.33

                 Time used:    0.016 Seconds
```

Under the `Measurement Equations` heading appear the unstandardized estimates, standard errors, t-values, and R^2 statistics. Statistical significance can be assessed by looking at the size of the standard errors (in parentheses) relative to the unstandardized parameter estimates. When the latter are twice as large as the former the estimates can be considered significant at the .05 level. For this example all of the unconstrained path coefficients are significant. In addition, the (Satorra-Bentler Scaled) χ^2 statistic of 6.23

(df=7) has a corresponding p-value of .51. This is not large enough to reject the null of a good fit. The RMSEA is 0.00, indicating a very good model fit.

The standardized estimates of the loadings displayed in the path diagram can be interpreted as the correlations between the observed and the latent variables. The GOVTRESP variable has a relatively weak correlation with both ECONOMIC (.25) and MORALS (.20). The remaining variables, however, have moderate to strong loadings on the respective common factor. The standardized loading for PRIVTOWN is .65; for COMPETE it is .80; for HOMOSEX it is .69; for ABORTION it is .84; and for EUTHANAS it is .70. The R^2 statistics listed in the output file are interpreted as the amount of variance in the observed variables accounted for by the latent variables. Despite receiving a path from both ECONOMIC and MORALS, the GOVTRESP has the smallest R^2 (.10). The other observed variables have moderate to high R^2 statistics. The multiple correlation coefficient for PRIVTOWN is .42; for COMPETE it is .63; for HOMOSEX it is .48; for ABORTION it is .71; and for EUTHANAS it is .50. Finally, the correlation between the two factors is -.03. Comparing the covariance estimate to its standard error in the output, this is not significant.

4.3 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 2.4 and 3.4 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:      Factor Analysis with Categorical Outcome Variables;
DATA:       FILE IS 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 an extra line is added to the VARIABLE statement to define the observed variables as categorical. The default in Mplus is to assume all observed variables are continuous unless specified otherwise. 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. After clicking on **Run**  to carry out the estimation, a text output file is produced. A selection of the output file is the following:

```
INPUT READING TERMINATED NORMALLY

Factor Analysis with Categorical Outcome Variables;

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)
  values_ord.dat

Input data format  FREE


THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

        Value                             5.707*
        Degrees of Freedom                    6**
        P-Value                           0.4566

*   The chi-square value for MLM, MLMV, MLR, ULS, 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, ULS 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                          1245.185
        Degrees of Freedom                   10
        P-Value                          0.0000

CFI/TLI

        CFI                              1.000
        TLI                              1.000

Number of Free Parameters                        8

RMSEA (Root Mean Square Error Of Approximation)
```

```
        Estimate                       0.000

SRMR (Standardized Root Mean Square Residual)

        Value                          0.017

WRMR (Weighted Root Mean Square Residual)

        Value                          0.531


MODEL RESULTS

                Estimates    S.E.  Est./S.E.    Std    StdYX

ECONOMIC BY
   PRIVTOWN         1.000   0.000     0.000   0.645   0.645
   GOVTRESP         0.390   0.074     5.287   0.252   0.252
   COMPETE          1.237   0.292     4.239   0.797   0.797

 MORALS   BY
   HOMOSEX          1.000   0.000     0.000   0.690   0.690
   ABORTION         1.224   0.070    17.546   0.844   0.844
   EUTHANAS         1.022   0.051    20.182   0.706   0.706
   GOVTRESP         0.294   0.056     5.275   0.203   0.203

 MORALS   WITH
   LEFTRGHT        -0.015   0.021    -0.723  -0.034  -0.034

 Variances
   LEFTRGHT         0.416   0.103     4.019   1.000   1.000
   MORALS           0.476   0.038    12.625   1.000   1.000


R-SQUARE

   Observed  Residual
   Variable  Variance  R-Square

   PRIVTOWN    0.584    0.416
   GOVTRESP    0.899    0.101
   COMPETE     0.364    0.636
   HOMOSEX     0.524    0.476
   ABORTION    0.287    0.713
   EUTHANAS    0.502    0.498
```

The overall model fit appears quite good. The χ^2 test yields a value of 5.707 with 6 degrees of freedom, which has a corresponding p-value of .4566. This p-value is too high to reject the null of a good fit. In addition the RMSEA is 0.00, indicating a very good fit.

Under the `Model Results` heading the unstandardized loadings appear along with standard errors, the ratio of the estimates to their standard errors, and two standardized estimates. The `Est./S.E.` column can be used to evaluate significance. 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. In this case all of the unconstrained loadings estimates are significant.

The column `StdYX` can be interpreted as the correlation between the latent and observed variables and is comparable to the standardized estimates provided by LISREL. GOVTRESP has relatively low standardized loadings on both factors (.252 for ECONOMIC and .203 for MORALS). For PRIVTOWN the loading is .645, for

COMPETE it is .797, for HOMOSEX it is .690, for ABORTION it is .844, and for EUTHANAS it is .706.

The squared multiple correlations provide information on how much variance the factors account for in the observed variables. Despite receiving a path from both latent variables, GOVTRESP has a low R^2 of only .101. The remaining R^2 statistics are, in order of increasing magnitude, PRIVTOWN (.416), HOMOSEX (.476), EUTHANAS (.498), COMPETE (.636), and ABORTION (.713). Finally, the correlation between the two common factors is a very small -.034, and the covariance estimate of -.015 is not statistically distinguishable from zero.

4.4 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. However, in order to focus specifically on the issue of categorical indicators all cases with missing observations on at least one indicator were dropped, reducing the original sample of 1200 to 1160. It is possible to maximize the information available in the raw data file using Mplus by adding an ANALYSIS statement specifying TYPE = missing h1, which has the effect of using pairwise rather than listwise deletion for missing observations. Thus rather than losing all information about cases with missing data on a single variable, correlations will be estimated using all cases with complete observations available on both variables (that is, even if there is missingness on a third variable). The Mplus syntax is the following:

```
TITLE:      Single Factor Model;
DATA:       FILE IS values_full.dat;
VARIABLE:   NAMES ARE privtown govtresp compete
            homosex abortion euthanas;
            CATEGORICAL ARE privtown govtresp
            compete homosex abortion euthanas;
            MISSING ARE all (-1);
ANALYSIS:   TYPE = missing h1
MODEL:      leftrght BY privtown govtresp compete;
            morals BY homosex abortion euthanas govtresp;
OUTPUT:     standardized;
            modindices;
```

The full output will not be displayed, but the results are summarized in Table 4 below.

4.5 SUMMARY

Table 4 summarizes results from the LISREL and Mplus analyses using listwise deletion plus the Mplus results using pairwise deletion. Comparing the listwise columns it is evident that the parameter estimates are equal but that Mplus produces smaller standard errors. Additionally Mplus reports a different χ^2 than LISREL and uses a different formula for arriving at the degrees of freedom (the Mplus output includes a note pointing the user to the location of the formula in the *User's Guide*). The two Mplus columns

comparing results from listwise and pairwise deletion show roughly similar estimates, though they are not equivalent. The pairwise column should be considered more accurate because it is able to incorporate the most information from the raw data.

Table 4: Comparison of Unstandardized Estimates: Two Factor Model with Ordinal Indicators

	***LISREL* (listwise)**		***Mplus* (listwise)**		**Mplus (pairwise)**	
	Unstand.	Stand	Unstand	Stand.	Unstand.	Stand.
ECONOMIC → PRIVTOWN	1.00	.65	1.00	.645	1.00	.657
ECONOMIC → GOVTRESP	.39 (.084)	.25	.390 (.074)	.252	.390 (.072)	.256
ECONOMIC → COMPETE	1.23 (.32)	.80	1.237 (.292)	.797	1.197 (.270)	.787
MORALS → HOMOSEX	1.00	.69	1.00	.690	1.00	.690
MORALS → ABORTION	1.22 (.071)	.84	1.224 (.070)	.844	1.218 (.069)	.840
MORALS → EUTHANAS	1.02 (.057)	.70	1.022 (.051)	.706	1.028 (.051)	.709
MORALS → GOVTRESP	.29 (.055)	.20	.294 (.056)	.203	.279 (.055)	.193
ECONOMIC ←→ MORALS	-.02[a] (.02)	-.03[b]	-.015[a] (.021)	-.034[b]	-.019[a] (.021)	-.042[b]
χ^2	6.23 (df=7)		5.707 (df=6)		7.409	
RMSEA	.000		.000		.000	

Standard errors appear in parentheses
[a] covariance; [b] correlation

5. 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, available in most general statistics packages, is a technique for identifying structure in data and generating hypotheses. Confirmatory factor analysis differs in that it is much more theory driven and is generally used to test explicit hypotheses.

Confirmatory factor analysis is the basis of the measurement model in full structural equation modeling (SEM) and can be estimated using SEM software. Three SEM programs supported by Indiana University are Amos, LISREL, and Mplus. Of these, Amos and LISREL are the most user-friendly, although Mplus syntax is not at all difficult to learn. Both Amos and LISREL can read in raw data from a variety of different programs, and both allow the user to estimate models by simply drawing a path diagram. However, Amos cannot accurately estimate models when the observed variables are categorical. PRELIS/LISREL and Mplus handle ordinal indicators by first estimating a polychoric correlation matrix and an asymptotic covariance matrix and employing these within a weighted least squares estimator. In LISREL the user must create these matrices first using PRELIS; in Mplus all the work is done by the program once indicators have been declared as categorical. All three 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.

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.

6. Acknowledgements

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

7. References

Akaike, H. (1987), "Factor analysis and AIC," *Psychometrika*, 52, 317-332.

Arbuckle, J.L. (2006), *Amos 7.0 User's Guide*, Chicago, IL: SPSS Inc.

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.

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.

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.

Jöreskog, K.G. and Sörbom, D. (2004). *LISREL 8.7*. Scientific Software International, Inc.

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.