Covariance Structure Analysis (LISREL)
Lecture Notes

Professor Ross L. Matsueda

LECTURE 14: WLS AND MODELS FOR ORDINAL AND LIMITED VARIABLES

I. PROPERTIES AND ROBUSTNESS OF ESTIMATORS.
   A. PROPERTIES OF ESTIMATORS UNDER NONNORMAL AND ORDINAL VARIABLES.
   B. ROBUSTNESS OF MAXIMUM LIKELIHOOD: MONTE CARLO EVIDENCE.
II. NON-NORMAL DISTRIBUTIONS.
   A. TESTING FOR SKEWNESS AND KURTOSIS.
   B. ASYMPTOTIC DISTRIBUTION FREE ESTIMATOR (WLS).
III. CATEGORICAL AND CENSORED MEASURES.
   A. CORRELATIONS FOR CATEGORICAL AND LIMITED DEPENDENT VARIABLES.
   B. WEIGHTED LEAST SQUARES ESTIMATION.
IV. PRELIS AND LISREL APPROACH TO ESTIMATION.

I. PROPERTIES AND ROBUSTNESS OF ESTIMATORS.

A. PROPERTIES OF ESTIMATORS UNDER NONNORMAL AND ORDINAL VARIABLES.

The maximum likelihood estimator, asymptotic standard errors, and likelihood-ratio test statistic all assume observed variables are measured on interval scales and are distributed multivariate normal. Given this assumption, estimates of overidentified models are consistent and asymptotically efficient, and estimated standard errors and chi-square statistics are consistent. The following table (modified from Bollen 1989, p. 416) summarizes properties of estimators under different conditions:

Observed Variables	Consistent (θ)			Asyptotically Efficient (θ)			Consistent ACOV(θ)			Consistent χ^2		
	ML	ADF	WLS(P)	ML	ADF	WLS(P)	ML	ADF	WLS(P)	ML	ADF	WLS(P)
1. Continuous												
Multinormal	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
No Kurtosis	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Elliptical	yes	yes	yes	yes	yes	yes	no	yes	yes	no	yes	yes
Nonnormal	yes	yes	yes	no	yes	yes	no	yes	yes	no	yes	yes
Censored	no	no	yes	no	no	yes	no	no	yes	no	no	yes
2. Ordinal												
Nonnormal	no	no	yes	no	no	yes	no	no	yes	no	no	yes

These are general properties of estimators under the specific conditions; a given property may hold for an estimator for special models, but not generally. Note that kurtosis refers to the relative thinness or fatness of the tails of a distribution relative to the normal (positive kurtosis has thinner tails; negative kurtosis means fatter tails). An elliptical distribution is one in which observed variables have no skewness and the identical kurtosis. This table tells us about the asymptotic properties of estimators (properties of the limiting distribution of the estimator as the sample size approaches infinity), but says nothing about the finite-sample properties of estimators, which may be more important in empirical applications, where we're dealing with finite samples. Thus, we may want to know how robust an estimator is to departures from assumptions in finite samples.

B. ROBUSTNESS OF MAXIMUM LIKELIHOOD: MONTE CARLO EVIDENCE.

One way of examining how robust an estimator is to departures from assumptions is to use a Monte Carlo study. Given that we know of the asymptotic distribution of the ML estimator but not the finite-sample distribution, the Monte Carlo method *simulates* the sampling distribution of the estimator for a given sample size and model. So, begin with some multivariate distribution -- say a skewed or categorical distribution -- then randomly select a sample of size n and estimate the parameter of a model; then repeat this over and over for t trials, and plot the frequency distribution of estimates. This will be a sampling distribution simulated for a given model, a given multivariate distribution of observed variables, and a given sample size. By varying the form of the multivariate distribution (varying skewness or kurtosis) or the sample size, one can examine the robustness of an estimator to violations of assumptions (e.g., multinormality and large samples).

Most of the Monte Carlo studies of covariance structure models have focused on the effects of small sample size and nonnormality resulting from crude classification of continuous variables into ordinal variables. Bollen (1987) reviews this literature and draws four conclusions. First, Pearson product moment correlations are biased toward zero when variables are categorical; this bias worsens with fewer categories (less than 5). Second, parameter estimates are fairly robust to categorization (2-5 categories) and nonnormality, but standardized coefficients can be attenuated (worsening with fewer categories). Third, estimated standard errors and χ^2s are affected by skewness and kurtosis of categorized variables. Fourth, categorization of continuous indicators may cause correlated measurement errors. Boomsma also finds that for a variety of models, ML is robust to small sample size, $n \geq 200$. Of the two brief Monte Carlo studies of nonnormal continuous variables, it appears that ML estimates yield small biases, but that estimated standard errors and χ^2s are underestimated.

In sum, ML (and GLS) estimates of parameters in a covariance structure model appear to be somewhat robust to departures from normality, both with ordinal and continuous variables. Thus, the normality assumption appears much weaker than for other statistical models, such as sample selection models, in which estimates are very sensitive to departures from the assumption of bivariate normality between disturbances. But with respect to standard errors and χ^2s, ML (and GLS) appear to produce estimates that are biased downward. These findings, however, are not general, but are instead highly model-specific. Therefore, the safe strategy would be to investigate robustness for each specific research context (by checking the sample distribution of variables and using appropriate estimators when possible).

II. NONNORMAL DISTRIBUTIONS.

A. TESTING FOR SKEWNESS AND KURTOSIS.

Recall that if a variable is normally distributed, its distribution can be completely described by the first moment about the origin (the mean) and the second moment about the mean (the variance). This is because higher-order cumulants (moments) about the mean are zero for a normally-distributed variable. To describe the distributions of non-normal variables, we have to add higher-order moments about the mean, which take on the following general form:

$$\mu_r = E[(X - E(X)]^r = E(X - \mu_1)^r$$

Of course, we rarely have access to population moments, so we will have to rely on sample moments, which take the following form:

$$m_r = \sum_{i=1}^{n}(x_i - \bar{x})^r$$

To describe the distributions of non-normal variables, we can usually rely on the third and fourth moments about the mean. The third moment about the mean measures skewness:

$$\mu_3 = E(X - \mu_1)^3$$

The sample counterpart is:

$$m_3 = \sum_{i=1}^{n}(x_i - \bar{x})^3$$

We can standardize this by dividing μ_3 by $(\mu_2)^{3/2} = \sigma^3 = (\sigma_x^2)^{3/2}$, which makes the measure of skewness independent of the dispersion of x:

$$\gamma_1 = \mu_3/(\mu_2)^{3/2} = E(X - \mu_1)^3/E[(X - \mu_1)^2]^{3/2}$$

For normally-distributed variables, this measure is zero; if it is positive, the distribution is skewed to the left; if it is negative, the distribution is skewed to the right. The sample counterpart of this standardized measure of skewness is:

$$\sqrt{b_1} = m_3/(m_2)^{3/2} = \sum_{i=1}^{n}(x_i - \bar{x})^3/[\sum_{i=1}^{n}(x_i - \bar{x})^2]^{3/2}$$

where $m_3 = \sum_{i=1}^{n}(x_i - \bar{x})^3$, and $m_2 = s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, the sample variance.

The fourth moment about the mean measures kurtosis:

$$\mu_4 = E(X - \mu_1)^4$$

The sample counterpart to the fourth moment is:

$$m_4 = \sum_{i=1}^{n}(x_i - \bar{x})^4$$

We can standardize this by dividing μ_4 by $\mu_2^2 = (\sigma_x^2)^2$, the variance squared, which makes the measure of kurtosis independent of the dispersion of x:

$$\gamma_2 = \mu_4/\mu_2^2 = E(X - \mu_1)^4/E[(X - \mu_1)^2]^2$$

For normally-distributed variables this measure is equal to 3; if it is greater than 3, it has positive kurtosis, resulting in thinner tails than normal; if it is less than 3, it has negative kurtosis, resulting in fatter tails than normal. To center this variables around zero (for normally-distributed variables) we can subtract 3 (see Kendall and Stuart 1961, ch. 3):

$$\gamma_2 = (\mu_4/\mu_2^2) - 3 = [E(X - \mu_1)^4/[E(X - \mu_1)^2]^2 - 3$$

The sample counterpart of the standardized measure of kurtosis is:

$$b_2 - 3 = m_4/(m_2)^2 - 3 = \sum_{i=1}^{n}(x_i - \bar{x})^4/[\sum_{i=1}^{n}(x_i - \bar{x})^2]^2 - 3$$

where $m_2 = s^2 = \sum_{i=1}^{n}(x_i - x)^2$, the sample variance. By subtracting 3 from b_2, we obtain a sample measure of kurtosis that takes on a value of zero for no kurtosis ("mesokurtic"), a positive value for positive kurtosis ("leptokurtic"), and a negative value for negative kurtosis ("platykurtic"). PRELIS and EQS print these measures of skewness $\sqrt{b_1}$ and kurtosis $b_2 - 3$.

Mardia has proposed multivariate tests of skewness and kurtosis. For skewness, he proposes:

$$b_{1,p} = (1/N^2) \sum_{i=1}^{n} \sum_{j=1}^{n} [\mathbf{z}_i - \bar{\mathbf{z}})' \mathbf{S}^{-1} (\mathbf{z}_j - \bar{\mathbf{z}})]^3$$

and for kurtosis:

$$b_{2,p} = (1/N) \sum_{i=1}^{n} \sum_{j=1}^{n} [\mathbf{z}_i - \bar{\mathbf{z}})' \mathbf{S}^{-1} (\mathbf{z}_j - \bar{\mathbf{z}})]^2$$

where **z** is a vector of x and y variables with mean $\bar{z}$ and covariance matrix **S**. When normalized, each of these is asymptotically distributed as a normal variate with zero mean and unit variance. Thus, in large samples, under the null hypothesis of no skewness or kurtosis, these are distributed as a Z-statistic. To improve the approximation of a normal variate in moderate samples, Mardia has provided corrections (see Bollen 1989, p. 424). The corrected measure of kurtosis is given in PRELIS and EQS. Thus, if the Z-statistic exceeds 1.96 one would reject the null hypothesis of no kurtosis at the .05 level of significance.

B. ASYMPTOTIC DISTRIBUTION FREE ESTIMATOR (WLS).

Michael Browne has developed a very general estimation procedure, often referred to as "weighted least squares," (WLS) which turns out to be enormously useful for covariance structure analysis. It provides optimal estimates for models with non-normal observable variables or ordinal or truncated variables. Moreover, all the estimators we have discussed so far can be viewed as special cases of WLS. Let's first apply it to the case in which observed variables are not distributed multivariate normal. Recall that the problem of efficient estimation arose in overidentified models, when there is more than one way of computing parameters in terms of moments. For the multinormal case, we could ignore higher-order moments, since they are zero, and use the maximum likelihood procedure to give an optimal weighting of conflicting sample estimates. And maximum likelihood accomplished this by choosing values of θ which minimized the discrepancy between S and $\Sigma(\theta)$. Now, without the multinormal assumption, an adequate description of the distribution requires higher-order sample population moments. With regard to estimation, we now have additional information about the sample distribution, and need to incorporate that information (higher-order sample moments) to obtain an efficient estimator. This is precisely what weighted least squares does: it weights the discrepancy matrix S - $\Sigma(\theta)$ by a weight matrix, which in turn is a function of the fourth-order moments about the mean, and the second-order moments about the mean (covariances).

For the general case, Browne presents the following fitting function for weighted least squares:

$$F_{WLS} = [\mathbf{s} - \sigma(\theta)]' \mathbf{W}^{-1} [\mathbf{s} - \sigma(\theta)]$$

The fitting function can be expressed in scalar form:

$$F_{WLS} = \Sigma_{g=1}^{k} \Sigma_{h=1}^{g} \Sigma_{i=1}^{k} \Sigma_{j=1}^{i} w^{gh,ij} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij})$$

where **s** is a u = ½(p + q)(p + q) vector of nonredundant sample covariances; $\sigma(\theta)$ is a u = ½(p + q)(p + q) vector of nonredundant population covariances implied by the model; and $\mathbf{W}^{-1}$ is a u x u = ½(p + q)(p + q) x ½(p + q)(p + q) positive definite weight matrix, with typical element $w^{gh,ij}$. In other words, we have simply taken the (p + q) x (p + q) matrices **S** and $\Sigma(\theta)$ and placed the elements into a two vectors. Thus, **s** contains the lower triangular elements of **S**, $\mathbf{s}' = [s_{11}, s_{21}, s_{22}, s_{31},..., s_{kk}]$. For a 3 x 3 **S**, $\mathbf{s} = s_{11}, s_{21}, s_{22}, s_{31}, s_{32}, s_{33}$. Two typical elements of **s** can be designated s_{ij} and s_{gh}. Similarly, $\sigma(\theta)$ contains the corresponding lower triangular elements of $\Sigma(\theta)$, $\sigma(\theta)' = [\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31},..., \sigma_{kk}]$. For a 3 x 3 $\Sigma(\theta)$, $\sigma(\theta) = \sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, \sigma_{32}, \sigma_{33}$. Two typical elements of $\sigma(\theta)$ can be designated σ_{ij} and σ_{gh}.

While **W** can be any positive definite weight matrix, Browne (1984) has shown that there is an optimal weight matrix. If **W** is taken to be the (population) asymptotic covariance matrix of **s**, Σ_{ss}, then the resulting estimator is asymptotically efficient within the class of estimators that can be represened as WLS. But what is Σ_{ss}? This is the asymptotic covariance matrix of the sample moments. Our sample moment matrix contains sample estimates of population variances and covariances. Each element is a sample estimate which has a variance (the square root of which is the standard error). The elements also have covariances (since the sample estimates of moments are not in general independent). Σ_{ss} contains these asymptotic variances and covariances of the sample moments in a single matrix. The ½u(u + 1) elements of Σ_{ss} each represents the asymptotic covariance of s_{gh} with s_{ij}:

$$ACOV(s_{gh}, s_{ij}) = 1/N\ (\sigma_{ghij} - \sigma_{gh}\, \sigma_{ij})$$

Thus, for example, if we have a three-variable model, $\mathbf{s}'$ (1 x 6) = $[s_{11}, s_{21}, s_{22}, s_{31}, s_{32}, s_{33}]$, Σ_{ss} is 6 x 6 and has (6 x 7)/2 = 21 elements:

Σ_{ss} (6 x 6) =

$$\begin{bmatrix} ACOV(s_{11}, s_{11}) & ACOV(s_{11}, s_{21}) & ACOV(s_{11}, s_{22}) & \cdots & ACOV(s_{11}, s_{33}) \\ ACOV(s_{21}, s_{11}) & ACOV(s_{21}, s_{21}) & ACOV(s_{21}, s_{22}) & \cdots & ACOV(s_{21}, s_{33}) \\ \vdots & \vdots & \vdots & & \vdots \\ ACOV(s_{33}, s_{11}) & ACOV(s_{33}, s_{21}) & ACOV(s_{33}, s_{22}) & \cdots & ACOV(s_{33}, s_{33}) \end{bmatrix}$$

Thus, replacing **W** with Σ_{ss} gives us the following fitting function:

$$F_{WLS} = [s - \sigma(\theta)]' \Sigma_{ss}^{-1} [s - \sigma(\theta)]$$

In scalar form, this becomes

$$F_{WLS} = \Sigma_{g=1}^{k} \Sigma_{h=1}^{g} \Sigma_{i=1}^{k} \Sigma_{j=1}^{i} [ACOV(s_{gh}, s_{ij})]^{-1} (s_{gh} - \sigma_{gh})(s_{ij} - \sigma_{ij})$$

To obtain WLS estimates, we would differentiate F_{WLS} with respect to each free parameter, set the resulting derivatives equal to zero, and then solve for parameters. Browne has shown that for any multivariate distribution in continuous variables, minimizing this function produces consistent and asymptotically efficient estimates under mild assumptions (that the eighth-order moments of observed variables are finite). Thus, this estimator yields optimal estimates even when observed variables are not normally distributed. Therefore, Browne (1982) introduced this as an "asymptotically distribution-free best generalized least squares estimator." It is sometimes called an "arbitrary distribution estimator." Jöreskog and Sörbom (1988) refer to it as "weighted least squares" to distinguish it from normal theory "generalized least squares."

Furthermore, this function gives us an intuitively appealing method of estimation. Recall that the issue of estimation in covariance structure models arose in overidentified models, in which there was more than one way to estimate parameters from sample moments, and we wanted to choose parameter values that somehow minimized (a function of) the discrepancy between **S** and $\Sigma(\theta)$. WLS does exactly that, minimizing a quadratic form, consisting of the residual matrix (in vector form) **s** - $\sigma(\theta)$ weighted by the inverse of the asymptotic covariance matrix of the sample moments **s**. Thus, among the multiple ways of estimating parameters from sample moments, those with high sampling variability (estimates based on sample moments with large variances and covariances) receive less weight, and those with low sampling variability (estimates based on sample moments with small variances and covariances) receive more weight.

> Aside: The *principle* of WLS (weighting a minimized sum of squared discrepancies) here is analogous to weighted least squares regression to correct for heteroskedasticity. There, the assumption that error variances are constant across observations is violated. WLS yields an efficient estimator by weighting *observations* inverse to their disturbance variance: those that have large variances (and hence greater sampling variability) receive less weight; those with small variances receive more weight. The analogy *only holds as a principle*, since here we are minimizing the discrepancies among competing sample moment estimators, whereas, in the regression context one is minimizing weighted discrepancies between predicted and observed values of the dependent variable.

Furthermore, WLS produces optimal estimates for models with nonnormally-distributed observed variables by taking the departure from normality into consideration in the weight matrix. Above we noted that the elements of the optimal weight matrix, $ACOV(s_{gh}, s_{ij}) = 1/N\ (\sigma_{ghij} - \sigma_{gh}\,\sigma_{ij})$, subtracts off the product of two covariances from the fourth-order moment; thus the weights, reflecting sampling variability, contain the fourth-order moment, which measures kurtosis in observed variables. When the observed variables are multinormally distributed, the fourth-order moment is not needed and WLS reduces to estimators based on normal theory, such as maximum likelihood or normal-theory GLS.

What about test statistics and standard errors? Browne shows that (N - 1) F_{WLS} evaluated at θ is distributed as a χ^2 variate with t degrees of freedom, where $F_{WLS} = [s - \sigma(\theta)]' \Sigma_{ss}^{-1} [s - \sigma(\theta)]$. This is directly analogous to the likelihood ratio test statistic, and is used in the same way. Furthermore the asymptotic covariance matrix of θ_{WLS} is

$$ACOV(\theta_{WLS}) = 1/N \{[\partial\sigma(\theta)/\partial\theta] \Sigma_{ss}^{-1} [\partial\sigma(\theta)/\partial\theta]'\}^{-1}$$

The elements of the quadratic form take the partial derivative of an element of the covariance matrix implied by the model with respect to each parameter and weights it by the inverse of the asymptotic covariance matrix of the sample moment matrix. This matrix can be used to construct z-statistics and perform the usual tests of hypotheses.

In general, we will not have access to the population asymptotic covariance matrix of our sample moments Σ_{ss}. Instead, we will have to rely on an estimator of this matrix. Browne shows if Σ_{ss} is replaced by any consistent estimator of Σ_{ss}, say S_{ss}, the WLS estimator is still consistent and asymptotically efficient. The fitting function and estimated asymptotic covariance matrix become:

$$F_{WLS} = [s - \sigma(\theta)]' S_{ss}^{-1} [s - \sigma(\theta)]$$

$$ACOV(\theta_{WLS}) = 1/N \{[\partial\sigma(\theta)/\partial\theta] S_{ss}^{-1} [\partial\sigma(\theta)/\partial\theta]'\}^{-1}$$

where S_{ss} is a positive definite matrix of sample estimates of asymptotic covariances of the sample moment matrix S. A typical element of S_{ss} is the estimated asymptotic covariance of s_{gh} with s_{ij}:

$$AC\hat{O}V(s_{gh}, s_{ij}) = 1/N\ (s_{ghij} - s_{gh}\, s_{ij})$$

where $s_{ghij} = (1/N) \Sigma_{t=1}^{N} (z_{gt} - \bar{z}_g)(z_{ht} - \bar{z}_h)(z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)$

$s_{gh} = (1/N) \Sigma_{t=1}^{N} (z_{gt} - \bar{z}_g)(z_{ht} - \bar{z}_h)$

$s_{ij} = (1/N) \Sigma_{t=1}^{N} (z_{it} - \bar{z}_i)(z_{jt} - \bar{z}_j)$

and z_{gt} corresponds to the observed value of the t-th observation on the g-th X or Y variable.

Thus, s_{gh} is simply the sample covariance between z_g and z_h; s_{ij} is the sample covariance between z_i and z_j, and s_{ghij} is the fourth-order moment involving z_g, z_h, z_i, and z_j. Note that if g = h = i = j, then $s_{gh} = s_{ij}$ = sample variance and $s_{ghij} = (1/N) \Sigma_{t=1}^{N} (z_{gt} - \bar{z}_g)^4$ is the univariate sample fourth moment. Note also that these are consistent but biased estimates of covariances and fourth-order central moments, since they use N in the denominator rather than N - 1.

To estimate a covariance structure model using WLS in LISREL 8, you must begin with PRELIS to estimate sample moments and the asymptotic covariance matrix of sample moments, and save the matrices to disk. Then use LISREL 8, reading in both the sample moments and the asymptotic covariance matrix and asking for WLS as the estimation procedure. Then all estimates and test statistics are correct. EQS will provide the same estimator by reading in raw data (so EQS can compute the asymptotic covariance matrix), and select ME=AGLS, for method of estimation equals "arbitrary generalized least squares."

While in principle, WLS is perhaps the most desireable estimator, since it doesn't require strong assumptions about the distribution of observed variables, two very serious restrictions limit its practical use:

1. The biggest restriction is that it requires much larger samples for large models. For small models (with ten or fewer observable variables, it is probably fine. But for more than 15 observables, the sample size must be very large to obtain stable estimates of Σ_{ss}. The PRELIS program requires that $N \geq 1.5k(k + 1)$ where k is the number of observable variables (for 30 variables $N \geq 1395$). While this is arbitrary, and can be overridden by specifying

a user minimum sample size by specifying MS = N on the OU line, typically, such a strategy will yield unstable (high sampling variability) estimates.

2. A second restriction is that WLS requires large amounts of storage space on disk and computer memory for large models. This is because it must save the u x u matrix $\mathbf{S}_{ss}$, which has ½u(u + 1) elements. Recall that u = ½(p + q)(p + q). Thus, for 30 variables, $\mathbf{S}_{ss}$ has 108,345 elements; for 55 variables $\mathbf{S}_{ss}$ has 1,186,570 elements. Moreover, the program has to invert $\mathbf{S}_{ss}$, which can require a large amount of memory, which can be very expensive.

One compromise available in the LISREL progam is to use only the diagonal elements of $\mathbf{S}_{ss}$, the asymptotic variances of the sample moment matrix in estimation. Here the amount of storage and memory required is reduced dramatically. The fitting function of the estimator, called diagonally weighted least squares, simplifies to:

$$F_{DWLS} = \Sigma_{g=1}^{k} \Sigma_{h=1}^{g} (1/w_{gh})(s_{gh} - \sigma_{gh})^2$$

The diagonally-weighted least squares estimator is consistent but asymptotically inefficient. It is a compromise between unweighted least squares and WLS.

A nice property of Browne's general weighted least squares approach to estimation in covariance structure models is that all other estimators become special cases. Under multinormality, $\sigma_{ghij} = \sigma_{gh}\,\sigma_{ij} + \sigma_{gi}\,\sigma_{ih} + \sigma_{gj}\,\sigma_{hi}$, and ACOV($s_{gh}$, s_{ij}) = 1/N ($\sigma_{gi}\,\sigma_{ih} + \sigma_{gj}\,\sigma_{hi}$), and

$$F_{WLS} = \tfrac{1}{2}\mathrm{tr}\{[\mathbf{S} - \Sigma(\theta)]\mathbf{W}^{-1}\}^2$$

where W is now a k x k weight matrix. Brown shows that, under multinormality:

1. When $\mathbf{W} = \Sigma(\theta)$, F_{ML} is the maximum likelihood fitting function, yielding ML estimates that are consistent and asymptotically efficient.

2. When $\mathbf{W} = \mathbf{S}$, F_{GLS} is the normal theory generalized least squares fitting function, yielding GLS estimates that are consistent and asymptotically efficient.

3. When $\mathbf{W} = \mathbf{I}$, F_{ULS} is the normal theory unweighted least squares fitting function, yielding ULS estimates that are consistent but asymptotically inefficient.

LISREL 8 provides each of these estimators, which can be specified in the ME parameter on the OU line. EQS also provides each of these estimators, plus an elliptical theory estimator, which gives asymptotically efficient and consistent estimators for models with observed variables that have no skewness and the identical kurtosis (see Bentler 1989).

The asymptotic distribution free WLS estimator provide optimal estimates for models containing continuous variables with arbitrary distributions. Given this, how can we estimate models with limited, discrete, or ordinal variables? The trick is to (1) assume that each observed limited dependent variable reflects a normally-distributed continuous latent variable; (2) estimate polychoric and polyserial correlations for ordinal variables and product-moment correlations on normal scores for limited dependent variables; and (3) use WLS estimation to get optimal estimates and correct estimated standard errors and χ^2s.

III. CATEGORICAL AND CENSORED MEASURES.

When observed variables, ys and xs, are not continuous or inverval-scale measures, the covariance structure model is no longer formally appropriate. Since the observed ys and xs are not measured on interval scales, their metrics are incorrect, and all parameter estimates involving them will be biased and inconsistent, and therefore, the basic hypothesis of covariance structure analysis, based on continuous variables, $\Sigma = \Sigma(\theta)$ does not hold. We can see this by assuming that each observed y_i is a noncontinuous measure of a continuous latent variable y*, and similarly for x_i (where *i* indexes the variable number):

$$y_i = f(y_i^*)$$

$$x_i = f(x_i^*)$$

Assume further that the covariance structure model holds for y_i^* and x_i^*: $\Sigma^* = \Sigma(\theta)$ If y and x were linear functions of y* and x*, respectively, we could use conventional methods (only the metrics would be affected); however in general these functions will be nonlinear. Our sample estimates of observed moments S -- which assume continuous normally-distributed variables -- involving y_i and x_i will also be biased, as will our estimates of the asymptotic covariance matrix of S, Σ_{ss}. This implies that applying WLS to S will not solve the problem. Thus, the problems are:

1. $\Sigma \neq \Sigma(\theta)$, therefore, θ is an inconsistent estimator of θ: $\underset{n\to\infty}{\text{plim}}\ E(\theta_{ML}) \neq \theta$
2. $E(S) = \Sigma \neq \Sigma^*$. The estimates of sample moments are inconsistent.
3. $E(S_{xx}) \neq \Sigma^*_{ss}$. The estimates of the asymptotic covariance matrix of S will be inconsistent.
4. Therefore, test statistics and asymptotic standard errors will be biased.

Since our covariance structure hypothesis holds for y_i^* and x_i^*, a straightforward solution to this problem will be to estimate y_i^* and x_i^* from the observed data, y_i and x_i, and then proceed as usual. That is precisely what LISREL 8 (when used with PRELIS) and LISCOMP (Bengt Muthén's program) do. Most of this work was pioneered by Bengt Muthén, who also developed his LISCOMP program. Recently, Jöreskog and Sörbom have implemented much of this into their PRELIS and LISREL 8 programs.

CORRELATIONS AMONG ORDINAL, DICHOTOMOUS, CENSORED, AND CONTINUOUS VARIABLES

To incorporate ordinal, dichotomous, and censored observed (dependent) variables into a covariance structure model, we first must model the relationship between the observed variable and continuous latent variable, and then compute the proper sample moment (correlation coefficient), asymptotic covariance matrix of sample moments, and then estimate the covariance structure model using WLS. The following is a summary of correlations for ordinal, dichotomous, censored, and continuous variables:

	Continuous	Ordinal	Dichotomous	Censored
Continuous	Pearson			
Ordinal	Polyserial	Polychoric		
Dichotomous	Biserial	Polychoric	Tetrachoric	--
Censored	Tobit	--	--	Tobit

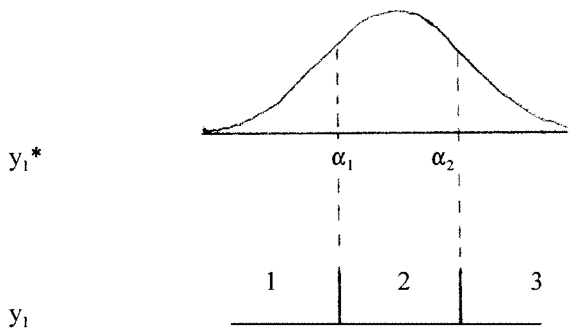
To handle dichotomous and ordinal variables, assume that y_1 is a three-category ordinal measure of a continuous latent variable y_1^*. We can think of the relationship between y_1 and y_1^* as a threshold model: when the latent continuous variable y_1^* passes a certain threshold α_i, y_1 moves from one category to another. Since y_1 has three categories, we assume two thresholds:

$y_1 = 1 \quad \text{if } y_1^* \leq \alpha_1$

$y_1 = 2 \quad \text{if } \alpha_1 < y_1^* \leq \alpha_2$

$y_1 = 3 \quad \text{if } \alpha_2 < y_1^*$

We can diagram this relationship:



In general for a variable y_1 with c categories, we define the thresholds as:

$y_1 = 1 \quad \text{if } y_1^* \leq \alpha_1$

$y_1 = 2 \quad \text{if } \alpha_1 < y_1^* \leq \alpha_2$

$y_1 = c - 1 \quad \text{if } \alpha_{c-1} < y_1^* \leq \alpha_c$

$y_1 = c \quad \text{if } \alpha_{c-1} < y_1^*$

To get to y_1^* from y_1 we simply need to specify a distribution for y_1^* and then estimate the thresholds. If we assume that y_1^* is normally distributed, then we can estimate the tresholds as follows:

$$\alpha_i = \Phi^{-1} \left(\textstyle\sum_{k=1}^{i} n_k/N\right),$$

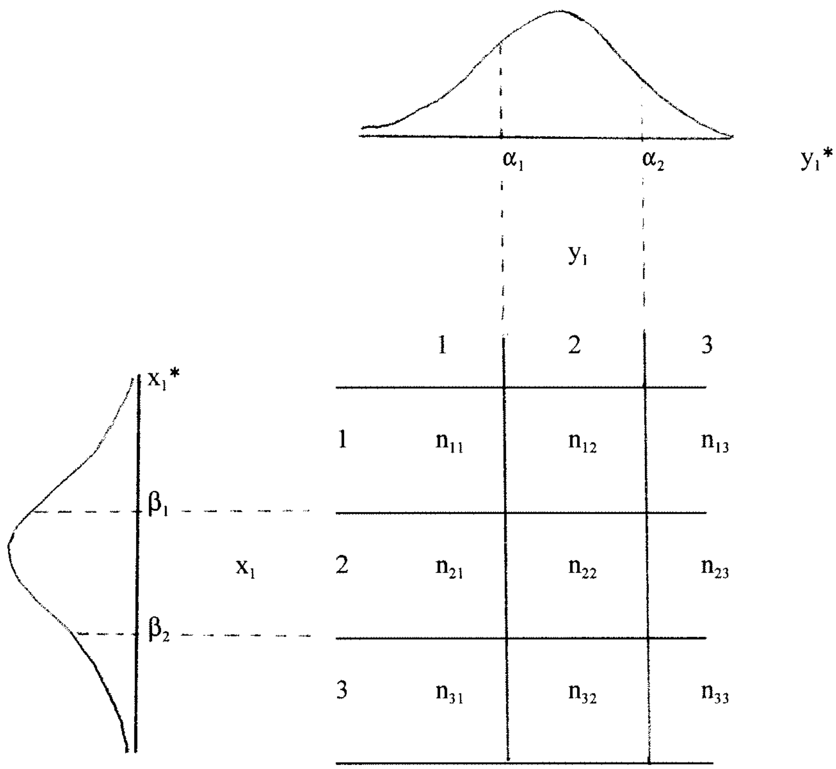
where $i = 1, 2, \ldots, c - 1$, c is the number of categories, Φ^{-1} is the inverse of the standard normal distribution function (cumulative distribution function), and n_k is the sample size of the kth category, $N = n_1 + n_2 + \ldots + n_{c-1}$.

For our example with three categories (j = 3), we would have:

$$\alpha_1 = \Phi^{-1} (n_1/N)$$

$$\alpha_2 = \Phi^{-1} (n_1/N + n_2/N)$$

Now, to estimate the correlation between two ordinal variables, assume both x_1 and y_1 are ordinal measures of continuous latent variables, which have a bivariate normal distribution. Let's assume that x_1 is an ordinal variable with d - 1 thresholds β_j. We can arrange the raw data in a cross-tabulation between x_1 and y_1. In the general case, assume that there are r categories of y_1 and s categories of x_1, and n_{ij} represents the observed frequences for i = 1, 2, ..., r categories of y_1 and j = 1, 2, ..., s categories of x_1. When r = 3 and s = 3, we have a 3 x 3 data matrix:



Furthermore, let π_{ij} be the population parameter denoting the probability that an observation falls into cell *(i, j)*. Then we can define the likelihood function of the sample as:

$$\mathcal{L}(n_{ij}\,|\pi_{ij}) = C\ \Pi_i^r\ \Pi_j^s\ \pi_{ij}^{n_{ij}} \qquad \text{where C is a constant}$$

Then, logging both sides yields:

$$\log \mathcal{L}(n_{ij}\,|\pi_{ij}) = C \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij} \log \pi_{ij}$$

It follows that:

$$\pi_{ij} = \Phi_2\,(\alpha_i, \beta_j) - \Phi_2\,(\alpha_{i-1}, \beta_j) - \Phi_2\,(\alpha_i, \beta_{j-1}) - \Phi_2\,(\alpha_{i-1}, \beta_{j-1})$$

where Φ_2 is the bivariate normal distribution function with population correlation ρ. One can obtain maximum likelihood estimates by differentiating $\log\mathcal{L}$ with respect to ρ, $\alpha_1 \ldots \alpha_{r-1}$, and $\beta_1 \ldots \beta_{s-1}$, set the equations equal to zero and solving for parameters. For example, Olsson (1979) shows that the first derivative with respec to ρ is:

$$\partial \log \mathcal{L}/\partial\rho = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}/\pi_{ij}\ \partial\pi_{ij}/\partial\rho$$

And, he shows that the last term is:

$$\partial\pi_{ij}/\partial\rho = \phi_2\,(\alpha_i, \beta_j) - \phi_2\,(\alpha_{i-1}, \beta_j) - \phi_2\,(\alpha_i, \beta_{j-1}) - \phi_2\,(\alpha_{i-1}, \beta_{j-1})$$

which simplifies the partial derivative, which we can set to zero:

$$\partial \log \mathcal{L}/\partial\rho = \sum_{i=1}^{r} \sum_{j=1}^{s} n_{ij}/\pi_{ij}\ [\phi_2\,(\alpha_i, \beta_j) - \phi_2\,(\alpha_{i-1}, \beta_j) - \phi_2\,(\alpha_i, \beta_{j-1}) - \phi_2\,(\alpha_{i-1}, \beta_{j-1})] = 0$$

Thus, solving this equation simultaneously with the equations for the thresholds yields maximum likelhood estimation of ρ. Alternatively, one can estimate the thresholds from marginal proportions of the table, and then solve for the polychoric correlation ρ:

$$\alpha_i = \Phi^{-1} (\textstyle\sum_{k=1}^{i} n_k/N), \qquad i = 1, 2, \ldots, c - 1,$$

$$\beta_j = \Phi^{-1} (\textstyle\sum_{l=1}^{j} n_l/N), \qquad j = 1, 2, \ldots, d - 1,$$

where c is the number of categories of y_1, d is the number of categories of x_1, Φ^{-1} is the inverse of the standard normal distribution function (cumulative distribution function), n_k is the sample size of the kth category of y_1, $N = n_1 + n_2 + \ldots + n_{c-1}$, and n_l is the sample size of the lth category of x_1, $N = n_1 + n_2 + \ldots + n_{d-1}$. We can substitute these estimates of the thresholds above and solve for ρ. Finally, the covariance matrix can be obtained by taking the expected value of the negative of the inverted matrix of second-order partial derivatives of $\log\mathcal{L}$ with respect to the parameters. Moreover, these results can be generalized to estimate a matrix of polychoric correlations by replacing y_1 with a vector Y where $Y' = [y_1' \, y_2' \ldots y_p']$ and replacing x_1 with a vector X where $X' = [x_1' \, x_2' \ldots x_p']$. Then the above strategy can be used to obtain ML estimates of the $\frac{1}{2}(p + q) \times (p + q)$ ρs -- call this matrix **R**. If the matrix of ρs are placed in a vector called $\hat{\rho}$, the covariance matrix of the ML estimates of polychoric correlations, $\Sigma\hat{\rho}\hat{\rho}$, which will be of order $\frac{1}{2}(p + q)(p + q) \times \frac{1}{2}(p + q)(p + q)$. As usual, in empirical applications we will not know the population asymptotic covariance matrix, $\Sigma\hat{\rho}\hat{\rho}$, so we will have to estimate it from sample data, using $S\hat{\rho}\hat{\rho}$. See Poon and Lee (1987) for ML estimation of multivariate polychoric correlations and a Fletcher-Powell optimization algorithm.

To estimate a polyserial correlation, a similar procedure is followed. Assume that x_1 is a continuous observed variable and y_1^* is a continuous latent variable, and the two are distributed bivariate normal with parameters $\mu_{x1} = \mu$, $\sigma_{x1}^2 = \sigma^2$, $\mu_{y1*} = 0$, $\sigma_{y1*} = 1$, and $\rho_{x1y1*} = \rho$. Furthermore let y_1 is an ordered categorical indicator (with c categories) of y_1^*. Then relationship between y_1 and y_1^* entails thresholds, τ_j:

$$y_1 = 1 \qquad \text{if } y_1^* \le \tau_1$$

$$y_1 = 2 \qquad \text{if } \tau_1 < y_1^* \le \tau_2$$

$$y_1 = c - 1 \qquad \text{if } \tau_{c-1} < y_1^* \le \tau_c$$

$$y_1 = c \qquad \text{if } \tau_{c-1} < y_1^*$$

Olsson et al. (1982) show that the distribution of y_1 is

$$P(y_1 = y_{1j}) = p_j = \Phi(\tau_j) - \Phi(\tau_{j-1})$$

where $\Phi(\tau) = 1/(2\pi)^{1/2} \int_{-\infty}^{\tau} \exp[-\tau^2/2]$. The polyserial correlation between a continuous x and y_1^* is:

$$\rho = \textstyle\sum_{j=1}^{c} \mu_j \, [\phi(\tau_{j-1}) - \phi(\tau_j)] / \sum_{j=1}^{c} \mu_j \, [\phi(\tau_{j-1}) - \phi(\tau_j)]^2/p_j$$

where $\phi(\tau) = 1/(2\pi)^{1/2} \exp[-\tau^2/2]$, μ_j is the expected value of x given $y_1 = y_j$. They obtain the log likelihood function for the polyserial correlation by defining x as a normally-distributed variable with probability density function

$$p(x_i) = 1/(2\pi\sigma) \exp\{[-\tfrac{1}{2}(x_i - \mu)/\sigma]^2$$

and Z as a standardized x, $z = (x - \mu)/\sigma$, and noting that the conditional distribution of y_1^* given x is normal with mean ρz and variance $1 - \rho^2$:

$$p(y_1^* | x) \sim N[\rho z, (1 - \rho^2)]$$

Then the conditional distribution of y_1 given x is:

$$p(y_1 = j|x) = \Phi(\tau_j^*) - \Phi(\tau_{j-1}^*)$$

where $\tau_j^* = (\tau_j - \rho z)/(1 - \rho^2)^{1/2}$. The log-likelihood function, then is:

$$\log \mathcal{L} = \sum_{i=1}^{N} [\log p(x_i) + \log p(y_{1i}|x_i)]$$

ML estimators are obtained by maximizing this function -- differentiating log $\mathcal{L}$ with respect to parameters ρ, μ, σ_2, and τ_j, setting the equations to zero, and solving for parameters. The asymptotic covariance matrix of parameters can be obtained by taking the expected value of minus one times the inverted matrix of second-order partial derivatives. As with the polychoric correlations, this can be extended to the multivariate case, by specifying vectors of ys and xs. See Poon and Lee (1987) for ML estimation of multivariate polyserial and polychoric correlations.

For observed variables censored from above or from below, the same logic is applied: specify a threshold model relating censored variables to underlying continuous normally-distributed variables, and then estimate correlations among latent variables. Censored variables have unusually high concentrations of cases at the upper or lower ends of the distribution. This is commonly caused by ceiling or floor effects on what otherwise would be continuous variables. For example, attitudinal measures using Likert scales could result in floor or ceiling effects if "strongly agree" or "strongly disagree" didn't distinguish respondents with very extreme attitudes. Or one could argue that self-reported criminal behavior is censored from below since most people report no crimes, and crime cannot take on negative values. We can specify the relationship between a censored observable variable and the underlying latent continuous variable. If we let y_1 be a variable censored from above (say, through a ceiling effect), y_1^* be the latent continuous variable, and a_1 be a threshold:

$$y_1 = y_1^* \qquad \text{if } y_1^* < a_1$$

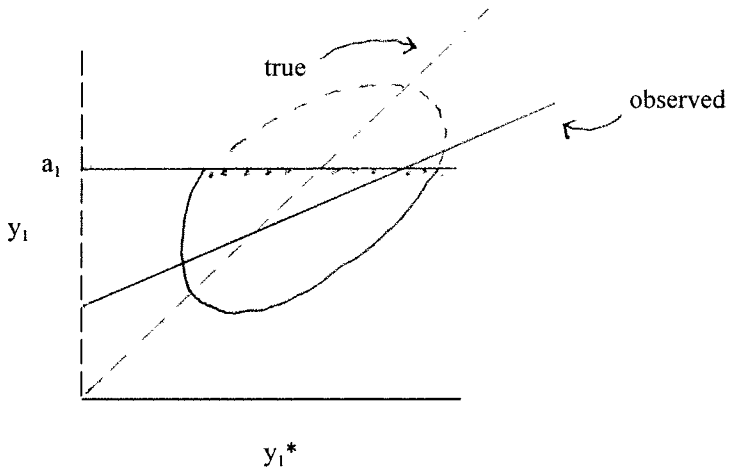
$$y_1 = a_1 \qquad \text{if } y_1^* \geq a_1$$

When y_1^* is less than the threshold, $y_1 = y_1^*$. But when y_1^* is greater than the threshold, observations on y_1 are censored, taking on the value of the upper bound (threshold). Clearly this will cause clumping of observations in the upper end of the distribution. A variable can be censored from below as well. Let y_1 be censored from below, y_1^* be the latent continuous variable, and b_1 be the the threshold:

$$y_1 = y_1^* \qquad \text{if } y_1^* > b_1$$

$$y_1 = b_1 \qquad \text{if } y_1^* \geq b_1$$

In both cases, censoring causes the observations to clump at the ends of the distributions. This will cause regressions on the censored variable to be biased toward zero. We can see this by examining a scatterplot of the observed variable y_1 censored from above with the continuous latent variable y_1^*:



If y_1 were not censored, the joint distribution would be bivariate normal, following the dotted lines. But censoring causes observations to clump at the threshold value. This causes the regression line to flatten out.
(Aside: Note that if the limit observations didn't take the on the threshold value, but instead could not be observed, they would be described as truncated observations, and the variable a truncated variable.)

To deal with censored variables, PRELIS uses a normal score transformation for censored observations. For a variable censored from above it uses the normal score $\hat{z}_a$ associated with the interval $y_1* \geq a_1$:

$$\hat{z}_a = \hat{\mu} + \hat{\sigma}\{\phi[(a_1 - \hat{\mu})/\hat{\sigma}]/\Phi[a_1 - \hat{\mu})/\hat{\sigma}]\}$$

where $\hat{\mu}$ and $\hat{\sigma}$ are ML estimates of μ and σ, respectively, ϕ is the standard normal density function, and Φ is the standard normal distribution function (cumulative distribution function). PRELIS does the same for variables censored from below, using the normal score $\hat{z}_b$ associated with the interval $y_1* \leq b_1$:

$$\hat{z}_b = \hat{\mu} + \hat{\sigma}\{\phi[(b_1 - \hat{\mu})/\hat{\sigma}]/\Phi[b_1 - \hat{\mu})/\hat{\sigma}]\}$$

Once the censored variables have been transformed to normal scores, PRELIS treats the variables as continuous.

B. WEIGHTED LEAST SQUARES ESTIMATION.

Once the matrix of maximum likelihood estimates of the polychoric, polyserial, tobit, and product-moment (Pearson) correlations are computed along with the asymptotic covariance matrix of this correlation matrix, then the usual covariance structure model can be estimated using weighted least squares. We could apply our normal theory estimators, ML, GLS, or ULS, and obtain consistent estimates, but the estimates would be asymptotically inefficient and the estimated asymptotic covariance matrix of the estimates would be biased; thus standard errors would be biased and test statistics would be incorrect. WLS will give consistent and asymptotically efficient estimates with unbiased estimates of the asymptotic covariance matrix of the estimates (and thus standard errors) and an unbiased estimate of the goodness-of-fit χ^2.

If we let $\hat{\rho}$ be a vector of our maximum likelihood estimates of various correlations, and $S\hat{\rho}\hat{\rho}$ be the estimated asymptotic covariance matrix of these correlation estimates, then

$$F_{WLS} = [\hat{\rho} - \sigma(\theta)]' \, S\hat{\rho}\hat{\rho}^{-1} \, [\hat{\rho} - \sigma(\theta)]$$

is our fitting function for WLS. The minimum value of this function yields a test statistic of the overall goodness of fit of the model, which has an asymptotic χ^2 distribution:

$(N-1)F_{WLS} \sim \chi^2 \quad df = t$

when F_{WLS} is evaluated at θ_{WLS}. The asymptotic covariance matrix of WLS estimates (note the difference with the asymptotic covariance matrix of correlations!) is:

$ACOV(\theta_{WLS}) = 1/N \ \{[\partial\sigma(\theta)/\partial\theta] \ S_{\hat{\rho}\hat{\rho}}^{-1} \ [\partial\sigma(\theta)/\partial\theta]'\}^{-1}$

IV. PRELIS AND LISREL APPROACH TO ESTIMATION.

Jöreskog and Sörbom (1986, 1989) have incorporated a WLS estimator for covariance structure models of continuous variables with nonnormal distributions and for models of ordinal, dichotomous, or censored variables. Their approach uses PRELIS to estimate the observed sample moment matrix and the asymptotic covariance matrix of sample moments, and LISREL 8 to obtain WLS estimates of parameters of the covariance structure model.

The first step is to use PRELIS to estimate correct sample moments and asymptotic covariances. To use PRELIS, first write out a raw data file with missing values coded as an out-of-range number, such as -9. Then read in the raw data file and specify the parameters of the run. PRELIS offers six types of correlations, corresponding to different kinds of observed variables. For continuous variables, it estimates Pearson product-moment correlations. For pairs of ordinal variables, it offers three correlations: Pearson correlations on ordinal variables replaced by normal scores, Pearson correlations on ordinal variables replaced by optimal scores (cannonical correlations), and polychoric correlations estimated by maximum likelihood. For correlations between ordinal and continuous variables, PRELIS provides polyserial correlations estimated either by maximum likelihood or an ad hoc consistent estimator. For censored variables, it provides correlations on censored variables replaced by normal scores. To define the scales of observed variables, use the following lines:

OR	*varlist*	for ordinal variables
CO	*varlist*	for continuous variables
CA	*varlist*	for variables censored from above
CB	*varlist*	for variables censored from below
CE	*varlist*	for variables censored from above and from below

Then to specify the form of correlation, use the following parameter on the OU line:

MA = MM *varlist*	for a matrix of moments about zero
MA = AM *varlist*	for a matrix of moments about zero augmented with a vector of means
MA = CM *varlist*	for a covariance matrix
MA = KM *varlist*	for a matrix of Pearson product moment correlations
MA = PM *varlist*	for a matrix of polyserial, polychoric, and Pearson correlations
MA = OM *varlist*	for a matrix of correlations based on optimal scores

After specifying the scale of the observed variable, PRELIS will keep track of the variable type and compute the correct correlation, given the specification of the form of correlation. Use KM and CM with continuous variables only. When OM is specified, PRELIS will provide correlations that vary depending on the definition of variables. It will compute Pearson correlations for continuous variables, consistent (but asymptotically inefficient) estimates of polyserial correlations for continuous and ordinal variables, and cannonical correlations for pairs of ordinal (ordered categorical) variables. When PM is specified, PRELIS again will provide correlations that vary depending on the definition of variables. It will compute Pearson correlations for pairs of continuous variables, maximum likelihood estimates of polyserial correlations for ordinal and continouous variables, and maximum likelihood estimates of polychoric correlations for pairs of ordinal (ordered categorical) variables. Note that tetrachoric correlations are included under polychoric correlations when both categorical variables are dichotomous; biserial correlations are included under polyserial correlations when one variable is dichotomous and the other is continuous. Note also that censored variables are treated as continuous after being transformed to normal scores.

Jöreskog and Sörbom (1986) present some Monte Carlo evidence on the different kinds of correlations produced by PRELIS for ordinal and continuous variables. They find that, in terms of bias and mean square error (bias plus inefficiency), polychoric and polyserial correlations estimated by maximum likelihood performed best (smallest bias and mean square error). Second best were correlations specified under OM (cannonical correlations based on optimal scores and consistent estimates of polychoric correlations). Pulling up the rear were product-moment correlations on normal scores and Pearson correlations on raw scores. Thus, based on their very limited Monte Carlo evidence, they recommend polychoric and polyserial correlations. In the absence of more evidence, this appears to be a reasonable recommendation. Note: You can run your own Monte Carlo simulation on your model using the GENRAW algorithm that comes with PRELIS.

In preparation for using WLS in LISREL 8, save the moment matrix estimated and the asymptotic covariance matrix of sample moments using SM and SA on the OU line:

SM = *filename* saves the matrix specified under MA (**S** or **R**)

SA = *filename* saves the asymptotic covariance matrix ($\mathbf{S}_{ss}$ or $\mathbf{S}\hat{\rho}\hat{\rho}$)

SV = *filename* saves the asymptotic variances (diagonal elements of $\mathbf{S}_{ss}$ or $\mathbf{S}\hat{\rho}\hat{\rho}$)

The latter matrix, asymptotic variances can be used to obtain diagonally-weighted least squares estimates, which are consistent but asymptotically inefficient, but require less storage space and less computing time.

To obtain WLS estimates in LISREL 8, read in both the sample moment matrix and the asymptotic covariance matrix. First, specify the input data, which takes two lines for each data file:
FI indicates filename; FO indicates the format of the data.

LA Variable labels.
RA Raw data.
MM Matrix of moments about zero.
CM Covariance matrix.
KM Correlation matrix (Pearson product-moment correlations).
OM Matrix of optimal scores from PRELIS.
PM Matrix of correlations, including polychoric or polyserial correlations (used in estimating a model with ordinal or dichotomous dependent variables).
ME Vector of means.
SD Vector of standard deviations.
AC Asymptotic covariance matrix of the covariance or correlation matrix.
AV Asymptotic variances of the covariance or correlation matrix.
DM User defined weight matrix.

The input data line has the following format (for e.g., a matrix of polychoric, polyserial or Pearson correlations):

PM FI = *filename* FO
(format)

If the matrix has been written by PRELIS, the fortran format will be (6D13.6), and will appear on the first line of the saved file (so you don't have to specify it). For example,

PM FI = BADCORR.DAT FO

will read in BADCORR.DAT, a correlation matrix estimated and written by PRELIS in format 6D13.6. Also include the file describing the location of the asymptotic covariance matrix (or asymptotic variances) estimated by PRELIS:

AC FI = ASYCOV.DAT FO

AV FI = ASYVAR.DAT FO

The second step is to specify the matrix to be analyzed (which must correspond to the input matrix when using WLS). On the data line, specify

DA MA = PM

There are six options for MA (default is CM):

MM Matrix of moments about zero.
AM Matrix of moments augmented by a vector of constants (for estimating means using LISREL 6).
CM Covariance matrix.
KM Correlation matrix (Pearson product-moment correlations).
OM Matrix of optimal scores from PRELIS.
PM Matrix of correlations, including polychoric or polyserial correlations (used in estimating a model with ordinal or dichotomous dependent variables) from PRELIS.

Then to specify the estimation procedure, select either WL (for weighted least squares, when reading in the asymptotic covariance matrix of sample moments and using it as the weight matrix) or DL (for diagonally weighted least squares when reading in the asymptotic variances of sample moments and using it as the diagonal weight matrix) as the estimation procedure using the ME (method of estimation) parameter on the OU card:

ME = *method of estimation*

WL for weighted least squares (WLS).
DW for diagonally weighted least squares (DWLS).

Again, the estimators based on normal theory will be consistent, but asymptotically inefficient, and yield biased estimates of standard errors and χ^2s. For a detailed outline of the PRELIS and LISREL 8 programs, see lecture 8 of the notes.

At this writing, EQS provides the asymptotic distribution free WLS estimator for continuous variables, but does not provide correlations or weight matrices for ordinal variables (this will probably be included in the next release of EQS). LISCOMP (Bengt Muthén's program) provides the WLS estimator for continuous as well as ordinal and censored variables. It also provides a probit-type of estimator for measurement models.