Scientific Software International

Methods of estimation

Contents

1. Introduction

For problems that do not include a structure on the means, LISREL can perform any of seven methods of estimation:

- Instrumental Variables (IV)
- Two-Stage Least Squares (TSLS)
- Unweighted Least Squares (ULS)
- Generalized Least Squares (GLS)
- Maximum Likelihood (ML)
- Generally Weighted Least Squares (WLS)
- Diagonally Weighted Least Squares (DWLS)

The purpose of these methods is to estimate the free and constrained parameters of the model from the sample covariance matrix **S**.

2. Instrumental variables (IV) and two-stage least squares (TSLS)

The ULS, GLS, ML, WLS and DWLS methods of estimation are iterative and require approximations to the parameters in order to begin the computations. These starting values of the parameters are computed in LISREL by the IV and TSLS methods. Although the resulting estimates are not *efficient* in the sense of having minimum large-sample variance, they are *consistent* and have the advantage of being very easy to compute. IV and TSLS do not provide standard errors of the estimated parameters.

When the model fits the data well, the starting values produced by the program are often so close to the iterated solution that only a few iterations are required to compute these solutions. For some models, the estimated starting values are identical to ML and other estimates. To emphasize the fact that the starting values are estimates in their own right, we call them *initial estimates* instead of starting values. One can choose to obtain only these initial estimates and not to compute ML or other estimates. In particular, this may be used with large models to save computer time, especially when the model is only tentative. In such a situation the initial estimates themselves or other information in the output may suggest ways to improve the model.

In most cases, one need only specify the non-zero fixed values in each column of $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$ necessary to fix the scales for $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, and leave it to the program to compute estimates of all the free parameters. However, one *can* specify starting values for any number of free parameters. These are sometimes necessary in non-standard models to give the program some help in starting the iterations.

A key concept in the computation of initial estimates (IV and TSLS) is that of a *reference variable*. A reference variable for a latent variable is an observed variable that represents the latent variable in the sense of being a valid and reliable measure of it. There can be only one reference variable for each latent variable. Although the selection of reference variables is done automatically by the program, there is a connection between the selection of reference variables and the assignment of scales to the latent variables. If one assigns scales to the latent variables by fixing a non-zero value in each column of $\boldsymbol{\Lambda}_x$ and $\boldsymbol{\Lambda}_y$, then the variables for which these non-zero values have been fixed will be reference variables. In this way, users can specify the reference variables explicitly.

The initial estimates are computed in four steps as follows:

1. *Reference variables* are determined as follows. If the scales for $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ have been fixed by assigning a non-zero fixed value in each column of $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$, LISREL will determine non-singular submatrices of $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ of m and n rows, respectively, if this is possible. The rows of these submatrices determine the reference variables. In this way $m+n$ reference variables can be determined, provided $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ contain m and n linearly independent rows, respectively. Note that this requires that $m \le p$ and $n \le q$.

   Remember that all parameters are zero by default. This means that, when this process is applied, all elements of $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are zero except those that have been assigned non-zero values by the user. If no non-zero values have been assigned, both $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are zero so that this procedure fails.

   If the scales of $\boldsymbol{\xi}$ have been fixed by standardizing $\boldsymbol{\Phi}$, the program will automatically assign fixed values in the columns of $\boldsymbol{\Lambda}_x$, relax the fixed diagonal elements of $\boldsymbol{\Phi}$ and use the same procedure as above to determine the reference variables. When the initial estimates for $\boldsymbol{\Lambda}_x$ have been determined in Step 2, the program will rescale the $\boldsymbol{\xi}$-variables so that $diag(\boldsymbol{\Phi}) = \mathbf{I}$.

2. For each row of $\boldsymbol{\Lambda}$ ($\boldsymbol{\Lambda}_y$ or $\boldsymbol{\Lambda}_x$), the free parameters, if any, are estimated from the linear relation between each observed variable and the reference variables using all other observed variables as instrumental variables.
3. For given $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$, the joint covariance matrix of $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, $\boldsymbol{\varepsilon}$, and $\boldsymbol{\delta}$ is estimated by unweighted least squares (ULS) applied to $\mathbf{S}-\boldsymbol{\Sigma}$. For given $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ this leads to a quadratic function which can be minimized easily. Parameters in $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}_\varepsilon$ and $\boldsymbol{\Theta}_\delta$ that have non-zero values are held fixed during this minimization.
4. When the joint covariance matrix of $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ has been estimated as in Step 3, the structural equation system can be estimated by instrumental variables methods. We estimate each equation separately using all the $\boldsymbol{\xi}$-variables as instrumental variables. Again, non-zero parameters in $\mathbf{B}$ and $\boldsymbol{\Gamma}$ are held fixed. The estimates computed in this step are identical to the well-known instrumental variables or two-stage least-squares estimators; see, for example, Goldberger (1964).

3. Unweighted least squares (ULS)

The ULS estimator minimizes the function

$$F = \frac{1}{2} tr\left[\left(\mathbf{S}-\boldsymbol{\Sigma}\right)^2\right] \qquad (1)$$

where $tr\left[\left(\mathbf{S}-\boldsymbol{\Sigma}\right)^2\right]$ represents the sum of squares of the elements in the (symmetric) residual matrix, $\mathbf{S}-\boldsymbol{\Sigma}$, of order $p+q$.

The ULS estimator is consistent and relatively quick to compute, but is not efficient – that is, does not attain minimum large-sample variances. Standard errors for ULS are estimated under normal theory. Robust ULS standard errors based on Browne (1984) are computed when the RO LISREL command or the Robust Estimation SIMPLIS command is used.

4. Generalized least squares (GLS)

GLS minimizes

$$F = \frac{1}{2} tr\left[\left(\mathbf{I}-\mathbf{S}^{-1}\boldsymbol{\Sigma}\right)^2\right] \qquad (2)$$

equivalent to minimizing the sum of squares of the residuals weighted by the inverse of the sample covariance matrix. The estimator is consistently efficient and has large-sample standard errors computed by LISREL under normal theory. Robust GLS standard errors based on Browne (1984) are computed when the RO LISREL command or the Robust Estimation SIMPLIS command is used.

5. Maximum likelihood (ML)

ML maximizes the likelihood of the parameters, given the data. In the present context, it is equivalent to minimizing

$$F = \log\|\boldsymbol{\Sigma}\| + tr(\mathbf{S}\boldsymbol{\Sigma}^{-1}) - \log\|\mathbf{S}\| - (p+q). \qquad (3)$$

ML has the same properties as GLS and is about equally time consuming to compute. Robust ML standard errors based on Browne (1984) are computed when the RO LISREL command or the Robust Estimation SIMPLIS command is used.

6. Generally weighted least-squares (WLS)

The ULS, GLS, and ML fit functions, are, in effect, special cases of the more general function for fitting covariance structures,

$$\begin{aligned} F(\boldsymbol{\theta}) &= (\mathbf{s}-\boldsymbol{\sigma})' \mathbf{W}^{-1} (\mathbf{s}-\boldsymbol{\sigma}) \\ &= \sum_{g=1}^{k}\sum_{h=1}^{g}\sum_{i=1}^{k}\sum_{j=1}^{i} w^{gh,ij} (s_{gh}-\sigma_{gh})(s_{ij}-\sigma_{ij}), \end{aligned} \qquad (4)$$

where

$$\mathbf{s}' = (s_{11}, s_{21}, s_{22}, s_{31}, ..., s_{kk}),$$

is a vector of the elements in the lower half, including the diagonal, of the covariance matrix $\mathbf{S}$ of order $k \times k$ used to fit the model to the data;

$$\boldsymbol{\sigma}' = (\sigma_{11}, \sigma_{21}, \sigma_{22}, \sigma_{31}, ..., \sigma_{kk}),$$

is the vector of corresponding elements of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ reproduced from the model parameters $\boldsymbol{\theta}$;

$$w^{gh,ij}$$

is a typical element of a positive definite matrix $\mathbf{W}^{-1}$ of order $u \times u$, where $u = k(k+1)/2.$ In most cases, the elements of $\mathbf{W}^{-1}$ are obtained by inverting a matrix $\mathbf{W}$ whose typical element is denoted $w_{gh,ij}$. The usual way of choosing $\mathbf{W}$ in weighted least squares is to let $w_{gh,ij}$ be a consistent estimate of the asymptotic covariance between s_{gh} and s_{ij}. If this is the case, we say that $\mathbf{W}^{-1}$ is the *correct weight matrix*. To estimate the model parameters $\boldsymbol{\theta}$, the fit function is minimized with respect to $\boldsymbol{\theta}$.

To obtain consistent estimates, any positive definite matrix $\mathbf{W}$ may be used. Under very general assumptions, if the model holds in the population and if the sample variances and covariances in $\mathbf{S}$ converge in probability to the corresponding elements in the population covariance matrix $\boldsymbol{\Sigma}$ as the sample size increases, any fit function with a positive definite $\mathbf{W}$ will be a consistent estimator of $\boldsymbol{\theta}$. In practice. numerical results obtained by one fit function are often close enough to the results that would be obtained by another fit function to give the same substantive interpretation of the results.

Further assumptions must be made, however, if one needs an asymptotically correct chi-square measure of goodness-of-fit and asymptotically correct standard errors of parameter estimates.

"Classical" theory for covariance structures (see, for example, Browne, 1974, or Jöreskog, 1981) assumes that the asymptotic variances and covariances of the elements of **S** are of the form

$$\mathrm{ACov}\left(s_{gh}, s_{ij}\right) = (1/N)(\sigma_{gh}\sigma_{hj} + \sigma_{gj}\sigma_{hi}), \tag{5}$$

where *N* is the total sample size. This holds, in particular, if the observed variables have a multivariate normal distribution, or if **S** has a Wishart distribution. The GLS and ML methods and their chi-square values and standard errors are based on (5). The GLS method corresponds to using in (4) a matrix $\mathbf{W}^{-1}$, which has as a general element

$$w^{gh,ij} = N(2-\delta_{gh})(2-\delta_{ij})(s^{gi}s^{hj} + s^{gj}s^{hi}), \tag{6}$$

where δ_{gh} and δ_{ij} are Kronecker deltas. The fit function (3) for ML is not of the form (4) but may be shown to be equivalent to using a $\mathbf{W}^{-1}$ of the form (6), with *s* replaced by an estimate of σ, which is then updated in each iteration.

In fundamental work by Browne (1982, 1984), this classical theory for covariance structures has been generalized to any multivariate distribution for continuous variables satisfying very mild assumptions. This approach uses a **W** matrix with typical element

$$w_{gh,ij} = m_{ghij} - s_{gh}s_{ij}, \tag{7}$$

where

$$m_{ghij} = (1/N)\sum_{a=1}^{N}(z_{ag} - \bar{z}_g)(z_{ah} - \bar{z}_h)(z_{ai} - \bar{z}_i)(z_{aj} - \bar{z}_j) \tag{8}$$

are the fourth-order central moments. Using such a **W** in (4) gives what Browne calls "asymptotically distribution-free best GLS estimators" for which asymptotically correct chi-squares and standard errors may be obtained. As shown by Browne, this **W** matrix may also be used to compute asymptotically correct chi-squares and standard errors for estimates which have been obtained by the classical ML and GLS methods.

When **W** is defined by (7), we call the fit function WLS (Weighted Least Squares) to distinguish it from GLS, where **W** is defined by (6). WLS and GLS are different forms of weighted le3ast squares: WLS is asymptotically distribution free, while GLS is based on normal theory.

While WLS is attractive in theory it presents several difficulties in practical applications. First, the matrix **W** is of order $u \times u$ and has $u(u+1)/2$ distinct elements. This increases rapidly with *k*, demanding large amounts of computer memory when *k* is at all large. For example, when *k* = 20, **W** has 22,155 distinct elements. Second, to estimate moments of fourth order with reasonable precision requires very large samples. Third, when there are missing observations in the data, different moments involved in (7) may be based on different numbers of cases unless listwise deletion is used. When pairwise deletion is used, it is not clear how to deal with this problem.

Browne's (1984) development is a theory for sample covariance matrices for continuous variables. In practice, correlation matrices are often analyzed, *i.e.*, covariance matrices scaled by stochastic standard deviations. The elements of such a correlation matrix do not have asymptotic variances and covariances of the form (5), even if **S** has a Wishart distribution. In PRELIS, we have extended Browne's (1984) work so that an estimate of the asymptotic covariance matrix of estimated correlations can also be obtained under the same general assumptions of non-normality. This approach can also be used when some or all of the variables are ordinal or censored, if the raw scores are replaced by normal scores. PRELIS can also compute estimates of the asymptotic variances and covariances of estimated polychoric and polyserial correlations.

7. Diagonally weighted least-squares (DWLS)

Computation of asymptotic covariance matrices of estimated coefficients is very time consuming and demands large amounts of memory when the number of variables is large. An alternative approach, which may be used even when the number of variables is large, is to compute only the asymptotic variances of the estimated coefficients.

Let w_{gh} be an estimate of the asymptotic variance of s_{gh}. These estimates may be used with a fit function of the form

$$F(\boldsymbol{\theta}) = \sum_{g=1}^{k}\sum_{h=1}^{g}(1/w_{gh})\left(s_{gh}-\sigma_{gh}\right)^2. \qquad (9)$$

This corresponds to using a diagonal weight matrix $\mathbf{W}^{-1}$ in (4). In LISREL we call this DWLS (Diagonally Weighted Least Squares). This does not lead to asymptotically efficient estimates of model parameters but is offered as compromise between unweighted least squares (ULS) and fully weighted least squares (WLS). Robust DWLS estimates, and standard errors based on Browne (1984) are computed when the RO LISREL command or the Robust Estimation SIMPLIS command is used.

8. The ridge option

The GLS, ML, and WLS methods of estimation require that the sample covariance matrix, **S**, be positive-definite. To provide for situations in which **S** is not quite positive-definite, as sometimes happens in econometric models containing identities and in regression problems with near-multicollinearity among predictors, LISREL includes a provision for so-called *ridge* estimation.

In ridge estimation, a constant times the diagonal of **S** is added to **S**; the constant is determined by the program, but may also be set by the user (the RO option on the OU command). The ridge option may be used with ULS, GLS, ML, WLS and DWLS; it has no effect on IV and TSLS.

9. The Information matrix

Associated with each of the iterative estimation procedures, that minimize a fit function, is an information matrix for the parameters. The order of this matrix is equal; to the number of free parameters in the model. The elements of the matrix are the expected values of the second derivatives of the fit function at the solution point (*i.e.*, the expected Hessian matrix).

The inverse of the information matrix contains the sampling variances, or large-sample variances, of the parameters as its diagonal elements. The off-diagonal elements are the covariances between all of the possible pairs of parameter estimates.

The standard errors of LISREL parameter estimates are the square roots of the diagonal elements of the inverse information matrix; the correlations between the estimates are the off-diagonal elements of the inverse information matrix divided by the corresponding pairs of standard errors. The information matrix must be positive-definite for the model to be identified.

10. Admissibility of the estimates

LISREL does not constrain the parameter estimates to be admissible (for example, diagonal elements of $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}_{\varepsilon}$, or $\boldsymbol{\Theta}_{\delta}$ can become negative if the data are unfavorable relative to the assumed model). The program has a built-in check on

admissibility of the solution, however, and will stop if the solution becomes non-admissible, unless the check is turned off (AD = OFF). The *admissibility* check is that

1. $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ have full column ranks and no rows of only zeros,
2. $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$, $\boldsymbol{\Theta}_\varepsilon$, or $\boldsymbol{\Theta}_\delta$ are positive definite.

It is often possible, though not always, to use various tricks to force the program to stay within the admissible parameter space (see Rindskopf, 1983, 1984, or for an example, see Jöreskog, 1981).

11. References

Browne, M.W. (1974). Generalised Least Squares Estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1-24.

Browne, M.W. (1982). Covariance Structures. In D.M. Hawkins (Ed.), *Topics in Applied Multivariate Analysis*, pp. 72-141. Cambridge: Cambridge University Press.

Browne, M.W. (1984). Asymptotically Distribution-free Methods in the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Goldberger, A.S. (1964). *Econometric Theory*. New York: Wiley.

Jöreskog, K.G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics*, **8**, 65-92.

Rindskopf, D.M. (1983). Using inequality constraints to prevent Heywood cases: The LISREL parameterization. *Data Analyst*, **1**, 1-3.

Rindskopf, D.M. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, **49**, 37-47.