1. Limited Information Estimation

The model considered here is a particular case of the linear simultaneous-equations model. There is a set of g endogenous variables, for which the data-generating process is assumed to take the form

$$\boldsymbol{y}_i = \boldsymbol{Y}_i\boldsymbol{\beta}_i + \boldsymbol{Z}_i\boldsymbol{\gamma}_i + \boldsymbol{u}_i, \quad i = 1, \ldots, g. \tag{1}$$

Here, the columns of the $n \times g_i$ matrix $\boldsymbol{Y}_i$ are a subset of the $\boldsymbol{y}_j$, $j = 1, \ldots, g$, not including $\boldsymbol{y}_i$. The $n \times k_i$ matrix $\boldsymbol{Z}_i$ has columns that are exogenous or predetermined explanatory variables, and the $\boldsymbol{u}_i$ are vectors of random disturbances that are serially uncorrelated but contemporaneously correlated, with contemporaneous covariance matrix $\boldsymbol{\Sigma}$, of dimensions $g \times g$.

All but the first of the g equations (1) are just identified. This means that $g_i + k_i = l$ for $i = 2, \ldots, g$. The first equation is overidentified, so that $g_1 + k_1 < l$. Further, l is the dimension of the space spanned by the columns of all the $\boldsymbol{Z}_i$ matrices. We let $\boldsymbol{W}$ denote an $n \times l$ matrix that spans that space. In order that the equations $i = 2, \ldots, g$ should in fact be identified, we make the assumption that

$$\mathcal{S}(\boldsymbol{Z}_i, \boldsymbol{P_W}\boldsymbol{Y}_i) = \mathcal{S}(\boldsymbol{W}) \tag{2}$$

for $i = 2, \ldots, g$, and that

$$\mathcal{S}(\boldsymbol{Z}_1, \boldsymbol{P_W}\boldsymbol{Y}_1) \subset \mathcal{S}(\boldsymbol{W}),$$

where the notation $\mathcal{S}(\boldsymbol{A})$ denotes the linear span of the columns of the matrix $\boldsymbol{A}$.

It is known that the 3SLS estimates of the parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\gamma}_1$ of the overidentified first equation are the same as the 2SLS estimates of those parameters, where the matrix $\boldsymbol{W}$ is used as the matrix of instruments. However, the estimates of all the other regression coefficients, the $\boldsymbol{\beta}_i$ and the $\boldsymbol{\gamma}_i$, for $i = 2, \ldots, g$, are not the 2SLS estimates of the corresponding equations with instruments $\boldsymbol{W}$. Thus 3SLS and 2SLS are the same only for the overidentified equation, but not for the others. An analogous relation is true for FIML and LIML for the system (1). The parameter estimates of these two methods are the same for the overidentified equation, but not for the others.

The 2SLS estimating equations for the just-identified equations can be written as

$$\boldsymbol{W}^\top(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\delta}_i) = \boldsymbol{0},$$

where $\boldsymbol{X}_i \equiv [\boldsymbol{Y}_i \;\; \boldsymbol{Z}_i]$ and $\boldsymbol{\delta}_i \equiv [\boldsymbol{\beta}_i \,\vdots\, \boldsymbol{\gamma}_i]$. If $\boldsymbol{\delta}_i^0$ denotes the true parameter vector, then the estimating equations can be rewritten as

$$\boldsymbol{W}^\top(\boldsymbol{u}_i - \boldsymbol{X}_i(\boldsymbol{\delta}_i - \boldsymbol{\delta}_i^0)) = \boldsymbol{0},$$

from which it can be seen clearly that the 2SLS estimates, $\hat{\boldsymbol{\delta}}_i$ say, are correlated with the estimates of the parameters of the first equation, unless the covariance matrix $\boldsymbol{\Sigma}$

is diagonal. For bootstrapping, this means that using the 2SLS estimates of equations 2 through g in the definition of the bootstrap DGP gives rise to a correlation between any statistic used to test a hypothesis about the parameters of the first equation and the bootstrap DGP.

If 3SLS estimates are used to define the bootstrap DGP, then this problem goes away, as we will show in a moment. Similarly, if FIML estimates are used for the bootstrap DGP, this DGP is asymptotically independent of test statistics for the parameters of the first equation.

The 3SLS estimating equations for system (1) are as follows:

$$\sum_{j=1}^{g} \sigma^{ij} \begin{bmatrix} \boldsymbol{Z}_i^\top \\ \boldsymbol{Y}_i^\top \boldsymbol{P}_{\boldsymbol{W}} \end{bmatrix} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}, \tag{3}$$

for $i = 1, \ldots, g$. Here, σ^{ij} is element (i, j) of $\boldsymbol{\Sigma}^{-1}$, or, in practice, of an estimate of $\boldsymbol{\Sigma}^{-1}$. From (2), equations 2 through g of (3) are equivalent to

$$\sum_{j=1}^{g} \sigma^{ij} \boldsymbol{W}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}. \tag{4}$$

Equations (4) imply that

$$\sum_{j=1}^{g} \sigma^{ij} \boldsymbol{Z}_1^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}, \quad i = 2, \ldots, g,$$

and these equations, combined with the first k_1 equations of (3) for $i = 1$, that is,

$$\sum_{j=1}^{g} \sigma^{1j} \boldsymbol{Z}_1^\top (\boldsymbol{y}_1 - \boldsymbol{X}_1 \boldsymbol{\delta}_1) = \boldsymbol{0},$$

imply that

$$\boldsymbol{Z}_1^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0} \tag{5}$$

for all $j = 1, \ldots, g$.

From (4), we see that

$$\sum_{j=1}^{g} \sigma^{ij} \boldsymbol{Y}_1^\top \boldsymbol{P}_{\boldsymbol{W}} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}, \quad i = 2, \ldots, g,$$

and these equations, combined with the last g_1 equations of (3) for $i = 1$, imply that

$$\boldsymbol{Y}_1^\top \boldsymbol{P}_{\boldsymbol{W}} (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}, \quad j = 1, \ldots, g. \tag{6}$$

Equations (5) and (6) then imply that (3) for $i = 1$ reduces to

$$\boldsymbol{X}_1^\top \boldsymbol{P}_{\boldsymbol{W}} (\boldsymbol{y}_1 - \boldsymbol{X}_1 \boldsymbol{\delta}_1) = \boldsymbol{0},$$

thereby confirming that 3SLS and 2SLS for the first equation give the same estimator.

Let $\boldsymbol{W}^1$ be an $n \times (l - k_1)$ matrix the columns of which span the orthogonal complement of k_1–dimensional space $\mathcal{S}(\boldsymbol{Z}_1)$ within the l–dimensional space $\mathcal{S}(\boldsymbol{W})$. Then the equations of (4) that have not as yet been used can be written as

$$\sum_{j=1}^{g} \sigma^{ij} (\boldsymbol{W}^1)^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}, \quad i = 2, \ldots, g.$$

Along with (5), this shows that, for $i = 2, \ldots, g$,

$$\sum_{j=1}^{g} \sigma^{ij} \boldsymbol{W}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \boldsymbol{0}.$$

These equations can be rewritten as

$$\sum_{j=2}^{g} \sigma^{ij} \boldsymbol{W}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = -\sigma^{i1} \boldsymbol{W}^\top (\boldsymbol{y}_1 - \boldsymbol{X}_1 \boldsymbol{\delta}_1). \tag{7}$$

If the elements of $\boldsymbol{\Sigma}^{-1}$ are given appropriately, equations (7) serve as estimating equations for the 3SLS estimates of the parameters of the just-identified equations, where $\boldsymbol{\delta}_1$ is given by the 2SLS estimates of the first equation.

Let $\boldsymbol{\Sigma}^{(1,1)}$ denote the matrix $\boldsymbol{\Sigma}^{-1}$ without its first row and column, let $\boldsymbol{\Sigma}^{(1),1}$ be the first column of $\boldsymbol{\Sigma}^{-1}$ without its first element, and let $\boldsymbol{\Sigma}_{(1),1)}$ be the first column of $\boldsymbol{\Sigma}$ without its first element. Then standard manipulations of the inverses of partitioned matrices show that

$$\boldsymbol{\Sigma}^{(1),1} = -\boldsymbol{\Sigma}^{(1,1)} \boldsymbol{\Sigma}_{(1),1} \frac{1}{\sigma_{11}},$$

from which we see that

$$\left(\boldsymbol{\Sigma}^{(1,1)}\right)^{-1} \boldsymbol{\Sigma}^{(1),1} = -\frac{1}{\sigma_{11}} \boldsymbol{\Sigma}_{(1),1}. \tag{8}$$

For each $k = 1, \ldots, l$, let a_{kj} denote the k^{th} component of $\boldsymbol{W}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j)$. Then taking the k^{th} components of each of the $g - 1$ equations of (7) gives a system of $g - 1$ linear simultaneous equations with unknowns the a_{kj}, k fixed, $j = 2, \ldots, g$, with coefficient matrix $\boldsymbol{\Sigma}^{(1,1)}$, and with right-hand sides the a_{k1} times the negative of the vector $\boldsymbol{\Sigma}^{(1),1}$.

Solving these equations gives the result

$$\boldsymbol{a}_k = -\left(\boldsymbol{\Sigma}^{(1,1)}\right)^{-1} \boldsymbol{\Sigma}^{(1),1} a_{k1}, \tag{9}$$

where $\boldsymbol{a}_k$ is the $g - 1$–vector with typical component a_{kj}, $j = 2, \ldots, g$. From (8) and (9), we find that

$$a_{kj} = \frac{\sigma_{j1}}{\sigma_{11}} a_{k1},$$

or, recalling the definition of the a_{kj},

$$\boldsymbol{W}^\top (\boldsymbol{y}_j - \boldsymbol{X}_j \boldsymbol{\delta}_j) = \frac{\sigma_{j1}}{\sigma_{11}} \boldsymbol{W}^\top (\boldsymbol{y}_1 - \boldsymbol{X}_1 \boldsymbol{\delta}_1).$$

Thus the 3SLS estimating equations for the $\boldsymbol{\delta}_j$, $j = 2, \ldots, g$ are

$$\boldsymbol{W}^\top\Big(\boldsymbol{y}_j - \boldsymbol{X}_j\boldsymbol{\delta}_j - \frac{\sigma_{j1}}{\sigma_{11}}(\boldsymbol{y}_1 - \boldsymbol{X}_1\hat{\boldsymbol{\delta}}_1)\Big) = \mathbf{0}, \tag{10}$$

where $\hat{\boldsymbol{\delta}}_1$ is the vector of 2SLS estimates of the parameters of the overidentified equation, and the σ_{ij} are defined by the equations

$$\sigma_{ij} = \frac{1}{n}\hat{\boldsymbol{u}}_i^\top\hat{\boldsymbol{u}}_j,$$

where the $\hat{\boldsymbol{u}}_i$ are the 2SLS residuals obtained by equation-by-equation estimation of the g equations of the system.

Now consider artificial regressions of the sort

$$\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\delta}}_j = \boldsymbol{P_W}\boldsymbol{X}_j\boldsymbol{d}_j + \alpha_j(\boldsymbol{y}_1 - \boldsymbol{X}_1\hat{\boldsymbol{\delta}}_1) + \text{residuals}, \quad j = 2, \ldots, g, \tag{11}$$

where $\hat{\boldsymbol{\delta}}_j$ is the vector of 2SLS estimates for equation j. These regressions should be run using ordinary least squares, where the parameters to be estimated are the elements of the vectors $\boldsymbol{d}_j$ and the α_j. A subset of the OLS estimating equations for (11) is given by

$$\boldsymbol{X}_j^\top\boldsymbol{P_W}\big(\boldsymbol{y}_j - \boldsymbol{X}_j\hat{\boldsymbol{\delta}}_j - \boldsymbol{P_W}\boldsymbol{X}_j\boldsymbol{d}_j - \alpha_j(\boldsymbol{y}_1 - \boldsymbol{X}_1\hat{\boldsymbol{\delta}}_1)\big) = \mathbf{0}. \tag{12}$$

Since, for $j \neq 1$, $\mathcal{S}(\boldsymbol{P_W}\boldsymbol{X}_j) = \mathcal{S}(\boldsymbol{W})$, by (2), and since $\boldsymbol{W}^\top\boldsymbol{P_W}\boldsymbol{X}_j = \boldsymbol{W}^\top\boldsymbol{X}_j$, equations (12) are equivalent to

$$\boldsymbol{W}^\top\big(\boldsymbol{y}_j - \boldsymbol{X}_j(\hat{\boldsymbol{\delta}}_j + \boldsymbol{d}_j) - \alpha_j(\boldsymbol{y}_1 - \boldsymbol{X}_1\hat{\boldsymbol{\delta}}_1)\big) = \mathbf{0}. \tag{13}$$

If we premultiply the regressions (11) by the projection $\boldsymbol{M_W}$, then by the FWL Theorem, the estimates of the α_j are the same as those obtained from the OLS regressions

$$\hat{\boldsymbol{u}}_j = \alpha_j\boldsymbol{M_W}\hat{\boldsymbol{u}}_1 + \text{residuals},$$

that is,

$$\hat{\alpha}_j = \frac{\hat{\boldsymbol{u}}_1^\top\hat{\boldsymbol{u}}_j}{\hat{\boldsymbol{u}}_1^\top\boldsymbol{M_W}\hat{\boldsymbol{u}}_1}.$$

For the purposes of estimating $\boldsymbol{\Sigma}$, we may just as well use the projected residuals $\boldsymbol{M_W}\hat{\boldsymbol{u}}_1$ for the overidentified equation as the residuals $\hat{\boldsymbol{u}}_1$ themselves. Thus the artificial regressions (11) yield estimates that differ from the 3SLS estimates only by the fact that the σ_{1j} are defined using these projected residuals for the first equation. Asymptotic equivalence of the $\hat{\boldsymbol{\delta}}_j + \hat{\boldsymbol{d}}_j$ and the 3SLS estimates of $\boldsymbol{\delta}_j$ is immediate, and, indeed, one can expect that the numerical values of the estimates will be very close in most cases. (If one changes the residuals of the overidentified equation to the projected residuals in a standard 3SLS estimation algorithm, the estimates thus obtained are in fact numerically identical to those given by the artificial regressions.)

Exactly similar reasoning can be used for FIML and LIML. The LIML estimating equations are in the present setup exactly the same as the FIML ones, since there is only one overidentified equation. The estimates of the parameters of the equations other than the first are defined by estimating equations of the form (7), with $\boldsymbol{\delta}_1$ now given by the LIML estimates of the parameters of the first equation rather than the 2SLS ones. In addition, the estimates used for the elements of $\boldsymbol{\Sigma}^{-1}$ are different from those used for 3SLS. However, use of any of the usual consistent estimates of $\boldsymbol{\Sigma}^{-1}$ leads to asymptotically equivalent estimators.

We now show that the estimates defined by the equations (7) are asymptotically uncorrelated with the estimates, 2SLS or LIML, of the parameters of the first equation. Since the latter are random only through the disturbances $\boldsymbol{u}_1$, it is enough to show that the linear combinations

$$\sum_{j=1}^{g} \sigma^{ij} \boldsymbol{u}_j \tag{14}$$

that occur in the estimating equations (7) are uncorrelated with $\boldsymbol{u}_1$ for $i \neq 1$. The covariance of $\boldsymbol{u}_1$ and the linear combination (14) is

$$\sum_{j=1}^{g} \sigma^{ij} \mathrm{E}(\boldsymbol{u}_j \boldsymbol{u}_1) = \sum_{j=1}^{g} \sigma^{ij} \sigma_{j1} = 0 \text{ for } i \neq 1.$$

since the σ^{ij} and the σ_{j1} are elements of inverse matrices. This proves the desired result. If $\boldsymbol{\Sigma}^{-1}$ is an estimate, then the above reasoning shows that the estimates of the parameters of the first equation are asymptotically independent of those given by (7) or the artificial regressions (11).

This method of estimation, first by performing equation-by-equation estimation by instrumental variables, or by ordinary least squares for those equations that have no endogenous explanatory variables, followed by re-estimation of the just identified equations with one additional regressor given by the residuals from the overidentified equation, plainly leads to estimators that are asymptotically equivalent to 3SLS or FIML whenever the first estimation procedure is root-n consistent. Thus all k–class estimators such that $\operatorname{plim}_{n\to\infty} k = 1$ can be used for the first set of estimators. A particularly interesting case of such a k–class estimator is Fuller's modification of the LIML estimator. All these estimators have different finite-sample properties of course.