Simultaneous equations systems

Since economic variables are jointly dependent, econometric models are often specified as systems of simultaneous equations, represented in what is sometimes called a ‘structural’ form, denoting a model of the conditional system. However, alternative values of the coefficients in the system may yield the same likelihood, precluding estimation: this is the identification problem. Conditions for identifiability are applied in §11.1, where methods of estimation and testing for over-identifying restrictions are suggested for the identified equations. The score of the likelihood function yields a set of equations for full-information maximum likelihood (FIML). However, appropriate choices in the score for the error-variance matrix and the system parameters (often called reduced form) also lead to many alternative estimators, some at a cost of reduced efficiency. In particular, three-stage least squares (3SLS) is obtained by estimating the error variance by two-stage least squares (2SLS) and the parameters of the system by OLS. 3SLS is asymptotically equivalent to FIML, whereas 2SLS is consistent, but inefficient if the error-variance matrix is non-diagonal (§11.2). Equally, 2SLS is IV using optimal instruments, in that any other set of instruments different from the whole set of conditioning variables in the system, leads to an estimator with further reduced efficiency (see §11.3). In finite samples, the numerical value of IV changes with the choice of normalization, whereas LIML is unaffected (see §11.4). Simultaneity precludes OLS estimation of the simultaneous form, but §11.5 reminds us that for just-identification, indirect least-squares coincides with 2SLS which is consistent, and OLS of the system parameters provides FIML of the simultaneous-form parameters.

11.1 Identification

Consider the simultaneous-equations model:

\[ \mathbf{B} y_t + \mathbf{C} z_t = \epsilon_t \]  
(11.1)

where \( \mathbb{E}[z_t \epsilon_t'] = 0 \) \( \forall t, s \). In (11.1), \( y_t \) and \( z_t \) are endogenous and strongly exogenous for \( (\mathbf{B}, \mathbf{C}) \) respectively, and the detailed specification of the equations is given by:

\[ y_{1,t} + \beta_{12} y_{2,t} + \beta_{13} y_{3,t} + \gamma_{11} z_{1,t} + \gamma_{13} z_{3,t} = \epsilon_{1,t} \]  
(11.2)

\[ \beta_{21} y_{1,t} + y_{2,t} + \gamma_{21} z_{1,t} + \gamma_{23} z_{3,t} = \epsilon_{2,t} \]  
(11.3)
\[ \beta_{31} y_{1,t} + y_{3,t} + \gamma_{33} z_{3,t} = \varepsilon_{3,t}. \] (11.4)

(1) Discuss the identification of the parameters in terms of rank and order conditions.
(2) How would you estimate any identifiable unknown parameters if \( \varepsilon_t \sim \mathcal{N}(0, \Sigma) \)?
(3) How would you test any available over-identifying restrictions?

11.1.1 Parameter identification

System coefficients are the parameters in the joint distribution of the data, and hence are always identified. Thus, identification of coefficients in the simultaneous form involves solving uniquely for those coefficients from the system parameters. The total number of restrictions needed for identifying the simultaneous-form parameters can be computed as the difference between the number of parameters in those two sets. In general, when \( \mathbf{B} \) and \( \mathbf{C} \) in (11.1) are \( n \times n \) and \( n \times k \) matrices, then \( \mathbf{P} = -\mathbf{B}^{-1} \mathbf{C} \) is \( n \times k \), and the variance matrix of the system disturbances is \( n \times n \), so that there are \( nk + n (n + 1) / 2 \) system parameters from which to derive \( n(n + k - 1) + n (n + 1) / 2 \) simultaneous-form parameters. Thus, at least \( n(n - 1) \) restrictions are needed for identification after the normalization restrictions have been imposed.

There are two conditions for identifying an equation: an order condition which is just necessary, and a rank condition which is both necessary and sufficient. The order condition states that the number of excluded exogenous variables should be no less than the number of included endogenous variables less one. Denoting by \( (\cdot)^v \) the vectoring operation of a matrix, defining \( \alpha = \mathbf{A}^v = (\mathbf{B} : \mathbf{C})^v \) and writing the restrictions as \( \Phi \alpha = 0 \), where \( \Phi \) is a selection matrix of zeros and ones, the rank condition states that the simultaneous form is identified if and only if \( \text{rank}(\Phi) \geq n(n - 1) \). The \( j \)th equation is identified if and only if \( \text{rank}(\Phi_j \mathbf{A}^j) = n - 1, j = 1, \ldots, n, \) where \( \Phi_j \) is such that \( \Phi_j \alpha^j = 0 \) and \( \alpha^j \) is the \( j \)th row of \( \mathbf{A} \).

The order condition is an implication of the rank condition. This is so because the total number of variables in the system is \( n + k \), so that a particular equation can have at most \( n + k - 1 \) right-hand side variables. However, if the equation is identified, the rank condition establishes that the number of excluded variables must be \( n - 1 \), so that the equation being considered must have \( n + k - 1 - (n - 1) = k \) variables at most in its right-hand side. In addition, \( k \) is the total number of exogenous variables in the system, and so the order condition for identification may be viewed in terms of the number of optimal instruments available for estimation.

Let us next apply these conditions to the system (11.2)–(11.4). To do so let us write the system as:

\[
\begin{pmatrix}
1 & \beta_{12} & \beta_{13} \\
\beta_{21} & 1 & 0 \\
\beta_{31} & 0 & 1
\end{pmatrix}
\begin{pmatrix}
y_{1,t} \\
y_{2,t} \\
y_{3,t}
\end{pmatrix}
+ 
\begin{pmatrix}
\gamma_{11} & 0 & \gamma_{13} \\
\gamma_{21} & \gamma_{22} & 0 \\
0 & 0 & \gamma_{33}
\end{pmatrix}
\begin{pmatrix}
z_{1,t} \\
z_{2,t} \\
z_{3,t}
\end{pmatrix}
= \varepsilon_t,
\]
so that:
\[ A = \begin{pmatrix} 1 & \beta_{12} & \beta_{13} & \gamma_{11} & 0 & \gamma_{13} \\ \beta_{21} & 1 & 0 & : & \gamma_{21} & \gamma_{22} & 0 \\ \beta_{31} & 0 & 1 & 0 & 0 & \gamma_{33} \end{pmatrix}. \]

The first equation does not satisfy the order condition because there are fewer restrictions (one) than endogenous variables less one (2) and hence it is not identified. Besides the normalization restriction, there are 2 and 3 restrictions in the second and third equations, respectively, so that the order condition is satisfied by both equations. To check for the rank condition, construct the selection matrices that specify where the restrictions occur (in columns) and how many there are (the number of rows):

\[
\Phi_2 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \Phi_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix},
\]

so that:

\[
\Phi_2 A' = \begin{pmatrix} \beta_{13} & 0 & 1 \\ \gamma_{13} & 0 & \gamma_{33} \end{pmatrix} \quad \text{and} \quad \Phi_3 A' = \begin{pmatrix} \beta_{12} & 1 & 0 \\ \gamma_{11} & \gamma_{21} & 0 \\ 0 & \gamma_{22} & 0 \end{pmatrix},
\]

hence \( \text{rank}(\Phi_2 A') = 2 \) and \( \text{rank}(\Phi_3 A') = 2 \) implying that these two equations are identified. In fact, the second equation is just-identified, and the third is over-identified, subject to non-zero values for the relevant parameters.

### 11.1.2 Estimation of identified parameters

The last two equations can be estimated by single equation or subsystem IV. However, if \( \epsilon \sim \mathcal{N}_n[0, \Sigma] \) then we can compute LIML. In a single-equation approach, LIML equals 2SLS in the second equation because this equation is just-identified, and both estimators are asymptotically equivalent in the third equation due to over-identification. Since the error-variance matrix is not diagonal there would be a gain in efficiency by applying sub-system LIML or 3SLS to the last two equations jointly. These two estimators are asymptotically equivalent if \( \Sigma \) is unrestricted.

### 11.1.3 Testing for over-identifying restrictions

If LIML is available then a LR test can be computed to test for over-identifying restrictions. The statistic is \( T \log \lambda \) where \( \lambda \) is the smallest root of the determinantal equation:

\[
\det (M_i - \lambda M) = 0,
\]

where \( M = I_T - Z(Z'Z)^{-1}Z' \) and \( M_i = I_T - Z_i(Z'_iZ_i)^{-1}Z'_i \) with \( Z_i \) being the matrix of exogenous variables in the \( i \)th equation. \( T \log \lambda \) is asymptotically distributed as a \( \chi^2(m) \) where \( m \) is the number of over-identifying restrictions in the \( i \)th equation.
Over-identifying restrictions may also be tested after IV estimation: the discrepancy between the minimized criterion function $\epsilon'W(W'W)^{-1}W'\epsilon$ under the null and the alternative is also asymptotically distributed as a $\chi^2(m)$. An LM test can also be performed, based on the Lagrange multiplier $\lambda$ obtained by $\min_\alpha \epsilon'Z(Z'Z)^{-1}Z'\epsilon$ subject to $\Phi\alpha = 0$.

11.2 Estimator generating equation

In the linear over-identified simultaneous-equations model:

$$Ax_t = By_t + Cz_t = \epsilon_t$$ (11.5)

where $\epsilon_t \sim \mathcal{N}[0, \Sigma]$, $A^{nu} = \theta$ are the $p$ unrestricted parameters, $(\cdot)^{nu}$ is a combined vectoring and selection operator, and $z_t$ is strongly exogenous for $\theta$ with:

$$E[x_t | z_t] = Pz_t$$ (11.6)

when $P^t = (I': I_k)$ and $P = -B^{-1}C$, the score vector is (using $X^t = (x_1, \ldots, x_T)$ etc.):

$$(\Sigma^{-1}AX'ZP^t)^{nu} = 0.$$ (11.7)

$(p$ equations to be solved for $\theta$) with:

$$\Sigma = T^{-1}AX'XA'.$$ (11.8)

1. Explain why it is sensible to solve an expression like (11.7) for $\theta$ given (11.5).
2. Under what conditions are solutions to (11.7) as efficient as FIML and why?
3. Describe three solutions to (11.7) which yield consistent estimates of $\theta$, not all of the same asymptotic efficiency, and derive the limiting distribution of one of these.

(Oxford M.Phil., 1982, 1989)

11.2.1 FIML estimates

Solving (11.7) yields FIML estimates of $B$, $C$ and $\Sigma$ which, under appropriate conditions, are consistent and reach the Cramér–Rao bound asymptotically. To see that (11.7) yields FIML, since conditional upon the $z_s$, $\epsilon_t \sim \mathcal{N}[0, \Sigma]$:

$$y_t = B^{-1}Cz_t + B^{-1}\epsilon_t = \Pi z_t + v_t,$$

so $v_t$ is a linear combination of the $\epsilon_t$, and so is also independently normally distributed with:

$$E[y_t | z_t] = \Pi z_t,$$
and $\mathbb{V}[y_t | z_t]$ given by:

$$E[(y_t - \Pi z_t)(y_t - \Pi z_t)' | z_t] = B^{-1}E[\epsilon_t \epsilon_t' | z_t](B^{-1})' = B^{-1}\Sigma (B^{-1})',$$

which we denote by $\Omega$ so that:

$$D_{y_t | z_t}(y_t | z_t) = (2\pi)^{-\frac{T}{2}}(\det \Omega^{-1})^\frac{T}{2} \exp\left(-\frac{1}{2}v_t'\Omega^{-1}v_t\right).$$

Because of independence, $D_{y_t | z_t}$ is:

$$(2\pi)^{-\frac{T}{2}}(\det \Omega^{-1})^\frac{T}{2} \exp\left(-\frac{1}{2}v_t'\Omega^{-1}v_t\right) = (2\pi)^{-\frac{T}{2}}(\det \Sigma^{-1})^\frac{T}{2} \det B^T \exp\left(-\frac{1}{2}v_t'\Omega^{-1}v_t\right).$$

But:

$$\sum_{t=1}^{T} v_t'\Omega^{-1}v_t = \sum_{t=1}^{T} \epsilon_t' (B^{-1})'B'\Sigma^{-1}BB^{-1}\epsilon_t = \sum_{t=1}^{T} \epsilon_t'\Sigma^{-1}\epsilon_t$$

$$= \text{tr} \left[ \Sigma^{-1} \sum_{t=1}^{T} \epsilon_t\epsilon_t' \right] = \text{tr} \left[ \Sigma^{-1} A \sum_{t=1}^{T} x_t x_t' A' \right]$$

$$= \text{tr} \left[ \Sigma^{-1} AX'XA' \right],$$

so that:

$$D_{y_t | z_t}(y_t | z_t) = (2\pi)^{-\frac{T}{2}}(\det \Sigma^{-1})^\frac{T}{2} \det B^T \exp\left(-\frac{1}{2}\text{tr} \left[ \Sigma^{-1} AX'XA' \right]\right).$$

Hence, the log-likelihood is:

$$\ell(y | z) = -\frac{T}{2} \log 2\pi + \frac{T}{2} \log \det \Sigma^{-1} + T \log \det B - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} AX'XA' \right].$$

To find the score, notice that for any non-singular matrix $F$ and for any matrices $G, H, K$ and $M$:

$$\frac{\partial \log \det F^{-1}}{\partial (F^{-1})^u} = (F^v)'$$

$$\frac{\partial \text{tr} [F^{-1}G]}{\partial (F^{-1})^u} = (G^v)'$$

$$\text{tr} [G'H] = G^v'H^v$$

$$\text{tr} [G'H'KM] = G^v'(H \otimes M')K^v.$$
Hence (with suitable accounting for the symmetry – see Hendry, 1995a, p.634):
\[
\frac{\partial \ell}{\partial (\Sigma^{-1} \nu)} = \frac{T}{2} \Sigma^{-1} - \frac{1}{2} (AX'XA')^{-\nu},
\]
so that equating to zero:
\[
\hat{\Sigma} = T^{-1} \hat{A}X'\hat{A}'.
\tag{11.9}
\]
To differentiate with respect to \(B\) and \(C\), write:
\[
\text{tr} [\Sigma^{-1} AX'XA'] = \text{tr} \left[ \Sigma^{-1} (B : C) \left( \begin{array}{c} Y' \\ Z' \end{array} \right) \left( \begin{array}{c} Y \\ Z \end{array} \right) \left( \begin{array}{c} B' \\ C' \end{array} \right) \right] = \text{tr} \left[ \Sigma^{-1} (BY'YB' + BY'ZC' + CZ'YB' + CZ'ZC') \right].
\]
So ignoring terms which vanish on differentiation:
\[
\frac{\partial \ell}{\partial (B')^{-\nu}} = T \frac{\partial \log |\det B'|}{\partial (B')^{-\nu}} - \frac{1}{2} \text{tr} \left[ \Sigma^{-1} (BY'YB' + BY'ZC' + CZ'YB') \right] - \frac{1}{2} \text{tr} \left( \Sigma^{-1} (BY'YB')B' + 2 (Y'ZC'\Sigma^{-1})^{-\nu} B' \right)
\]
\[
= T (B^{-1})^{-\nu} - \frac{1}{2} \frac{\partial (B')^{-\nu}}{\partial (B')^{-\nu}} (Y'\otimes \Sigma^{-1}) B'^{\nu} + \frac{1}{2} (Y'ZC'\Sigma^{-1})^{-\nu} B'^{\nu}
\]
\[
= T (B^{-1})^{-\nu} - (Y'\otimes \Sigma^{-1}) B'^{\nu} - (Y'ZC'\Sigma^{-1})^{-\nu}
\]
\[
= T (B^{-1})^{-\nu} - (Y'\otimes \Sigma^{-1})^{-\nu}.
\]
A more convenient expression for this derivative is obtained by using (11.9) to write:
\[
TB^{-1} = B^{-1}AX'XA'\Sigma^{-1}
\]
\[
= B^{-1}(B : C) \left( \begin{array}{c} Y' \\ Z' \end{array} \right) XA'\Sigma^{-1}
\]
\[
= (I_n : -\Pi) \left( \begin{array}{c} Y' \\ Z' \end{array} \right) XA'\Sigma^{-1}
\]
\[
= (Y' - \Pi Z) XA'\Sigma^{-1},
\]
and so:
\[
\frac{\partial \ell}{\partial (B')^{-\nu}} = \left( [Y' - \Pi Z] XA'\Sigma^{-1} \right)^{-\nu} - \left( Y'XA'\Sigma^{-1} \right)^{-\nu}
\]
\[
= -(\Pi Z'XA'\Sigma^{-1} - \nu).
\]
Finally:
\[
\frac{\partial \ell}{\partial (C')^{-\nu}} = -\frac{1}{2} \frac{\partial (2 \text{tr} [\Sigma^{-1} BY'ZC'] + \text{tr} [\Sigma^{-1} CZ'ZC'])}{\partial (C')^{-\nu}}
\]
\[
= -\frac{1}{2} \frac{\partial \left( 2 (Z'YB'\Sigma^{-1})^{-\nu} C'^{\nu} + (C'^{\nu})' (Z'Z'\otimes \Sigma^{-1}) C'^{\nu} \right)}{\partial (C')^{-\nu}}
\]
\[
= - (Z'YB'\Sigma^{-1})^{-\nu} - (Z'Z'\otimes \Sigma^{-1}) C'^{\nu}
\]
\[
= - (Z'XAXB^{-1} \Sigma^{-1})^{-\nu}.
\]
Hence, the complete score vector $\partial \ell / \partial A^\nu$ is:

$$
\left( \begin{array}{c}
\frac{\partial \ell}{\partial B^\nu} \\
\frac{\partial \ell}{\partial C^\nu}
\end{array} \right) = - \left( (\Pi Z' X A' \Sigma^{-1})^\nu \right) = - (PZ' X A' \Sigma^{-1})^\nu .
$$

Many of the elements in $A$ are restricted a priori, so FIML is obtained by solving $(\Sigma^{-1} AX' ZP')^\nu = 0$ for the $p$ unrestricted $\theta$, which is the expression in (11.7).

11.2.2 Efficiency

The FIML solution to (11.7) requires iterative computation. Alternative, non-iterative, methods replace $P$ and $\Sigma$ by estimates which are not based on FIML of $A$, and that is why (11.7) is known as an estimator generating equation (EGE). Asymptotically, efficiency is retained provided $P$ and $\Sigma$ are consistently estimated as illustrated below in §11.2.3. For single-equation methods, the error variance becomes irrelevant in $(\sigma^{-2} \alpha X' ZP)^\nu = 0$, so methods based on a consistent estimate of $P$ remain single-equation asymptotically efficient.

11.2.3 Alternative estimation methods

A joint solution of (11.7) for $P$, $\Sigma$ and $A$ provides FIML and so consistent and asymptotically efficient estimates of $\theta$.

3SLS can also be obtained from solving (11.7) for particular choices of $\Sigma$ and $P$. $\Sigma$ is computed in terms of the 2SLS estimates of $A$ so that its $(i,j)$th element is:

$$
\tilde{\sigma}_{ij} = T^{-1} \tilde{e}_i \tilde{e}_j,
$$

where $\tilde{e}_i = X\tilde{a}_i$, $\tilde{a}_i$ is the 2SLS estimate of $a_i$, and $a_i$ is the $i$th row of $A$. To see what $P$ is, notice that 3SLS can be derived by minimizing the criterion function:

$$
\text{tr} \left[ \tilde{\Sigma}^{-1} e' Q e \right],
$$

where $Q = Z(Z'Z)^{-1}Z'$, with respect to $\theta$. Thus:

$$
\text{tr} \left[ \tilde{\Sigma}^{-1} e' Q e \right] = \text{tr} \left[ \tilde{\Sigma}^{-1} AX' Q A' \right]
= \text{tr} \left[ A' \tilde{\Sigma}^{-1} AX' Q X \right]
= A^\nu \left( \tilde{\Sigma}^{-1} \otimes X' Q X \right) A^\nu \quad (11.10)
$$

so that:

$$
\frac{\partial \text{tr} \left[ \tilde{\Sigma}^{-1} e' Q e \right]}{\partial A^\nu} = 2 \left( \tilde{\Sigma}^{-1} \otimes X' Q X \right) A^\nu = 2 \left( \tilde{\Sigma}^{-1} A X' Q X \right)^\nu ,
$$
which yields (11.7) with $\Pi$ in $P$ replaced by $\hat{\Pi} = Y'Z(Z'Z)^{-1}$ (the OLS estimator of the system matrix of parameters).

$\Sigma$ is consistent for $\Sigma$ because $\bar{e}_i$ can be written as:

$$\bar{e}_i = X(\bar{a}_i - a_i) + Xa_i,$$

and so:

$$\bar{\sigma}_{ij} = T^{-1} (X(\bar{a}_i - a_i) + Xa_i)' (X(\bar{a}_i - a_i) + Xa_j)$$

$$= (\bar{a}_i - a_i)' T^{-1} X' X (\bar{a}_i - a_i) + (\bar{a}_i - a_i)' T^{-1} X' Xa_j$$

$$+ a_i'T^{-1} X' (\bar{a}_i - a_i) + T^{-1} e_i' e_j,$$

$$\xrightarrow{A.S.} \sigma_{ij}.$$

Also:

$$T^{-1} \Sigma^{-1} AX'QX = T^{-1} \left( \Sigma^{-1} - \Sigma^{-1} \right) AX'Z \left( \hat{\Pi}' - \Pi' : I_k \right)$$

$$+ T^{-1} \Sigma^{-1} AX'Z \left( \hat{\Pi}' - \Pi' : I_k \right)$$

$$+ T^{-1} \left( \Sigma^{-1} - \Sigma^{-1} \right) AX'ZP' + T^{-1} \Sigma^{-1} AX'ZP'$$

$$= T^{-1} \Sigma^{-1} AX'ZP' + o_p(1)$$

so that the score of the likelihood and the criterium function from which 3SLS is derived differ by $o_p(1)$, implying that 3SLS and FIML are asymptotically equivalent.

2SLS can also be obtained from the EGE, as 2SLS of the $i$th equation minimizes $e_i'Qe_i$ with respect to the unrestricted elements in $a_i$. Thus:

$$\frac{\partial}{\partial a_i} e_i'Qe_i = \frac{\partial}{\partial a_i} a_i'X'QXa_i = 2X'QXa_i. \quad (11.11)$$

But, (11.7) with $\Pi$ replaced by $\hat{\Pi}$ can also be written as:

$$(\Sigma^{-1} \otimes X'QX) A^v = (\Sigma^{-1} \otimes X'QX) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

When $\Sigma = \text{diag}(\sigma_{ii})$ then:

$$(\Sigma^{-1} \otimes X'QX) A^v = \begin{pmatrix} \sigma_{11}^{-1}X'QX a_1 \\ \vdots \\ \sigma_{nn}^{-1}X'QX a_n \end{pmatrix} \quad (11.12)$$

which yields (11.11) for all equations in the system.
We now provide a heuristic proof of the limiting distribution of 3SLS. To do so, let us find its expression from (11.10) by writing:

$$A^\nu = s - S\theta,$$

where $s$ is a vector of zeros and ones selecting the left-hand side variable in each equation and $S$ is a selection matrix of zeros and ones picking up the unrestricted parameters in $A$. Thus, (11.10) can be written as:

$$\text{tr} \left[ \Sigma^{-1} \epsilon' Q \epsilon \right] = (s - S\theta)' \left( \Sigma^{-1} \otimes X'QX \right) (s - S\theta),$$

so that:

$$\frac{\partial \text{tr} \left[ \Sigma^{-1} \epsilon' Q \epsilon \right]}{\partial \theta} = -2S' \left( \Sigma^{-1} \otimes X'QX \right) s + 2S' \left( \Sigma^{-1} \otimes X'QX \right) S\theta,$$

and hence the equations to solve are:

$$S' \left( \Sigma^{-1} \otimes X'QX \right) S (\tilde{\theta} - \theta) = S' \left( \Sigma^{-1} \otimes X'QX \right) (s - S\theta).$$

Subtracting $S'(\Sigma^{-1} \otimes X'QX)S\theta$ from both sides:

$$S' \left( \Sigma^{-1} \otimes X'QX \right) S (\tilde{\theta} - \theta) = S' \left( \Sigma^{-1} \otimes X'QX \right) (s - S\theta)$$

which is also:

$$S' \left( \Sigma^{-1} \otimes X'QX \right) S (\tilde{\theta} - \theta) = S' \left( \Sigma^{-1} \otimes X'Z (Z'Z)^{-1} \right) (I_n \otimes Z') (I_n \otimes X) A^\nu.$$

But $\epsilon^\nu = (I_n \otimes X)A^\nu$ so that:

$$S' \left( \Sigma^{-1} \otimes X'QX \right) S (\tilde{\theta} - \theta) = S' \left( \Sigma^{-1} \otimes X'Z (Z'Z)^{-1} \right) (I_n \otimes Z') \epsilon^\nu.$$

and because:

$$T^{-1}V [(I_n \otimes Z') \epsilon^\nu] = \Sigma \otimes \Sigma_{zz},$$

where $\Sigma_{uw} = E[T^{-1}U'W]$, under appropriate conditions:

$$(\Sigma \otimes \Sigma_{zz})^{-\frac{1}{2}} T^{-\frac{1}{2}} (I_n \otimes Z') \epsilon^\nu \overset{D}{\rightarrow} N_{nk}[0, I_{nk}].$$

Also:

$$\tilde{\Sigma}^{-1} \otimes T^{-1}X'Z (T^{-1}Z'Z)^{-1} \overset{A^S}{\rightarrow} \Sigma \otimes \Sigma_{xx} \Sigma_{zz}^{-1},$$

and:

$$S' \left( \Sigma^{-1} \otimes X'QX \right) S \overset{A^S}{\rightarrow} S' \left( \Sigma^{-1} \otimes \Sigma_{xx} \Sigma_{zz}^{-1} \Sigma_{zz} \right) S := V.$$
Hence, by Cramér’s theorem and (11.13) $T^{-1/2}(\hat{\theta} - \theta)$ tends in distribution to:

$$\mathcal{N}_p\left[0, \mathbf{V}^{-1}\mathbf{S}(\Sigma^{-1} \otimes \Sigma_{zz}^{-1}) (\Sigma \otimes \Sigma_{zz}) (\Sigma^{-1} \otimes \Sigma_{zz}^{-1} \Sigma_{xx}) \mathbf{SV}^{-1}\right],$$

or simplifying:

$$T^{-1/2} \left(\hat{\theta} - \theta\right) \overset{D}{\to} \mathcal{N}_p[0, \mathbf{V}^{-1}].$$

A necessary and sufficient condition for $\mathbf{V}$ to be non-singular is that the model is identified: see Sargan (1988, p.76).

### 11.3 2SLS viewed as IV

Consider the system:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$  \hspace{1cm} (11.14)

where $\mathbf{y}' = (y_1, \ldots, y_T)$, $\mathbf{X}' = (x_1, \ldots, x_T)$ but $E[x_t\epsilon_t] \neq 0$:

$$\mathbf{X} = \mathbf{Z}\Pi' + \mathbf{V}$$  \hspace{1cm} (11.15)

where $\mathbf{Z}' = (z_1, \ldots, z_T)$, $E[\mathbf{Z}'\epsilon] = 0$ and $E[z_t\epsilon_t] = 0 \forall t, s$, when $x_t$ and $z_t$ are $k \times 1$ and $n \times 1$ respectively, with $k \leq n$, and rank($\Pi$) = $k$.

1. Show that the two-stage least-squares (2SLS) estimator $\hat{\beta}$ of $\beta$ can be obtained by:

$$\arg\min_\beta \epsilon'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\epsilon$$  \hspace{1cm} (11.16)

2. Derive the limiting distribution of $\hat{\beta}$, stating any assumptions and theorems required.

3. Show that a loss of asymptotic efficiency results if only a subset $\mathbf{Z}_a$ of $\mathbf{Z}$ is used in (11.16).

(Oxford M.Phil., 1982)

#### 11.3.1 2SLS is IV

2SLS was at first defined as OLS applied in two stages: (11.15) is estimated by OLS to obtain a predictor of all current endogenous variables in the right-hand side of (11.14), those current endogenous variables in $\mathbf{X}$ are replaced by their predictors, and (11.14) is estimated equation by equation by OLS. Alternatively, 2SLS can be obtained by minimizing (11.16) because it can be shown to coincide with IV when $\mathbf{Z}$ is the matrix of instruments. Let us show that.

The $i$th equation in (11.15) may be written for $i = 1, \ldots, k$ as:

$$x_i = \mathbf{Z}\pi_i + v_i,$$
where \( \mathbf{x}_i = (x_{1i}, \ldots, x_{Ti})' \), \( \pi_i = (\pi_{1i}, \ldots, \pi_{ni})' \) and \( \mathbf{v}_i = (v_{1i}, \ldots, v_{Ti})' \). So, the OLS estimator of \( \pi_i \) is:

\[ \hat{\pi}_i = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{x}_i, \]

and hence the first stage applied to all equations in (11.15) yields:

\[ \hat{\Pi}' = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X}. \]

Replacing \( \mathbf{X} \) by \( \mathbf{Z} \hat{\Pi}' \) in (11.14) yields:

\[ \mathbf{y} = \mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X}\beta + \epsilon, \]

and so the OLS estimator of \( \beta \) in this latter equation is:

\[ \hat{\beta} = (\mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}, \]

which is known as 2SLS.

To show that the same estimator is obtained by minimizing (11.16), let \( \mathbf{Q} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \) and differentiate the quadratic form in (11.16) with respect to \( \beta \):

\[ \frac{\partial \mathbf{e}' \mathbf{Q} \mathbf{e}}{\partial \beta} = \frac{\partial (\mathbf{y} - \mathbf{X}\beta)' \mathbf{Q} (\mathbf{y} - \mathbf{X}\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{Q}\mathbf{y} + 2\mathbf{X}'\mathbf{Q}\mathbf{X}\beta, \]

so that equating to zero and replacing \( \mathbf{Q} \):

\[ \hat{\beta} = (\mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}, \]

which coincides with \( \hat{\beta} \). So, 2SLS equals IV when the set of instruments consists of all variables in \( \mathbf{Z} \).

11.3.2 Limiting distribution of 2SLS

To derive the limiting distribution of 2SLS, write:

\[ \hat{\beta} = (\mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y} \]
\[ = (\mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'(\mathbf{X}\beta + \epsilon) \]
\[ = \beta + (\mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\epsilon, \]

and consider:

\[ T^{-1}\mathbf{Z}'\epsilon = T^{-1}\sum_{t=1}^{T} \mathbf{z}_t\epsilon_t. \]
So, if $z$ is stationary and ergodic and $\epsilon_t \sim \text{IID}(0, \sigma^2)$, Mann–Wald’s theorem implies:

$$T^{-\frac{1}{2}} \sum_{t=1}^{T} z_t \epsilon_t \xrightarrow{D} N_n \left[ 0, \sigma^2 \Sigma_{zz} \right],$$

and:

$$T^{-1}Z'Z = T^{-1} \sum_{t=1}^{T} z_t z_t' \xrightarrow{AS} \Sigma_{zz},$$

by the ergodic theorem. Assuming $\Sigma_{zz}$ is non-singular, Slutsky’s theorem implies that:

$$(T^{-1}Z'Z)^{-1} \xrightarrow{AS} \Sigma_{zz}^{-1}.$$

In addition:

$$X'Z = (\Pi Z' + V') Z = \Pi Z'Z + V'Z,$$

and if the $\{z_t v_{jt}\}$s are IID:

$$T^{-1}V'Z = T^{-1} \sum_{t=1}^{T} z_t v_t' \xrightarrow{AS} \Sigma_{zv} = 0,$$

because $\Sigma_{zv} = 0$ by assumption. Hence:

$$T^{-1}X'Z \xrightarrow{AS} \Pi \Sigma_{zz}.$$

So, by Slutsky’s theorem:

$$T \left( X'Z (Z'Z)^{-1} Z'X \right)^{-1} \xrightarrow{AS} \left( \Pi \Sigma_{zz} \Pi' \right)^{-1} \Pi,$$

and applying Cramér’s theorem:

$$T^\frac{1}{2} \left( \tilde{\beta} - \beta \right) \xrightarrow{AS} N_k \left[ 0, \sigma^2 \left( \Pi \Sigma_{zz} \Pi' \right)^{-1} \right] := N_k \left[ 0, V \right]. \hspace{1cm} (11.17)$$

### 11.3.3 Optimality of 2SLS

Partition $Z = (Z_a : Z_b)$ to compute IV using $Z_a$ as $n > n_1 \geq k$ instruments. The IVE of $\beta$ is defined as:

$$\tilde{\beta} - \beta = \left( X'Z_a (Z_a'Z_a)^{-1} Z_a'X \right)^{-1} X'Z_a (Z_a'Z_a)^{-1} Z_a' \epsilon.$$  

But:

$$T^{-1}X'Z_a = \Pi T^{-1}Z'Z_a + T^{-1}V'Z_a,$$
and because \( T^{-1}V'Z \overset{AS}{\to} 0 \) then \( T^{-1}V'Z_a \overset{AS}{\to} 0 \). Also:

\[
T^{-1}Z'Z = T^{-1} \begin{pmatrix} Z'_1Z_a & Z'_2Z_b \\ Z'_1Z_a & Z'_2Z_b \end{pmatrix} \overset{AS}{\to} \Sigma_{zz} := \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix},
\]

where \( \Sigma_{ab} = E[z_a z'_b] \) so that:

\[
T^{-1}X'Z_a \overset{AS}{\to} \Pi \begin{pmatrix} \Sigma_{aa} \\ \Sigma_{ba} \end{pmatrix},
\]

and so:

\[
T \left( X'Z_a (Z'_a Z_a)^{-1} Z'a X \right)^{-1} X'Z_a (Z'_a Z_a)^{-1} \overset{AS}{\to} \Pi \left( \begin{pmatrix} \Sigma_{aa} \\ \Sigma_{ba} \end{pmatrix} \Sigma_{aa}^{-1} (\Sigma_{aa} \cdot \Sigma_{ab}) \Pi' \right)^{-1} \Pi \begin{pmatrix} \Sigma_{aa} \\ \Sigma_{ba} \end{pmatrix} \Sigma_{aa}^{-1} = \Pi \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Sigma_{aa}^{-1} \Sigma_{ab} \Pi'.
\]

Also:

\[
T^{-\frac{1}{2}}Z'_a \epsilon \overset{D}{\to} N_{n_1} \begin{bmatrix} 0, \sigma_z^2 \Sigma_{aa} \end{bmatrix}.
\]

Thus, by Cramér’s theorem, the asymptotic distribution of \( T^{1/2}(\hat{\beta} - \beta) \) is:

\[
N_{n_1} \begin{bmatrix} 0, \sigma_z^2 \Pi \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \Sigma_{aa}^{-1} \Sigma_{ab} \end{pmatrix}^{-1} \Pi' = N_{n_1} \begin{bmatrix} 0, V^* \end{bmatrix}. \tag{11.18}
\]

We next compare the asymptotic variances in (11.17) and (11.18). Since:

\[
V - V^* = \Pi \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} - \Pi \begin{pmatrix} 0 & 0 \\ 0 & \Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab} \end{pmatrix} \Pi' \tag{11.19}
\]

which is positive semi-definite, because \( \Sigma_{bb} - \Sigma_{ba} \Sigma_{aa}^{-1} \Sigma_{ab} \) is positive definite as it is the inverse of element \( (2, 2) \) in the positive-definite matrix \( \Sigma_{zz}^{-1} \). Thus, the asymptotic variance of IV exceeds that of 2SLS by a positive semi-definite matrix, and therefore there is a potential loss in efficiency from excluding the variables in \( Z_b \) from the set of instruments.

**11.4 Normalization in IV**

Consider the structural equation:

\[
y_{1,t} = \beta y_{2,t} + \nu_t \tag{11.20}
\]
where \( \nu_t \sim \text{IN}[0, \sigma^2_t] \), when there are \( k > 1 \) valid instrumental variables \( z_t \) which can be used for an instrumental variables estimator (IVE) \( \hat{\beta} \) of \( \beta \). Alternatively, the same model could have been formulated as:

\[
y_{2,t} = \gamma y_{1,t} + \epsilon_t \tag{11.21}
\]

where \( \epsilon_t \sim \text{IN}[0, \sigma^2_t] \), with the IV estimator \( \hat{\gamma} \) again based on \( z_t \). Prove that \( \hat{\beta} \hat{\gamma} \leq 1 \). Discuss this result, considering whether or not the estimator \( 1/\hat{\gamma} \) of \( \beta \) is consistent. Is any method of estimating \( \beta \) independent of the normalization choice between (11.20) and (11.21)?


Denoting by \( Q = Z(Z'Z)^{-1}Z' \), the IV estimator of \( \beta \) in (11.20) is defined by:

\[
\hat{\beta} = (y'_{2}Qy_{2})^{-1}y'_{2}Qy_{1},
\]

and that of \( \gamma \) in (11.21) is:

\[
\hat{\gamma} = (y'_{1}Qy_{1})^{-1}y'_{1}Qy_{2},
\]

so that:

\[
\hat{\beta} \hat{\gamma} = (y'_{2}Qy_{2})^{-1}(y'_{1}Qy_{1})^{-1}(y'_{1}Qy_{2})^2.
\]

But, \( Z'Z \) is positive definite and so \( Z'Z = HH' \) where \( H \) is non-singular so that:

\[
y_{2}'Qy_{2} = y_{2}'Z(HH')^{-1}Z'y_{2} = y_{2}'Z(H')^{-1}H^{-1}Z'y_{2} = u'v,
\]

where \( u = H^{-1}Z'y_{2} \). Similarly:

\[
y_{1}'Qy_{1} = v'v,
\]

with \( v = H^{-1}Z'y_{1} \) and:

\[
y_{1}'Qy_{2} = v'u.
\]

So, applying the Cauchy–Schwartz inequality:

\[
\hat{\beta} \hat{\gamma} = \frac{(v'u)^2}{(v'v)(u'u)} \leq 1,
\]

showing that the numerical value of IVE changes with the choice of normalization. However, \( \hat{\beta} \hat{\gamma} \rightarrow^A S 1 \) as \( T \rightarrow \infty \) which is to be shown next.

Let us write:

\[
\frac{1}{\hat{\gamma}} = \frac{y'_{1}Qy_{1}}{y'_{1}Qy_{2}} = \frac{y'_{1}Q(y_{2}\beta + \nu)}{y'_{1}Qy_{2}} = \beta + \frac{y'_{1}Q\nu}{y'_{1}Qy_{2}}.
\]

If the \( z \)s are valid instruments in (11.20) then \( T^{-1}Z'\nu \rightarrow^A S 0 \) and so \( \hat{\gamma}^{-1} \rightarrow^A S \beta \). So, normalization has no effect on the consistency of IV.
Estimators robust to normalization can be found. In (11.20) LIML minimizes:

\[
\lambda = \frac{b'_1 W_1 b_1}{b'_1 W b_1},
\]

where \( b'_1 = (-1 : \beta) \), \( W_1 = Y'_1 Y_1 \), \( W = Y'_1 M_2 Y_1 \), \( M_2 = I_T - Q \), \( Y_1 = (y_1 : y_2) \) and \( Z \) is the matrix of all exogenous variables in the system. LIML of (11.21) minimizes:

\[
\lambda^* = \frac{b'_1^* W_1 b_1^*}{b'_1^* W b_1^*},
\]

where \( b'_1^* = (\gamma : -1) \). But, \( \gamma = 1 / \beta \) so that \( b'_1^* = -\gamma(-1 : \beta) = -\gamma b'_1 \). Hence:

\[
\lambda^* = \frac{b'^*_1 W_1 b_1^*}{b'^*_1 W b_1^*} = \frac{\gamma^2 b'^*_1 W_1 b_1^*}{\gamma^2 b'^*_1 W b_1^*} = \lambda
\]

and so LIML minimizes the same function under either normalization.

### 11.5 Simultaneity

In the simultaneous system:

\[
\begin{align*}
y_1 - \beta y_2 &= u_1, \\
y_2 - \gamma z &= u_2,
\end{align*}
\]

where:

\[
\begin{pmatrix}
  u_{1,t} \\
  u_{2,t}
\end{pmatrix}
\sim \text{IN}_2[0, \Sigma],
\]

for \( t = 1, \ldots, T \), \( y_1 \) and \( y_2 \) are \( T \times 1 \) vectors of observations on endogenous variables, \( z \) is a \( T \times 1 \) vector of observations on a variable strongly exogenous for \( (\beta, \gamma) \) and \( u_1 \) and \( u_2 \) are \( T \times 1 \) vectors of errors with an unrestricted covariance matrix \( \Sigma \), which is symmetric and positive definite.

1. Derive the reduced-form equations, and obtain least-squares estimators of their coefficients and error variance matrix.
2. Show that the indirect and two-stage least-squares estimators of \( \beta \) are equivalent.
3. Another estimator of \( \beta \) is obtained by minimizing the determinant of the reduced-form error variance matrix, and solving for \( \beta \) from the resulting estimates of the reduced-form coefficients. What is this estimator? Is it also equivalent to the estimators in \( \S 11.5.1 \)?
4. Briefly describe one other equivalent estimator. Are all of these estimators equivalent when it is known that \( \Sigma \) is diagonal?

(Oxford M.Phil., 1992)
### 11.5.1 OLS of the reduced form

The second equation is in its reduced form so that the reduced form of the system is obtained by replacing $y_2$ in the first equation:

$$ y_1 - \beta \gamma z = u_1 + \beta u_2 = \epsilon_1. $$

Hence, the OLS estimator for the reduced form is:

\[
\hat{\beta} = \left( \sum_{t=1}^{T} z_t^2 \right)^{-1} \sum_{t=1}^{T} z_t y_{1,t}, \\
\hat{\gamma} = \left( \sum_{t=1}^{T} z_t^2 \right)^{-1} \sum_{t=1}^{T} z_t y_{2,t}.
\]

The reduced-form error-variance matrix $\Omega$ can be estimated by:

\[
\hat{\Omega} = \begin{pmatrix}
T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_{1,t}^2 & T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_{1,t} \hat{\epsilon}_{2,t} \\
T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_{1,t} \hat{\epsilon}_{2,t} & T^{-1} \sum_{t=1}^{T} \hat{\epsilon}_{2,t}^2
\end{pmatrix}.
\]

### 11.5.2 Indirect least-squares

Because the system is just identified, indirect least squares uniquely solves $B \hat{\Pi} + \hat{C} = 0$ where:

\[
\hat{B} = \begin{pmatrix}
1 & -\hat{\beta} \\
0 & 1
\end{pmatrix} \quad \text{and} \quad \hat{C} = \begin{pmatrix}
0 \\
-\hat{\gamma}
\end{pmatrix},
\]

so that solving for $\hat{\beta}$ and $\hat{\gamma}$ from:

\[
\begin{pmatrix}
1 & -\hat{\beta} \\
0 & 1
\end{pmatrix} \begin{pmatrix}
\hat{\beta} \\
\hat{\gamma}
\end{pmatrix} + \begin{pmatrix}
0 \\
-\hat{\gamma}
\end{pmatrix} = \begin{pmatrix}
0 \\
0
\end{pmatrix},
\]

or:

\[
\hat{\beta} \hat{\gamma} - \hat{\beta} \hat{\gamma} = 0,
\]

this implies for $\hat{\gamma} \neq 0$:

\[
\hat{\beta} = \frac{\hat{\beta} \hat{\gamma}}{\hat{\gamma}} = \frac{\sum_{t=1}^{T} z_t y_{1,t}}{\sum_{t=1}^{T} z_t y_{2,t}},
\]

which is IV of the first equation using $z$ as instrument, and hence is 2SLS.

### 11.5.3 FIML estimator

FIML estimation of the system parameters are obtained by minimizing $\log \det \Omega$. Because the system is just-identified, these FIML estimates equal OLS.
11.5.4 LI-MLE

LI-MLE, 2SLS and 3SLS are all equal to one another because the system is just-identified. 3SLS is asymptotically equivalent to FIML. So, LI-MLE and 2SLS are also asymptotically equivalent to FIML. If Σ is diagonal the system becomes recursive, and OLS of the simultaneous form is consistent because:

\[ E[y_{2,t}|u_{1,t}] = \gamma E[z_{t}|u_{1,t}] + E[u_{2,t}|u_{1,t}] = 0, \]

since \( z_t \) is strongly exogenous, and correlation across equations is zero. However, OLS of the first equation does not coincide with LI-MLE because OLS minimizes \( u_0'u_1 \) with respect to \( \beta \) whereas LI-MLE minimizes:

\[ \lambda = \frac{b_1'W_1b_1}{b_1'W_1b_1}, \]

where \( b_1' = (1 - \beta) \), \( W_1 = Y_1'Y_1 \), \( W = Y_1'M_zY_1, M_z = I_T - z(z'z)^{-1}z' \) and \( Y_1 = (y_1 : y_2) \), so that:

\[ \lambda = \frac{u_1'b_1}{u_1'M_zu_1} \neq u_1'u_1. \]