Class Notes: Problem Set 5

EC337

Pinjas Albagli

Last updated: March 26, 2025

Contents

1 Some useful properties

1. **Bilinearity and associativity of the Kronecker product:**

   For arbitrary conformable matrices A, B, and C and scalar k,

$$A \otimes (B + C) = A \otimes B + A \otimes C,$$

$$(B + C) \otimes A = B \otimes A + C \otimes A,$$

$$(kA) \otimes B = A \otimes (kB) = k(A \otimes B),$$

$$(A \otimes B) \otimes C = A \otimes (B \otimes C),$$

$$A \otimes 0 = 0 \otimes A = 0.$$

2. **Mixed-product property of the Kronecker product:**

   For arbitrary conformable matrices A, B, C, and D,

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD).$$

3. **Mixed Kronecker matrix-vector product:**

   For arbitrary conformable matrices A, B, and C,

$$\text{vec}(ABC) = \left(C' \otimes A\right) \text{vec}(B).$$

4. **Inverse of a Kronecker product:**

   For arbitrary nonsingular matrices A and B,

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

5. **Transpose of a Kronecker product:**

   For arbitrary matrices A and B,

$$(A \otimes B)' = A' \otimes B'.$$

6. **Determinant of a Kronecker product:**

   For arbitrary $n \times n$ matrix A and $k \times k$ matrix B,

$$|A \otimes B| = |A|^k |B|^n.$$

7. **Trace of a Kronecker product:**

   For arbitrary square matrices A and B,

$$\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B).$$

8. **Vectorization is a unitary transformation (preserves the inner product):**

   For arbitrary $m \times n$ matrices A and B,

$$\operatorname{tr}\left(A'B\right) = \operatorname{vec}(A)' \operatorname{vec}(B).$$

9. **Basic properties of the trace:**

   For arbitrary square matrices A and B and scalar k,

$$\operatorname{tr}(A+B) = \operatorname{tr}(A) + \operatorname{tr}(B),$$

$$\operatorname{tr}(kA) = k\operatorname{tr}(A),$$

$$\operatorname{tr}\left(A'\right) = \operatorname{tr}(A),$$

$$\operatorname{tr}(I_k) = k.$$

10. **The trace is invariant under cyclic permutations (the cyclic property):**

    For arbitrary suitably conformable matrices A, B, C, and D,

$$\operatorname{tr}(ABCD) = \operatorname{tr}(BCDA) = \operatorname{tr}(CDAB) = \operatorname{tr}(DABC).$$

11. **(Some) basic properties of the determinant:**

    For arbitrary $n \times n$ matrices A and B and scalar k,

$$|I_k| = 1,$$

$$|kA| = k^n |A|,$$

$$\left|A'\right| = |A|,$$

$$\left|A^{-1}\right| = |A|^{-1} \text{ (if } |A| \neq 0\text{)},$$

$$|AB| = |A|\,|B|.$$

12. **Determinant of a triangular matrix:**

For arbitrary triangular matrix A with diagonal entries $\{a_{ii}\}_{i=1}^{n}$,

$$|A| = \prod_{i=1}^{n} a_{ii}.$$

Moreover,

$$|A| = 0 \iff \exists\, i \in \{1, \ldots, n\} : a_{ii} = 0,$$

and

$$a_{ii} = 1 \;\forall\, i \in \{1, \ldots, n\} \implies |A| = 1.$$

13. **Basic properties of triangular matrices:**

Let $A, B \in \Delta$ be two arbitrary suitably conformable lower (upper) triangular matrices and k a scalar, where Δ is the set of all lower (upper) triangular matrices and ∇ is the set of all upper (lower) triangular matrices. Then,

$$kA \in \Delta,$$

$$A + B \in \Delta,$$

$$AB \in \Delta,$$

$$A^{-1} \in \Delta \text{ with diagonal } \left\{a_{ii}^{-1}\right\}_{i=1}^{n} \text{ (provided } |A| \neq 0 \text{)},$$

$$A' \in \nabla$$

2 Notation

Consider the following system of N equations in N endogenous variables and K exogenous variables for a sample of T observations $t \in \{1, \ldots, T\}$

$$\tilde{a}_1^1\, y_{1t} + \cdots + \tilde{a}_N^1\, y_{Nt} + \tilde{b}_1^1\, x_{1t} + \cdots + \tilde{b}_K^1\, x_{Kt} = u_{1t} \tag{S.1}$$

$$\vdots$$

$$\tilde{a}_1^N\, y_{1t} + \cdots + \tilde{a}_N^N\, y_{Nt} + \tilde{b}_1^N\, x_{1t} + \cdots + \tilde{b}_K^N\, x_{Kt} = u_{Nt}. \tag{S.N}$$

Let

$$y_t \equiv \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix}_{N\times 1} \quad x_t \equiv \begin{pmatrix} x_{1t} \\ \vdots \\ x_{Kt} \end{pmatrix}_{K\times 1} \quad u_t \equiv \begin{pmatrix} u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix}_{N\times 1}$$

$$\tilde{A} \equiv \begin{bmatrix} \tilde{a}_1^1 & \cdots & \tilde{a}_N^1 \\ \vdots & \ddots & \vdots \\ \tilde{a}_1^N & \cdots & \tilde{a}_N^N \end{bmatrix}_{N\times N} \qquad \tilde{B} \equiv \begin{bmatrix} \tilde{b}_1^1 & \cdots & \tilde{b}_K^1 \\ \vdots & \ddots & \vdots \\ \tilde{b}_1^N & \cdots & \tilde{b}_K^N \end{bmatrix}_{N\times K}$$

Then, we can compactly write the system for observation t as

$$\tilde{A}y_t + \tilde{B}x_t = u_t \tag{1}$$

and, transposing and stacking the system for the T observations, we obtain the matrix equation

$$Y\tilde{A}' + X\tilde{B}' = U, \tag{2}$$

where

$$Y \equiv \begin{bmatrix} y_{11} & \cdots & y_{N1} \\ \vdots & \ddots & \vdots \\ y_{1T} & \cdots & y_{NT} \end{bmatrix}_{T\times N} = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}$$

$$X \equiv \begin{bmatrix} x_{11} & \cdots & x_{K1} \\ \vdots & \ddots & \vdots \\ x_{1T} & \cdots & x_{KT} \end{bmatrix}_{T\times K} = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix}$$

$$U \equiv \begin{bmatrix} u_{11} & \cdots & u_{N1} \\ \vdots & \ddots & \vdots \\ u_{1T} & \cdots & u_{NT} \end{bmatrix}_{T\times N} = \begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix}$$

As discussed in the lecture notes, the system is not identified without further restrictions on the structural parameters

$$\tilde{\delta} \equiv \begin{bmatrix} \text{vec}\,(\tilde{A}) \\ \text{vec}\,(\tilde{B}) \end{bmatrix}_{N(N+K)\times 1}.$$

From now on we will suppose the system is identified through the imposition of

(i) N normalizations $\tilde{a}_i^i = 1$, $i \in \{1, \ldots, N\}$

(ii) At least $N(N-1)$ exclusion restrictions of the form $\tilde{a}_j^i = 0$ or $\tilde{b}_k^i = 0$, $i, j \in \{1, \ldots, N\}$. $i \neq j$, $k \in \{1, \ldots, K\}$.

For $i \in \{1, \ldots, N\}$, define:

- $y_t^{(i)}$: $N_i \times 1$ subvector of y_t, containing $N_i \le N-1$ of the $N-1$ endogenous variables in y_t excluding y_{it} —i.e., the endogenous variable on the LHS of equation $(S.i)$. The endogenous variables y_{jt} (where $j \neq i$) included in $y_t^{(i)}$ are those with $\tilde{a}_j^i \neq 0$.

- $a^{(i)}$: $N_i \times 1$ subvector of

$$a^i \equiv \begin{pmatrix} a_1^i \\ \vdots \\ a_N^i \end{pmatrix}_{N \times 1}$$

  corresponding to the coefficients on the endogenous variables y_{jt} in $y_t^{(i)}$, i.e., those with $a_j^i \neq 0$, where $a_j^i \equiv -\tilde{a}_j^i$ (for $i \neq j$).

- $x_t^{(i)}$: $K_i \times 1$ subvector of x_t, containing $K_i \le K$ of the K exogenous variables in x_t. These are the exogenous variables on the RHS of equation $(S.i)$. The exogenous variables x_{kt} included in $x_t^{(i)}$ are those with $\tilde{b}_k^i \neq 0$.

- $b^{(i)}$: $K_i \times 1$ subvector of

$$b^i \equiv \begin{pmatrix} b_1^i \\ \vdots \\ b_K^i \end{pmatrix}_{K \times 1}$$

  corresponding to the coefficients on the exogenous variables x_{kt} in $x_t^{(i)}$, i.e., those with $b_k^i \neq 0$, where $b_k^i \equiv -\tilde{b}_k^i$.

Then, we can write the system after imposing the normalizations and exclusion restrictions as

$$y_{1t} = y_t^{(1)\prime} a^{(1)} + x_t^{(1)\prime} b^{(1)} + u_{1t} \tag{S.1}$$

$$\vdots$$

$$y_{Nt} = y_t^{(N)\prime} a^{(N)} + x_t^{(N)\prime} b^{(N)} + u_{Nt}. \tag{S.N}$$

Now, let

$$z_t^{(i)} \equiv \begin{pmatrix} y_t^{(i)} \\ x_t^{(i)} \end{pmatrix}_{(N_i+K_i) \times 1}$$

and

$$\delta^{(i)} \equiv \begin{pmatrix} a^{(i)} \\ b^{(i)} \end{pmatrix}_{(N_i+K_i) \times 1}$$

so that equation $(S.i)$ can be written as

$$y_{it} = z_t^{(i)\prime} \delta^{(i)} + u_{it}$$

for $i \in \{1, \ldots, N\}$. Stacking equation $(S.i)$ for the T observations $t \in \{1, \ldots, T\}$, we obtain the matrix form

$$Y_i = Z^{(i)}\delta^{(i)} + U_i,$$

where Y_i and U_i are the i^{th} columns of matrices Y and U, respectively,

$$Z^{(i)} \equiv \begin{bmatrix} Y^{(i)} & X^{(i)} \end{bmatrix}_{T\times(N_i+K_i)}$$

$$Y^{(i)} \equiv \begin{bmatrix} y_1^{(i)\prime} \\ \vdots \\ y_T^{(i)\prime} \end{bmatrix}_{T\times N_i}$$

$$X^{(i)} \equiv \begin{bmatrix} x_1^{(i)\prime} \\ \vdots \\ x_T^{(i)\prime} \end{bmatrix}_{T\times K_i} .$$

3 Question 1

3.1 General derivations

3.1.1 2SLS estimator

First, consider equation $(S.i)$ alone, where $i \in \{1, \ldots, N\}$. The goal here is estimation of $\delta^{(i)}$ and, as stated above, identification is assumed. Since we are interested in the i^{th} column of Y, Y_i, we need excluded instruments for $Y^{(i)}$ —$X^{(i)}$ act as included instruments for themselves.

Solving for Y in equation (2), we obtain the reduced form

$$Y = X\Pi' + V, \tag{3}$$

where

$$\Pi' \equiv -\tilde{B}'\tilde{A}'^{-1}$$

and

$$V \equiv U\tilde{A}'^{-1}.$$

Notice that exogeneity of X implies that

$$\mathbb{E}\left[X'V\right] = \mathbb{E}\left[X'U\right]\tilde{A}'^{-1}$$

$$= 0_{K\times N},$$

so $\widehat{Y} = X\widehat{\Pi}'$ and U —where $\widehat{\Pi}' = (X'X)^{-1}X'Y$ is the OLS estimator of Π' in multivariate regression (3)— are asymptotically uncorrelated since $\widehat{\Pi} \xrightarrow{p} \Pi$. Therefore, for the best linear predictor of $Y^{(i)}$, $\widehat{Y}^{(i)}$, is a valid excluded instrument, and the full matrix of instruments is

$$\widehat{Z}^{(i)} \equiv \begin{bmatrix} \widehat{Y}^{(i)} & X^{(i)} \end{bmatrix}_{T \times (N_i + K_i)}.$$

Note that $\widehat{Y}^{(i)}$ is a submatrix of

$$\begin{aligned} \widehat{Y} &= X\widehat{\Pi}' \\ &= X(X'X)^{-1}X'Y \\ &= X(X'X)^{-1}X' \begin{bmatrix} Y_1 & \cdots & Y_N \end{bmatrix} \\ &= \begin{bmatrix} X(X'X)^{-1}X'Y_1 & \cdots & X(X'X)^{-1}X'Y_N \end{bmatrix}, \end{aligned}$$

formed by the N_i columns corresponding to the variables in $y_t^{(i)}$, i.e.,

$$\widehat{Y}^{(i)} = X(X'X)^{-1}X'Y^{(i)}.$$

Moreover, since $X^{(i)} \in \text{Col}\,(X) \implies X(X'X)^{-1}X'X^{(i)} = X^{(i)}$,

$$\begin{aligned} \widehat{Y}^{(i)\prime}X^{(i)} &= \left(X(X'X)^{-1}X'Y^{(i)}\right)'X^{(i)} \\ &= Y^{(i)\prime}X(X'X)^{-1}X'X^{(i)} \\ &= Y^{(i)\prime}X^{(i)} \end{aligned}$$

and

$$\begin{aligned} \widehat{Y}^{(i)\prime}\widehat{Y}^{(i)} &= \left(X(X'X)^{-1}X'Y^{(i)}\right)'X(X'X)^{-1}X'Y^{(i)} \\ &= Y^{(i)\prime}X(X'X)^{-1}X'X(X'X)^{-1}X'Y^{(i)} \\ &= Y^{(i)\prime}X(X'X)^{-1}X'Y^{(i)} \\ &= \widehat{Y}^{(i)\prime}Y^{(i)} \\ &= Y^{(i)\prime}\widehat{Y}^{(i)}. \end{aligned}$$

Therefore,

$$\widehat{Z}^{(i)\prime}Z^{(i)} = \begin{bmatrix} \widehat{Y}^{(i)\prime} \\ X^{(i)\prime} \end{bmatrix} \begin{bmatrix} Y^{(i)} & X^{(i)} \end{bmatrix}$$

$$= \begin{bmatrix} \widehat{Y}^{(i)\prime} Y^{(i)} & \widehat{Y}^{(i)\prime} X^{(i)} \\ X^{(i)\prime} Y^{(i)} & X^{(i)\prime} X^{(i)} \end{bmatrix}$$

$$= \begin{bmatrix} \widehat{Y}^{(i)\prime} \widehat{Y}^{(i)} & \widehat{Y}^{(i)\prime} X^{(i)} \\ X^{(i)\prime} \widehat{Y}^{(i)} & X^{(i)\prime} X^{(i)} \end{bmatrix}$$

$$= \begin{bmatrix} \widehat{Y}^{(i)\prime} \\ X^{(i)\prime} \end{bmatrix} \begin{bmatrix} \widehat{Y}^{(i)} & X^{(i)} \end{bmatrix}$$

$$= \widehat{Z}^{(i)\prime} \widehat{Z}^{(i)},$$

so the IV estimator is

$$\widehat{\delta}^{(i)}_{\text{2SLS}} = (\widehat{Z}^{(i)\prime} Z^{(i)})^{-1} \widehat{Z}^{(i)\prime} Y_i$$

$$= (\widehat{Z}^{(i)\prime} \widehat{Z}^{(i)})^{-1} \widehat{Z}^{(i)\prime} Y_i,$$

i.e., the OLS estimator of the regression of Y_i on $\widehat{Y}^{(i)}$ and $X^{(i)}$.

Now, consider estimation of the full system. To this end, write the system in matrix form by stacking the N previously derived matrix equations

$$Y_i = Z^{(i)} \delta^{(i)} + U_i$$

for $i \in \{1, \ldots, N\}$ as follows:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_N \end{pmatrix}_{NT \times 1} = \begin{bmatrix} Z^{(1)} & \cdots & 0_{T \times (N_N + K_N)} \\ \vdots & \ddots & \vdots \\ 0_{T \times (N_1 + K_1)} & \cdots & Z^{(N)} \end{bmatrix}_{NT \times \left(\sum_{i=1}^{N} N_i + K_i \right)} \begin{pmatrix} \delta^{(1)} \\ \vdots \\ \delta^{(N)} \end{pmatrix}_{\left(\sum_{i=1}^{N} N_i + K_i \right) \times 1} + \begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix}_{NT \times 1}$$

$$\Longleftrightarrow \qquad Y^* = Z^* \delta + U^*,$$

where Y^*, Z^*, δ, and U^* are defined in the obvious way.[1]

Similarly, let

$$\widehat{Z}^* \equiv \begin{bmatrix} \widehat{Z}^{(1)} & \cdots & 0_{T \times (N_N + K_N)} \\ \vdots & \ddots & \vdots \\ 0_{T \times (N_1 + K_1)} & \cdots & \widehat{Z}^{(N)} \end{bmatrix}_{NT \times \left(\sum_{i=1}^{N} N_i + K_i \right)}$$

[1]Note that $Y^* = \text{vec}\,(Y)$.

$$= \begin{bmatrix} X(X'X)^{-1}X'Z^{(1)} & \cdots & 0_{T\times(N_N+K_N)} \\ \vdots & \ddots & \vdots \\ 0_{T\times(N_1+K_1)} & \cdots & X(X'X)^{-1}X'Z^{(N)} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} X(X'X)^{-1}X' & \cdots & 0_{T\times T} \\ \vdots & \ddots & \vdots \\ 0_{T\times T} & \cdots & X(X'X)^{-1}X' \end{bmatrix}}_{NT\times NT} \underbrace{\begin{bmatrix} Z^{(1)} & \cdots & 0_{T\times(N_N+K_N)} \\ \vdots & \ddots & \vdots \\ 0_{T\times(N_1+K_1)} & \cdots & Z^{(N)} \end{bmatrix}}_{NT\times\left(\sum_{i=1}^{N} N_i+K_i\right)}$$

$$= \Big(I_N \otimes X(X'X)^{-1}X'\Big)Z^*$$

$$= \Big(I_N\, I_N \otimes X\left((X'X)^{-1}X'\right)\Big)Z^*$$

$$= \Big(I_N \otimes X\Big)\Big(I_N \otimes (X'X)^{-1}X'\Big)Z^*$$

$$= \Big(I_N \otimes X\Big)\Big(I_N\, I_N \otimes \left((X'X)^{-1}\right)X'\Big)Z^*$$

$$= \Big(I_N \otimes X\Big)\Big(I_N \otimes (X'X)^{-1}\Big)\Big(I_N \otimes X'\Big)Z^*.$$

Finally, following analogous arguments to those discussed above for the single equation case, the system-2SLS estimator is the OLS estimator of the regression of Y^* on $\widehat{Z}^*$,

$$\widehat{\delta}_{\text{2SLS}} = (\widehat{Z}^{*\prime} Z^*)^{-1}\widehat{Z}^{*\prime} Y^*$$

$$= (\widehat{Z}^{*\prime}\widehat{Z}^*)^{-1}\widehat{Z}^{*\prime} Y^*.$$

3.1.2 3SLS estimator

Now, to obtain the 3SLS estimator —which amounts to a GLS-style transformation—, notice that the assumption that

$$u_t \overset{iid}{\sim} \big(0, \Sigma\big)$$

implies that, for $i, j \in \{1, \dots, N\}$,

$$\mathbb{E}\left[U_i U_j'\right] = \mathbb{E}\left[\begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix}\begin{pmatrix} u_{j1} & \cdots & u_{jT} \end{pmatrix}\right]$$

$$
= \begin{bmatrix} \mathbb{E}\left[u_{i1}u_{j1}\right] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbb{E}\left[u_{iT}u_{jT}\right] \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma_{ij} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{ij} \end{bmatrix}
$$

$$
= \sigma_{ij}\, I_T,
$$

where

$$
\Sigma \equiv \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix}.
$$

Therefore,

$$
\mathbb{E}\left[U^* U^{*\prime}\right] = \mathbb{E}\left[\begin{pmatrix} U_1 \\ \vdots \\ U_N \end{pmatrix} \begin{pmatrix} U_1' & \cdots & U_N' \end{pmatrix}\right]
$$

$$
= \begin{bmatrix} \mathbb{E}\left[U_1 U_1'\right] & \cdots & \mathbb{E}\left[U_1 U_N'\right] \\ \vdots & \ddots & \vdots \\ \mathbb{E}\left[U_N U_1'\right] & \cdots & \mathbb{E}\left[U_N U_N'\right] \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma_{11}\, I_T & \cdots & \sigma_{1N}\, I_T \\ \vdots & \ddots & \vdots \\ \sigma_{N1}\, I_T & \cdots & \sigma_{NN}\, I_T \end{bmatrix}
$$

$$
= \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix} \otimes I_T
$$

$$
= \Sigma \otimes I_T.
$$

Next, applying the GLS-style transformation,

$$
\left(\Sigma^{-1/2} \otimes I_T\right) Y^* = \left(\Sigma^{-1/2} \otimes I_T\right) Z^* \delta + \left(\Sigma^{-1/2} \otimes I_T\right) U^*
$$

with

$$\mathbb{E}\left[\left(\Sigma^{-1/2}\otimes I_T\right)U^*\left(\left(\Sigma^{-1/2}\otimes I_T\right)U^*\right)'\right] = \left(\Sigma^{-1/2}\otimes I_T\right)\mathbb{E}\left[U^*U^{*\prime}\right]\left(\Sigma^{-1/2}\otimes I_T\right)'$$

$$= \left(\Sigma^{-1/2}\otimes I_T\right)\left(\Sigma\otimes I_T\right)\left(\Sigma^{-1/2\prime}\otimes I_T'\right)$$

$$= \left(\Sigma^{-1/2}\,\Sigma\otimes I_T I_T\right)\left(\Sigma^{-1/2}\otimes I_T\right)$$

$$= \left(\Sigma^{-1/2}\,\Sigma\,\Sigma^{-1/2}\otimes I_T I_T I_T\right)$$

$$= \left(I_N\otimes I_T\right)$$

$$= I_{NT},$$

we obtain the 3SLS estimator

$$\widehat{\delta}_{\text{3SLS}} = \left(\left(\left(\Sigma^{-1/2}\otimes I_T\right)\widehat{Z}^*\right)'\left(\Sigma^{-1/2}\otimes I_T\right)\widehat{Z}^*\right)^{-1}\left(\left(\Sigma^{-1/2}\otimes I_T\right)\widehat{Z}^*\right)'\left(\Sigma^{-1/2}\otimes I_T\right)Y^*$$

$$= \left(\widehat{Z}^{*\prime}\left(\Sigma^{-1/2\prime}\otimes I_T{}'\right)\left(\Sigma^{-1/2}\otimes I_T\right)\widehat{Z}^*\right)^{-1}\widehat{Z}^{*\prime}\left(\Sigma^{-1/2\prime}\otimes I_T{}'\right)\left(\Sigma^{-1/2}\otimes I_T\right)Y^*$$

$$= \left(\widehat{Z}^{*\prime}\left(\Sigma^{-1/2}\,\Sigma^{-1/2}\otimes I_T I_T\right)\widehat{Z}^*\right)^{-1}\widehat{Z}^{*\prime}\left(\Sigma^{-1/2}\,\Sigma^{-1/2}\otimes I_T I_T\right)Y^*$$

$$= \left(\widehat{Z}^{*\prime}\left(\Sigma^{-1}\otimes I_T\right)\widehat{Z}^*\right)^{-1}\widehat{Z}^{*\prime}\left(\Sigma^{-1}\otimes I_T\right)Y^*$$

$$= \left(Z^{*\prime}X(X'X)^{-1}X'\left(\Sigma^{-1}\otimes I_T\right)X(X'X)^{-1}X'Z^*\right)^{-1}Z^{*\prime}X(X'X)^{-1}X'\left(\Sigma^{-1}\otimes I_T\right)Y^*$$

$$= \left(Z^{*\prime}\left(1\otimes X(X'X)^{-1}X'\right)\left(\Sigma^{-1}\otimes I_T\right)X(X'X)^{-1}X'Z^*\right)^{-1}Z^{*\prime}\left(1\otimes X(X'X)^{-1}X'\right)\left(\Sigma^{-1}\otimes I_T\right)Y^*$$

$$= \left(Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)X(X'X)^{-1}X'Z^*\right)^{-1}Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Y^*$$

$$= \left(Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)\left(1\otimes X(X'X)^{-1}X'\right)Z^*\right)^{-1}Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Y^*$$

$$= \left(Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'X(X'X)^{-1}X'\right)Z^*\right)^{-1}Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Y^*$$

$$
\begin{aligned}
&= \left(Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Z^{*}\right)^{-1} Z^{*\prime}\left(\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Y^{*} \\
&= \left(Z^{*\prime}\left(I_N\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Z^{*}\right)^{-1} Z^{*\prime}\left(I_N\Sigma^{-1}\otimes X(X'X)^{-1}X'\right)Y^{*} \\
&= \left(Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}\otimes (X'X)^{-1}X'\right)Z^{*}\right)^{-1} Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}\otimes (X'X)^{-1}X'\right)Y^{*} \\
&= \left(Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}I_N\otimes (X'X)^{-1}X'\right)Z^{*}\right)^{-1} Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}I_N\otimes (X'X)^{-1}X'\right)Y^{*} \\
&= \left(Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}\otimes (X'X)^{-1}\right)\left(I_N\otimes X'\right)Z^{*}\right)^{-1} Z^{*\prime}\left(I_N\otimes X\right)\left(\Sigma^{-1}\otimes (X'X)^{-1}\right)\left(I_N\otimes X'\right)Y^{*}.
\end{aligned}
$$

3.2 Exactly identified system

Finally, notice that exact identification requires that $N_i + K_i = K \ \forall i \in \{1, \ldots, N\}$. The necessary order condition for identification of equation $(S.i)$ is that the number of restrictions we impose on the $N+K$ parameters on the LHS (of the first representation discussed above) through the normalization $\tilde{a}_i^i = 1$ and the exclusion restrictions —i.e., $(N - N_i) + (K - K_i)$— is at least N.[2] That is,

$$
N - N_i + K - K_i \geq N \iff K - K_i \geq N_i,
$$

which requires that the number of excluded exogenous variables appearing elsewhere in the system, $K - K_i$, is at least as large as the number of included endogenous variables, N_i. Since the equation is just identified, the order condition holds with equality and we get

$$
K - K_i = N_i \iff N_i + K_i = K.
$$

Therefore, block diagonal matrix

$$
\left(I_N \otimes X'\right)Z^{*} = \underbrace{\begin{bmatrix} X' & \cdots & 0_{K\times T} \\ \vdots & \ddots & \vdots \\ 0_{K\times T} & \cdots & X' \end{bmatrix}}_{NK\times NT} \underbrace{\begin{bmatrix} Z^{(1)} & \cdots & 0_{T\times K} \\ \vdots & \ddots & \vdots \\ 0_{T\times K} & \cdots & Z^{(N)} \end{bmatrix}}_{NT\times NK}
$$

[2]Recall that the corresponding equations obtained from the reduced form —the first row of the full set of equations in matrix form— comprises K linear equations on $N+K$ unknown structural parameters, so we need at least N additional equations.

$$= \underbrace{\begin{bmatrix} X'Z^{(1)} & \cdots & 0_{K\times K} \\ \vdots & \ddots & \vdots \\ 0_{K\times K} & \cdots & X'Z^{(N)} \end{bmatrix}}_{NK\times NK}$$

is a square matrix with square, nonsingular diagonal blocks $X'Z^{(i)}$ and is therefore nonsingular. Thus, in the case of an exactly identified system, the system-2SLS estimator simplifies to

$$\begin{aligned}
\widehat{\delta}_{\text{2SLS}} &= (\widehat{Z}^{*\prime} Z^*)^{-1} \widehat{Z}^{*\prime} Y^* \\
&= \left(\left(\left(I_N \otimes X\right) \left(I_N \otimes (X'X)^{-1}\right) \left(I_N \otimes X'\right) Z^* \right)' Z^* \right)^{-1} \left(\left(I_N \otimes X\right) \left(I_N \otimes (X'X)^{-1}\right) \left(I_N \otimes X'\right) Z^* \right)' Y^* \\
&= \left(Z^{*\prime} \big(I_N \otimes X\big) \big(I_N \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big) Z^* \right)^{-1} Z^{*\prime} \big(I_N \otimes X\big) \big(I_N \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big) Y^* \\
&= \left(\big(I_N \otimes X'\big) Z^* \right)^{-1} \left(I_N \otimes (X'X)^{-1} \right)^{-1} \left(Z^{*\prime} \big(I_N \otimes X\big) \right)^{-1} Z^{*\prime} \big(I_N \otimes X\big) \big(I_N \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big) Y^* \\
&= \left(\big(I_N \otimes X'\big) Z^* \right)^{-1} \left(I_N \otimes (X'X)^{-1} \right)^{-1} \big(I_N \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big) Y^* \\
&= \left(\big(I_N \otimes X'\big) Z^* \right)^{-1} \big(I_N \otimes X'\big) Y^*.
\end{aligned}$$

Simlarly, the 3SLS estimator simplifies to

$$\begin{aligned}
\widehat{\delta}_{\text{3SLS}} &= \left(\big(I_N \otimes X'\big)\, Z^* \right)^{-1} \left(\Sigma^{-1} \otimes (X'X)^{-1} \right)^{-1} \left(Z^{*\prime} \big(I_N \otimes X\big) \right)^{-1} Z^{*\prime} \big(I_N \otimes X\big) \big(\Sigma^{-1} \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big)\, Y^* \\
&= \left(\big(I_N \otimes X'\big)\, Z^* \right)^{-1} \left(\Sigma^{-1} \otimes (X'X)^{-1} \right)^{-1} \big(\Sigma^{-1} \otimes (X'X)^{-1}\big) \big(I_N \otimes X'\big)\, Y^* \\
&= \left(\big(I_N \otimes X'\big)\, Z^* \right)^{-1} \big(I_N \otimes X'\big)\, Y^* \\
&= \widehat{\delta}_{\text{2SLS}},
\end{aligned}$$

which coincides with the system-2SLS estimator for the exactly identified system.

4 Question 2

4.1 General derivations

Consider the MLE of (δ, Σ) under the assumption that

$$u_t \overset{iid}{\sim} N\big(0, \Sigma\big) \implies v_t \overset{iid}{\sim} N\big(0, \Omega\big),$$

where $\Omega = \tilde{A}^{-1}\Sigma\tilde{A}^{-1\prime}$ and $v_t \equiv \tilde{A}^{-1}u_t$ is the projection error in the reduced form

$$y_t = \Pi x_t + v_t$$

with $\Pi \equiv -\tilde{A}^{-1}\tilde{B}$.[3] We can obtain the matrix representation of the reduced form by transposing and stacking the reduced form for the T observations:

$$Y = X\Pi' + V,$$

where

$$V \equiv \begin{pmatrix} v_1' \\ \vdots \\ v_T' \end{pmatrix}_{T\times N} = U\tilde{A}^{-1\prime}.$$

Vectorize V to obtain multivariate-normal vector[4]

$$V^* \equiv \text{vec}\,(V) = \begin{pmatrix} v_{11} \\ \vdots \\ v_{1T} \\ \vdots \\ v_{N1} \\ \vdots \\ v_{NT} \end{pmatrix}_{NT\times 1} \sim N\big(0, \Omega \otimes I_T\big)$$

since

$$\mathbb{E}\left[V^*V^{*\prime}\right] = \mathbb{E}\left[\begin{pmatrix} v_{11} \\ \vdots \\ v_{1T} \\ \vdots \\ v_{N1} \\ \vdots \\ v_{NT} \end{pmatrix}\begin{pmatrix} v_{11} & \cdots & v_{1T} & \cdots & v_{N1} & \cdots & v_{NT} \end{pmatrix}\right]$$

[3] Notice that the variance of v_t, Ω, is a function of the structural parameters in δ and Σ, $\Omega\,(\delta, \Sigma)$.

[4] An alternative is to work directly with the matrix normal distribution.

$$= \underbrace{\begin{bmatrix} \underbrace{\mathbb{E}\left[v_{11}^2\right]}_{\omega_{11}} & \cdots & \underbrace{\mathbb{E}\left[v_{11}v_{1T}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{11}v_{N1}\right]}_{\omega_{1N}} & \cdots & \underbrace{\mathbb{E}\left[v_{11}v_{NT}\right]}_{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \underbrace{\mathbb{E}\left[v_{1T}v_{11}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{1T}^2\right]}_{\omega_{11}} & \cdots & \underbrace{\mathbb{E}\left[v_{1T}v_{N1}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{1T}v_{NT}\right]}_{\omega_{1N}} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \underbrace{\mathbb{E}\left[v_{N1}v_{11}\right]}_{\omega_{N1}} & \cdots & \underbrace{\mathbb{E}\left[v_{N1}v_{1T}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{N1}^2\right]}_{\omega_{NN}} & \cdots & \underbrace{\mathbb{E}\left[v_{N1}v_{NT}\right]}_{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \underbrace{\mathbb{E}\left[v_{NT}v_{11}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{NT}v_{1T}\right]}_{\omega_{N1}} & \cdots & \underbrace{\mathbb{E}\left[v_{NT}v_{N1}\right]}_{0} & \cdots & \underbrace{\mathbb{E}\left[v_{NT}^2\right]}_{\omega_{NN}} \end{bmatrix}}_{NT \times NT}$$

$$= \underbrace{\begin{bmatrix} \omega_{11} & \cdots & \omega_{1N} \\ \vdots & \ddots & \vdots \\ \omega_{N1} & \cdots & \omega_{NN} \end{bmatrix}}_{N \times N} \otimes \underbrace{\begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix}}_{T \times T}$$

$$= \Omega \otimes I_T.$$

Notice that

$$\begin{aligned} V^{*\prime}(\Omega \otimes I_T)^{-1} V^* &= V^{*\prime}\left(\Omega^{-1} \otimes I_T^{-1}\right) V^* \\ &= V^{*\prime}\left(\Omega^{-1\prime} \otimes I_T\right) \operatorname{vec}(V) \\ &= V^{*\prime} \operatorname{vec}\left(I_T\, V\, \Omega^{-1}\right) \\ &= \operatorname{vec}(V)' \operatorname{vec}\left(V\, \Omega^{-1}\right) \\ &= \operatorname{tr}\left(V'\, V\, \Omega^{-1}\right) \\ &= \operatorname{tr}\left(V\, \Omega^{-1}\, V'\right) \\ &= \operatorname{tr}\left(\left(Y - X\Pi'\right) \Omega^{-1} \left(Y - X\Pi'\right)'\right) \end{aligned}$$

and

$$\begin{aligned} |\Omega \otimes I_T| &= |\Omega|^T \, |I_T|^N \\ &= |\Omega|^T \, 1^N \end{aligned}$$

$$= |\Omega|^T.$$

Therefore, the log-likelihood is

$$\ell(\delta,\Sigma) = \ln\left((2\pi)^{-NT/2}|\Omega\otimes I_T|^{-1/2}\exp\left(-\frac{1}{2}V^{*\prime}(\Omega\otimes I_T)^{-1}V^*\right)\right)$$

$$= \ln\left((2\pi)^{-NT/2}|\Omega|^{-T/2}\exp\left(-\frac{1}{2}\operatorname{tr}\left(\left(Y-X\Pi'\right)\Omega^{-1}\left(Y-X\Pi'\right)'\right)\right)\right)$$

$$= -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\ln(|\Omega|) - \frac{1}{2}\operatorname{tr}\left(\left(Y-X\Pi'\right)\Omega^{-1}\left(Y-X\Pi'\right)'\right)$$

and we can minimize the objective function

$$Q(\delta,\Sigma) = \frac{NT}{2}\ln(2\pi) + \frac{T}{2}\ln(|\Omega|) + \frac{1}{2}\operatorname{tr}\left(\left(Y-X\Pi'\right)\Omega^{-1}\left(Y-X\Pi'\right)'\right)$$

since

$$\arg\max_{(\delta,\Sigma)} \ell(\delta,\Sigma) = \arg\min_{(\delta,\Sigma)} \; -\ell(\delta,\Sigma)$$

$$= \arg\min_{(\delta,\Sigma)} Q(\delta,\Sigma).$$

4.2 Recursive system

We define a recursive system as one where $\tilde{A}$ is lower triangular and Σ is diagonal:

$$\tilde{A} = \begin{bmatrix} \tilde{a}_1^1 & 0 & 0 & \cdots & 0 \\ \tilde{a}_1^2 & \tilde{a}_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_1^N & \tilde{a}_2^N & \tilde{a}_3^N & \cdots & \tilde{a}_N^N \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_{11} & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{NN} \end{bmatrix}.$$

The normalizations $\tilde{a}_{ii} = 1$ for $i \in \{1,\ldots.N\}$ further imply that A is unitriangular and so is A^{-1}. Moreover, A^{-1} lower unitriangular implies that $A^{-1\prime}$ is upper unitriangular. Therefore,

$$|\Omega| = \left|A^{-1}\Sigma A^{-1\prime}\right|$$

$$= \underbrace{\left|A^{-1}\right|}_{1}\left|\Sigma\right|\underbrace{\left|A^{-1\prime}\right|}_{1}$$

$$= |\Sigma|$$

$$= \prod_{i=1}^N \sigma_{ii}.$$

Also, notice that

$$Y - X\Pi' = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix} - \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix} \Pi'$$

$$= \begin{bmatrix} y_1' - x_1'\Pi' \\ \vdots \\ y_T' - x_T'\Pi' \end{bmatrix}$$

$$= \begin{bmatrix} (y_1 - \Pi x_1)' \\ \vdots \\ (y_T - \Pi x_T)' \end{bmatrix}$$

and

$$\Omega^{-1} = A' \, \Sigma^{-1} A$$

Therefore,

$$\left(Y - X\Pi'\right)\Omega^{-1}\left(Y - X\Pi'\right)' = \begin{bmatrix} (y_1 - \Pi x_1)' \, \Omega^{-1} \\ \vdots \\ (y_T - \Pi x_T)' \, \Omega^{-1} \end{bmatrix} \begin{bmatrix} (y_1 - \Pi x_1) & \cdots & (y_T - \Pi x_T) \end{bmatrix}$$

$$= \begin{bmatrix} (y_1 - \Pi x_1)' \, \Omega^{-1} (y_1 - \Pi x_1) & \cdots & (y_1 - \Pi x_1)' \, \Omega^{-1} (y_T - \Pi x_T) \\ \vdots & \ddots & \vdots \\ (y_T - \Pi x_T)' \, \Omega^{-1} (y_1 - \Pi x_1) & \cdots & (y_T - \Pi x_T)' \, \Omega^{-1} (y_T - \Pi x_T) \end{bmatrix}$$

and

$$\operatorname{tr}\left(\left(Y - X\Pi'\right)\Omega^{-1}\left(Y - X\Pi'\right)'\right) = \sum_{t=1}^{T} (y_t - \Pi x_t)' \, \Omega^{-1} (y_t - \Pi x_t)$$

$$= \sum_{t=1}^{T} \left(y_t - (-A^{-1}B)x_t\right)' A' \, \Sigma^{-1} A \left(y_t - (-A^{-1}B)x_t\right)$$

$$= \sum_{t=1}^{T} \left(Ay_t + AA^{-1}Bx_t\right)' \, \Sigma^{-1} \left(Ay_t + AA^{-1}Bx_t\right)$$

$$= \sum_{t=1}^{T} (Ay_t + Bx_t)' \, \Sigma^{-1} (Ay_t + Bx_t)$$

$$= \sum_{t=1}^{T} u_t' \Sigma^{-1} u_t$$

$$= \sum_{t=1}^{T} \left(y_{1t} - z_t^{(1)\prime} \delta^{(1)} \quad \cdots \quad y_{Nt} - z_t^{(N)\prime} \delta^{(N)} \right) \begin{bmatrix} \sigma_{11}^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{NN}^{-1} \end{bmatrix} \begin{pmatrix} y_{1t} - z_t^{(1)\prime} \delta^{(1)} \\ \vdots \\ y_{Nt} - z_t^{(N)\prime} \delta^{(N)} \end{pmatrix}$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}}.$$

These results imply that we can write the objective function as

$$Q(\delta, \Sigma) = \frac{NT}{2} \ln(2\pi) + \frac{T}{2} \ln \left(\prod_{i=1}^{N} \sigma_{ii} \right) + \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}}$$

$$= \frac{NT}{2} \ln(2\pi) + \frac{T}{2} \sum_{i=1}^{N} \ln(\sigma_{ii}) + \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}}$$

$$= \sum_{i=1}^{N} \left\{ \frac{T}{2} \ln(2\pi) + \frac{T}{2} \ln(\sigma_{ii}) + \frac{1}{2} \sum_{t=1}^{T} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}} \right\}.$$

Finally, notice that the MLE is

$$\left(\widehat{\delta}, \widehat{\Sigma} \right) = \underset{(\delta, \Sigma)}{\arg\min}\; Q(\delta, \Sigma)$$

$$= \underset{(\delta, \Sigma)}{\arg\min} \sum_{i=1}^{N} \left\{ \frac{T}{2} \ln(2\pi) + \frac{T}{2} \ln(\sigma_{ii}) + \frac{1}{2} \sum_{t=1}^{T} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}} \right\},$$

which can be obtained by minimizing

$$\frac{T}{2} \ln(2\pi) + \frac{T}{2} \ln(\sigma_{ii}) + \frac{1}{2} \sum_{t=1}^{T} \frac{\left(y_{it} - z_t^{(i)\prime} \delta^{(i)} \right)^2}{\sigma_{ii}}$$

for each $i \in \{1, \ldots, N\}$, i.e., OLS equation by equation (MLE based on the marginal distribution of u_{it}).