1 Setup

1.1 Observations notation

$N =$ number of observations, $n = 1, ..., N$
$R =$ number of equations, $r = 1, ..., R$
$K_r =$ number of regressors in equation r, $k_r = 1, ..., K_r$
$K =$ total number of regressors in the system, $K = \sum_{r=1}^{R} K_r$
$H =$ number of strata (provinces in our sample), $h = 1, ..., H$
$Q_h =$ number of clusters in stratum h, $q_h = 1, ..., Q_h$
$B_{qh} =$ number of observations in cluster q_h in stratum h, $b_{qh} = 1, ..., B_{qh}$
$N_h =$ number of observations in stratum h, $N_h = \sum_{q=1}^{Q_h} B_{qh}$.

Note that $N = \sum_{h=1}^{H} N_h = \sum_{h=1}^{H} \sum_{q=1}^{Q_h} B_{qh}$.

Index n identifies a unique observation, as does index b_{qh} for a given h and q. The latter index also identifies the cluster and stratum, while the former index does not. We'll use index n for the formula for parameter estimates, as these do not depend on clusters; we'll use index b_{qh} for the formula for the estimate of parameters' variance-covariance matrix that depends on clusters.

We will use the following matrix operators:
$A \otimes D$ is a Kronecker product of matrices A and D
$A \circ D$ is a Hadamard product of matrices A and D

1.2 Matrix notation

$1_{m \times n}$ is an $m \times n$ matrix of ones
I_N is an $N \times N$ identity matrix
$I_{RN \times R}$ is identity matrix I_R stacked vertically N times: $I_{RN \times R} = 1_{N \times 1} \otimes I_R$
$I_{R \times RN}$ is identity matrix I_R stacked horizontally N times: $I_{R \times RN} = 1_{1 \times N} \otimes I_R$
$0_{m \times n}$ is an $m \times n$ matrix of zeros
$Z_{RN \times N}$ is an $RN \times N$ block-diagonal matrix of zeros and ones. Each of its columns consists of N column vectors of size $R \times 1$, stacked vertically. For column $i \in \{1, \ldots, N\}$, its i^{th} component vector is $1_{R \times 1}$, and all its other component vectors are $0_{R \times 1}$:

$$Z_{RN \times N} = \begin{bmatrix} 1_{R\times1} & 0_{R\times1} & \dots & 0_{R\times1} \\ 0_{R\times1} & 1_{R\times1} & \dots & 0_{R\times1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{R\times1} & 0_{R\times1} & \dots & 1_{R\times1} \end{bmatrix}. \tag{1.1}$$

$Z_{Q_h K \times RN_h}$ is a $Q_h K \times RN_h$ block-diagonal matrix of zeros and ones:

$$Z_{Q_h K \times RN_h} = \begin{bmatrix} 1_{K\times RB_{1h}} & 0_{K\times RB_{2h}} & \dots & 0_{K\times RB_{Q_h h}} \\ 0_{K\times RB_{1h}} & 1_{K\times RB_{2h}} & \dots & 1_{K\times RB_{Q_h h}} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{K\times RB_{1h}} & 0_{K\times RB_{2h}} & \dots & 0_{K\times RB_{Q_h h}} \end{bmatrix}. \tag{1.2}$$

It consists of $(Q_h)^2$ blocks. Each block has K rows. Hence, the total number of rows is $Q_h K$. Blocks in the first vertical stack have RB_{1h} columns, blocks in the second vertical stack have RB_{2h} columns, and so on. Hence, the total number of columns is $R\left(\sum_{q=1}^{Q_h} B_{qh}\right) = RN_h$. All diagonal blocks are matrices of ones, and all off-diagonal blocks are matrices of zeroes.

1.3 System setup

All equations in the system are of the form $y_{nr} = x_{nr}\delta_r + e_{nr}$, where e_{nr} is the error term. (Please note that this notation does not match the notation in our model write-up). Note that each element in the equation is indexed by the equation id, r, as well as observation id, n. y_{nr} refers to the dependent (left-hand side) variables in the linear equations in the auxiliary model, and x_{nr} refers to explanatory (right-hand side) variables in the linear equations in the auxiliary model. y_{nr} and e_{nr} are scalars, x_{nr} is a $1 \times K_r$ row vector, and δ is a $K_r \times 1$ column vector. Note that each x_{nr} includes a constant. Generally, components of x_{nr} will vary across equations: $x_{ni} \neq x_{nj}$ for $i \neq j$, $i, j \in \{1, ..., R\}$.

Recall that for each observation $n \in \left\{1, \ldots, N\right\}$ there is a unique b_{qh} identifier, where observation n is the b_{qh}-th observation in cluster q_h in stratum h. So specification $y_{nr} = x_{nr}\delta_r + e_{nr}$ can be equivalently written as $y_{b_{qh}r} = x_{b_{qh}r}\delta_r + e_{b_{qh}r}$, with full equality between a_{nr} and $a_{b_{qh}r}$ for $a \in \left\{y, x, e\right\}$. For this equivalence to continue to hold, it is important to always preserve the ordering of observations in the whole sample, in each cluster, of clusters within each stratum, of strata, and the ordering of equations.

1.4 Stacking up equations for a given observation

Stack up all R equations for the n^{th} observation: $y_n = x_n\delta + e_n$, where y_n and e_n are $R \times 1$ column vectors, x_n is an $R \times K$ matrix, and δ is a $K \times 1$ column vector. Now each element is indexed only by the observation id n, as each element contains all equations for that observation. For example,

$$\delta = \begin{bmatrix} \delta_1 & \leftarrow & K_1 \times 1 \\ \delta_2 & \leftarrow & K_2 \times 1 \\ \vdots & \ddots & \vdots \\ \delta_R & \leftarrow & K_R \times 1 \end{bmatrix}.$$

Note that the $R \times K$ matrix x_n is block diagonal, with each block x_{nr} of the size $1 \times K_r$:

$$x_n = \begin{bmatrix} x_{n1} & 0_{1\times K_2} & 0_{1\times K_3} & \cdots & 0_{1\times K_{R-1}} & 0_{1\times K_R} \\ 0_{1\times K_1} & x_{n2} & 0_{1\times K_3} & \cdots & 0_{1\times K_{R-1}} & 0_{1\times K_R} \\ 0_{1\times K_1} & 0_{1\times K_2} & x_{n3} & \cdots & 0_{1\times K_{R-1}} & 0_{1\times K_R} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{1\times K_1} & 0_{1\times K_2} & 0_{1\times K_3} & \cdots & x_{n,R-1} & 0_{1\times K_R} \\ 0_{1\times K_1} & 0_{1\times K_2} & 0_{1\times K_3} & \cdots & 0_{1\times K_{R-1}} & x_{nR} \end{bmatrix}.$$

Note that, once again, there is full equivalence between using index n and index b_{qh}: for a given observation n and its corresponding index b_{qh}, $a_n = a_{b_{qh}}$ for $a \in \left\{y, x, e\right\}$.

This matrix notation describing the system of R equations for each observation n, $y_n = x_n\delta + e_n$, when expanded, looks as follows:

$$\underbrace{\begin{bmatrix} y_{n1} \\ y_{n2} \\ \vdots \\ y_{nR} \end{bmatrix}}_{=y_n} = \underbrace{\begin{bmatrix} x_{n1} & 0_{1\times K_2} & \cdots & 0_{1\times K_R} \\ 0_{1\times K_1} & x_{n2} & \cdots & 0_{1\times K_R} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{1\times K_1} & 0_{1\times K_2} & \cdots & x_{nR} \end{bmatrix}}_{=x_n} \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_R \end{bmatrix}}_{=\delta} + \underbrace{\begin{bmatrix} e_{n1} \\ e_{n2} \\ \vdots \\ e_{nR} \end{bmatrix}}_{=e_n}. \tag{1.3}$$

1.5 Stacking up equation systems over all observations

Next, stack up each n^{th} vector/matrix over all $N = \sum_{h=1}^{H}\sum_{q=1}^{Q_h} B_{qh}$ observations: matrix equation $y = x\delta + e$ expresses the whole system of R equations for N observations. y and e are $RN \times 1$ column vectors, x is an $RN \times K$ matrix, and as before δ is a $K \times 1$ column vector. This matrix notation describing the system of R equations for all N observations, $y = x\delta + e$, when expanded, looks as follows:

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{=y} = \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}}_{=x} \underbrace{\begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_R \end{bmatrix}}_{=\delta} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_N \end{bmatrix}}_{=e},$$

where each n-th component is described in equation (1.3).

1.6 Stacking up equation systems by cluster

Here, we will stack up all B_{qh} observation-specific matrices for observations belonging to the same cluster q_h in stratum h. Let

$$x_{(qh)} = \begin{bmatrix} x_{1_{qh}} \\ x_{2_{qh}} \\ \vdots \\ x_{B_{qh}} \end{bmatrix}_{RB_{qh}\times K},$$

where each $x_{p_{qh}}$ is a block-diagonal $R \times K$ matrix equivalent to x_n for $n = p_{qh}$-th observation in cluster q_h in stratum h. Define similarly $e_{(qh)}$, which is $RB_{qh} \times 1$. Note that $x_{(qh)}$ is quite distinct from x_{nr}: x_{nr} is a $1 \times K_r$ row vector of values of K_r explanatory variable for equation r for observation n; $x_{(qh)}$ is an $RB_{qh} \times K$ matrix of x_{nr} first stacked diagonally for all R equations, and then stacked vertically for all observations n in cluster q_h in stratum h.

1.7 Stacking up equation systems by stratum

Now, stack up all Q_h cluster-specific matrices belonging to the same stratum h:

$$x_{(h)} = \begin{bmatrix} x_{(1h)} \\ x_{(2h)} \\ \vdots \\ x_{(Q_h h)} \end{bmatrix}_{RN_h\times K}.$$

Define similarly $e_{(h)}$, which is $RN_h \times 1$. Note that in subscript notation, () denote cluster- and stratum-level stacking of matrices.

1.8 Preserving the ordering

Once again, for these derivations to have any meaning, it is very important to always preserve the ordering of observations in the whole sample, in each cluster, of clusters within each stratum, of strata, and the ordering of equations. Specifically, these orderings should be preserved when constructing the stacked matrices described above and during all computations described below.

2 Point estimates of coefficients for linear SUR

To estimate this SUR system:

1. Compute equation by equation OLS for each equation $r = 1, \ldots, R$. This produces OLS estimates of δ_r; let's call them $\hat{\delta}_r^{OLS}$. As before, stack up these estimates over all R equations to form a $K \times 1$ column vector $\hat{\delta}^{OLS}$.

2. Using this OLS estimate of δ, construct estimates of error terms: $\hat{e} = y - x\hat{\delta}^{OLS}$.

3. Using these estimates of the error terms, compute the estimate of the $R \times R$ error covariance matrix Ω (it measures the covariance of error terms across equations):

$$\hat{\Omega} = \frac{1}{N} \sum_{n=1}^{N} \hat{e}_n \hat{e}_n', \tag{2.1}$$

Note that $\hat{\Omega}$ is NOT the same as $\frac{1}{N}\hat{e}\hat{e}'$, which is $RN \times RN$. For matrix notation, define

$$\underset{RN\times N}{\hat{\varepsilon}} = \underbrace{(1_{1\times N} \otimes \hat{e})}_{RN\times N} \circ Z_{RN\times N} = \begin{bmatrix} \hat{e}_1 & 0_{R\times 1} & \ldots & 0_{R\times 1} \\ 0_{R\times 1} & \hat{e}_2 & \ldots & 0_{R\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{R\times 1} & 0_{R\times 1} & \ldots & \hat{e}_N \end{bmatrix}, \tag{2.2}$$

where $Z_{RN\times N}$ is defined in equation (1.1), $\hat{e}$ is $RN \times 1$ and each $\hat{e}_n$ is $R \times 1$. Then

$$\hat{\Omega} = \frac{1}{N}\left(I_{R\times RN} \underbrace{(\hat{\varepsilon}\hat{\varepsilon}')}_{RN\times RN} I_{RN\times R} \right). \tag{2.3}$$

4. Now compute the final SUR estimate of δ:

$$\hat{\delta} = \left(\sum_{n=1}^{N} x_n' \hat{\Omega}^{-1} x_n \right)^{-1} \left(\sum_{n=1}^{N} x_n' \hat{\Omega}^{-1} y_n \right), \tag{2.4}$$

or, in matrix notation,

matrix operation more efficient.

$$\hat{\delta} = \left(x' \left[I_N \otimes \hat{\Omega}^{-1} \right] x \right)^{-1} \left(x' \left[I_N \otimes \hat{\Omega}^{-1} \right] y \right). \tag{2.5}$$

This final $\hat{\delta}$ is our SUR estimate of the auxiliary parameter δ. Use this method to compute $\hat{\delta}$ with both 'real' and synthetic datasets.

construct X, y with b_0 → real data
we want to code for an algorithm
~~we guess b_1, we construct~~
M synthetic data sets, then
compare Wald matrix b/t b_0 and b_1.

$\tilde{\beta}$: true structural parameters
b_0: guess of parameters

$\underbrace{\text{1 'real' data} \quad \text{M synthetic data}}_{\text{M+1 error terms.}}$

3 Estimate of the asymptotic covariance matrix for SUR coefficient estimates

3.1 Allowing for clustering

a subset of data where error term comes from the same set of distributions.

Here, we will account for possible clustering of the error terms, and therefore use index b_{qh} to identify a unique observation.

Use this method to compute the covariance matrix for SUR coefficients estimated with 'real' data only.

1. Using the SUR estimate of δ described in equation (2.5) in section (2), construct new, more accurate estimates of error terms: $\check{e} = y - x\hat{\delta}$ (we'll denote these new estimates of e as $\check{e}$ to differentiate them from the estimates obtained with $\hat{\delta}^{OLS}$).

2. Using these new estimates of the error terms, compute the updated estimate of the $R \times R$ error covariance matrix Ω:

$$\check{\Omega} = \frac{1}{N}\sum_{n=1}^{N} \check{e}_n \check{e}_n' = \frac{1}{N}\left(I_{R\times RN} \underbrace{\left(\check{\varepsilon}\check{\varepsilon}'\right)}_{RN\times RN} I_{RN\times R} \right), \tag{3.1}$$

   where $\check{\varepsilon} = \underbrace{\left(1_{1\times N} \otimes \check{e}\right)}_{RN\times N} \circ Z_{RN\times N}$ is $RN \times N$.

3. Define $\check{u}_{b_{qh}} = x'_{b_{qh}} \check{\Omega}^{-1} \check{e}_{b_{qh}}$ [dimensions are, correspondingly, $K \times 1 = \left(K \times R\right)\left(R \times R\right)\left(R \times 1\right)$]. Define the following sum for each $q_h = 1, \ldots, Q_h$ and $h = 1, \ldots, H$ (this sum is over all observations in a given cluster):

$$\tilde{u}_{qh} = \sum_{b_{qh}=1}^{B_{qh}} \check{u}_{b_{qh}} = \sum_{b_{qh}=1}^{B_{qh}} x'_{b_{qh}} \check{\Omega}^{-1} \check{e}_{b_{qh}} = x'_{(qh)} \left[I_{B_{qh}} \otimes \check{\Omega}^{-1}\right] \check{e}_{(qh)}. \tag{3.2}$$

   Define $\tilde{u}_{(h)}$ to be a $Q_h K \times 1$ vector that is created by vertically stacking all Q_h of $\tilde{u}_{qh}$ belonging to stratum h:

$$\tilde{u}_{(h)} = \begin{bmatrix} \tilde{u}_{1h} \\ \tilde{u}_{2h} \\ \vdots \\ \tilde{u}_{Q_h h} \end{bmatrix}.$$

   For matrix notation, define the following $K \times RN_h$ matrix:

$$M = x'_{(h)} \left[I_{N_h} \otimes \check{\Omega}^{-1}\right]. \tag{3.3}$$

   Then

$$\tilde{u}_{(h)} = \left[\left(1_{Q_h \times 1} M\right) \circ Z_{Q_h K \times RN_h}\right] \check{e}_{(h)}, \tag{3.4}$$

   where $Z_{Q_h K \times RN_h}$ is defined in equation (1.2).

4. Average the sum $\tilde{u}_{qh}$ over all clusters in a given stratum h, for each $h = 1, \ldots, H$:

$$\bar{u}_h = \frac{1}{Q_h}\sum_{q=1}^{Q_h} \tilde{u}_{qh} = \frac{1}{Q_h}\left(x'_{(h)} \left[I_{N_h} \otimes \check{\Omega}^{-1}\right] \check{e}_{(h)}\right) = \frac{1}{Q_h}\left(M \check{e}_{(h)}\right). \tag{3.5}$$

   Note that $\bar{u}_h$ is $K \times 1$, while $\tilde{u}_{(h)}$ computed in the previous step is $Q_h K \times 1$.

5. Compute
$$\hat{G} \equiv \frac{N-1}{N-K}\sum_{h=1}^{H}\frac{Q_h}{Q_h-1}\sum_{q=1}^{Q_h}\left(\tilde{u}_{qh}-\bar{u}_h\right)\left(\tilde{u}_{qh}-\bar{u}_h\right)', \tag{3.6}$$
which is a $K \times K$ matrix. For matrix notation, define
$$u_{(h)} = \tilde{u}_{(h)} - 1_{Q_h \times 1} \otimes \bar{u}_h, \tag{3.7}$$
which is $Q_h K \times 1$, and
$$U_h = \underbrace{\left(1_{1\times Q_h} \otimes u_{(h)}\right)}_{Q_h K \times Q_h} \circ Z_{Q_h K \times Q_h} = \begin{bmatrix} \tilde{u}_{1h}-\bar{u}_h & 0_{K\times 1} & \dots & 0_{K\times 1} \\ 0_{K\times 1} & \tilde{u}_{2h}-\bar{u}_h & \dots & 0_{K\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{K\times 1} & 0_{K\times 1} & \dots & \tilde{u}_{Q_h h}-\bar{u}_h \end{bmatrix}, \tag{3.8}$$
which is $Q_h K \times Q_h$. Then
$$\hat{G} \equiv \frac{N-1}{N-K}\sum_{h=1}^{H}\frac{Q_h}{Q_h-1}\left(I_{K\times Q_h K}\underbrace{\left(U_h U_h'\right)}_{Q_h K \times Q_h K} I_{Q_h K \times K}\right). \tag{3.9}$$

6. Compute
$$\hat{D} = \left(\sum_{n=1}^{N} x_n' \check{\Omega}^{-1} x_n\right)^{-1} = \left(x'\left[I_N \otimes \check{\Omega}^{-1}\right]x\right)^{-1}. \tag{3.10}$$

7. The estimate of the covariance matrix for SUR coefficient estimates that accounts for clustering is equal to $\hat{D}\hat{G}\hat{D}$.

3.2 Robust covariance without clustering

Here, we will account for possible heteroskedasticity of error terms but assume no clustering of the error terms. We will therefore use index n to identify a unique observation.

Use this method to compute the covariance matrix for SUR coefficients estimated with synthetic data.

1. Compute
$$\widehat{GG} = \frac{N}{N-K}\sum_{n=1}^{N} x_n' \check{\Omega}^{-1} \check{e}_n \check{e}_n' \check{\Omega}^{-1} x_n, \tag{3.11}$$
which is a $K \times K$ matrix. In matrix notation,

e check

$$\widehat{GG} = \frac{N}{N-K}\left(\overbrace{x'\left[I_N \otimes \check{\Omega}^{-1}\right]}^{\text{temp2}}\left(\check{\varepsilon}\check{\varepsilon}'\right)\left[I_N \otimes \check{\Omega}^{-1}\right]' x\right), \tag{3.12}$$
where $\check{\varepsilon} = \underbrace{\left(1_{1\times N} \otimes \check{e}\right)}_{RN \times N} \circ Z_{RN \times N}$ is $RN \times N$ (same one that is used in equation (3.1)).

2. The estimate of the robust covariance matrix for SUR coefficient estimates without clustering is equal to $\hat{D}\widehat{GG}\hat{D}$.

4 Notes on computation

1. Note that the term $x_n'\check{\Omega}^{-1}$ is present in in both $\hat{D}$ and $\widehat{GG}$.
2. For matrix notation, pre-compute and then reuse:
   (a) $I_N \otimes \hat{\Omega}^{-1}$
   (b) $x' \left[I_N \otimes \hat{\Omega}^{-1} \right]$
   (c) $I_N \otimes \check{\Omega}^{-1}$
   (d) $x' \left[I_N \otimes \check{\Omega}^{-1} \right]$
   (e) M
   (f) Note that $\check{\varepsilon}$ in equations (3.1) and (3.12) is the same