Lecture 11

Generalized method of moments

Lectures in SDPE: *Econometrics I* on February 29, 2024

Markus Jäntti
Swedish Institute For Social Research
Stockholm University

- Consider the generic regression equation

$$Y = \boldsymbol{X}'\beta + e. \tag{1}$$

- The least-squares based approaches to estimation discussed hitherto rely on defining and using k moment conditions, as in

$$\mathrm{E}[\boldsymbol{X}e] = \mathbf{0} \tag{2}$$

to estimate k unknown parameters.

Introduction

- The parameters are *estimated* either by:
  - constructing the error-sum-of-squares function; the parameters are estimated by finding its minimum

$$S_n(\beta) = \sum_{i=1}^{n} (Y_i - \boldsymbol{X}_i\beta)^2 = (\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta) \tag{3}$$

    This can be thought of either as the sample analog of the expected value of the (quadratic) loss function or as the way to solve n equations involving k unknowns (when $n > k$).
  - constructing the *sample analog* of the *population moment conditions*; the parameters are estimated by solving the k moment conditions:

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i e_i = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i (Y_i - \boldsymbol{X}_i'\beta) = \frac{1}{n}\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\beta) = \boldsymbol{0} \tag{4}$$

    The method-of-moments (MM) estimator uses exactly k equations to solve k unknowns.
- In this lecture, we discuss MM estimation when the number of moment conditions we have, l, exceeds the number of parameters k, called *generalized method-of-moments* (GMM) estimation.

- Consider the linear regression

$$Y = \boldsymbol{X}'\beta + e = \boldsymbol{X}_1'\beta_1 + \boldsymbol{X}_2'\beta_2 + e,\ \mathrm{E}[\boldsymbol{X}e] = 0, \tag{5}$$

  where $\boldsymbol{X}_1, \beta_1$ are $k \times 1$ and $\boldsymbol{X}_2, \beta_2$ are $r \times 1$ ($k + r = l$).

- Suppose $\beta_2 = \mathbf{0}$, in which case we have

$$Y = \boldsymbol{X}_1'\beta_1 + e,\ \mathrm{E}[\boldsymbol{X}e] = 0, \tag{6}$$

- We now have k unknown parameters to estimate but we are equipped with $l > k$ moment conditions. The regression is *overidentified* with r *overidentifying restrictions*.

- A *moment condition model* (aka. *estimating equations*) takes moments $\boldsymbol{g}(Y, \boldsymbol{Z}, \boldsymbol{X}, \beta)$, $l \times 1$ functions of a $k \times 1$ ($k \leq l$) parameter vector whose expectation at the true parameter β_0 is

$$\mathrm{E}[\boldsymbol{g}(Y, \boldsymbol{Z}, \boldsymbol{X}, \beta_0)] = \mathbf{0}. \tag{7}$$

- The example in eq. 6 has $\boldsymbol{g}(Y, \boldsymbol{X}, \boldsymbol{X}_1, \beta_1) = \boldsymbol{X}(Y - \boldsymbol{X}_1' \beta_1)$ (so $\boldsymbol{X}$ is in place of $\boldsymbol{Z}$ and $\boldsymbol{X}_1$ is $\boldsymbol{X}$).

- In general for linear regression equations, we have

$$\begin{aligned} \mathrm{E}[\boldsymbol{g}(Y,\boldsymbol{Z},\boldsymbol{X},\beta_0)] &= \mathrm{E}[\boldsymbol{Z}(Y-\boldsymbol{X}'\beta)] \\ &= \mathrm{E}[\boldsymbol{Z}Y] - \mathrm{E}[\boldsymbol{Z}\boldsymbol{X}']\beta = 0. \end{aligned} \tag{8}$$

- If $l = k$, β can be solved directly and exactly (the method of moments estimator)
- If $l > k$, we have more conditions than parameters and we need a trick. Take $l \times l$ matrix $\boldsymbol{W} > 0$ and multiply from the left with $\mathrm{E}[\boldsymbol{X}\boldsymbol{Z}']\boldsymbol{W}$

$$\mathrm{E}[\boldsymbol{X}\boldsymbol{Z}']\boldsymbol{W}\mathrm{E}[\boldsymbol{Z}Y] = \mathrm{E}[\boldsymbol{X}\boldsymbol{Z}']\boldsymbol{W}\mathrm{E}[\boldsymbol{Z}\boldsymbol{X}']\beta \tag{9}$$

- The coefficients are “defined” by

$$\beta = (\mathrm{E}[\boldsymbol{X}\boldsymbol{Z}']\boldsymbol{W}\mathrm{E}[\boldsymbol{Z}\boldsymbol{X}'])^{-1}\mathrm{E}[\boldsymbol{X}\boldsymbol{Z}']\boldsymbol{W}\mathrm{E}[\boldsymbol{Z}Y] \tag{10}$$

1. iid observations (to be relaxed towards the end)
2. $\mathrm{E}[Y^2] < \infty$
3. $\mathrm{E}[||\boldsymbol{X}||^2] < \infty$
4. $\mathrm{E}[||\boldsymbol{Z}||^2] < \infty$
5. $\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'} = \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']$ is positive definite
6. $\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'} = \mathrm{E}[\boldsymbol{Z}\boldsymbol{X}']$ has full rank k
7. $\mathrm{E}[\boldsymbol{Z}e] = 0$
8. $\mathrm{E}[Y^4] < \infty$, $\mathrm{E}[||\boldsymbol{X}||^4] < \infty$, and $\mathrm{E}[||\boldsymbol{Z}||^4] < \infty$

GMM Estimator

We show three different ways of arriving at the GMM estimator, all of which have their uses:

- As the “plugin”-estimator using equation 10
- Solving for β from the sample analog of equation 7 (or 8)
- Explicitly minizing the squared sum of errors

- First, the GMM estimator is formulated as the plug-in estimator of the moment condition model
- The coefficients is estimated by plugging in the sample analogs for the expectations and a feasible $\boldsymbol{W}_n$:

$$\begin{aligned}\widehat{\beta}_{GMM} =& \left(\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i'\right)\boldsymbol{W}_n\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{Z}_i\boldsymbol{X}_i'\right)\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{Z}_i\right)\boldsymbol{W}_n\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Z}_iY_i\right) \\ =&(\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{Y}\end{aligned} \tag{11}$$

- An alternative formulation solves β from the sample analog of the LHS of equation 7:

$$\begin{aligned}\mathbf{0} = \overline{\boldsymbol{g}}_n(\beta_1) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i(\beta_1) = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i(Y_i - \boldsymbol{X}'_{1,i}\beta_1) \\ &= \frac{1}{n}(\boldsymbol{Z}'\boldsymbol{Y} - \boldsymbol{Z}'\boldsymbol{X}\beta_1).\end{aligned} \tag{12}$$

- When $l > k$, this requires the use of a "help" matrix $\boldsymbol{W}$ to allow for solving the β.

- We can also formulate it as the solution to a quadratic minimization problem of the sum of squared residuals (which can be useful for purposes of testing):

$$\widehat{\beta}_{\text{GMM}} = \arg\min_{\beta} J_n(\beta) = \arg\min_{\beta} n(\overline{\boldsymbol{g}}(\beta))' \boldsymbol{W}_n (\overline{\boldsymbol{g}}(\beta)) \tag{13}$$

  where $\boldsymbol{W}_n$ is a $l \times l$ weight matrix.

- The FOC are

$$\begin{aligned} \mathbf{0} =& \frac{\partial}{\partial \beta} J_n(\widehat{\beta}) = 2 \frac{\partial}{\partial \beta} \overline{\boldsymbol{g}}(\widehat{\beta})' \boldsymbol{W}_n \overline{\boldsymbol{g}}(\widehat{\beta}) \\ =& -2\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'(\boldsymbol{Y}-\boldsymbol{X}\widehat{\beta})\right) \end{aligned} \tag{14}$$

- Solving for β yields

$$\widehat{\beta}_{\text{GMM}} = (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{Y} \tag{15}$$

Consistency of GMM Estimator

- To examine the properties of the GMM estimator, we substitute in $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$ to get

$$\begin{aligned}
\widehat{\beta}_{GMM} =& (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'(\boldsymbol{X}\beta + \boldsymbol{e}) \\
=& (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X}\beta + (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{e} \\
=& \beta + (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{e} \Rightarrow \\
\widehat{\beta}_{GMM} - \beta =& (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{e}
\end{aligned} \tag{16}$$

- By the assumptions made, the sample averages (suitably divided by ns) on the RHS of the last line converge to their population expectations and the product of the limits is zero, so $\widehat{\beta}_{GMM}$ is consistent:

$$\begin{aligned}
\widehat{\beta}_{GMM} - \beta =& (\underbrace{n^{-1}\boldsymbol{X}'\boldsymbol{Z}}_{\rightarrow \boldsymbol{Q}'_{ZX'}}\ \underbrace{\boldsymbol{W}_n}_{\rightarrow \boldsymbol{W}}\ \underbrace{n^{-1}\boldsymbol{Z}'\boldsymbol{X}}_{\rightarrow \boldsymbol{Q}_{ZX'}})^{-1}\ \underbrace{n^{-1}\boldsymbol{X}'\boldsymbol{Z}}_{\rightarrow \boldsymbol{Q}'_{ZX'}}\ \underbrace{\boldsymbol{W}_n}_{\rightarrow \boldsymbol{W}}\ \underbrace{n^{-1}\boldsymbol{Z}'\boldsymbol{e}}_{\boldsymbol{0}} \\
=& (\boldsymbol{Q}'_{ZX'}\boldsymbol{W}\boldsymbol{Q}_{ZX'})^{-1}\boldsymbol{Q}'_{ZX'}\boldsymbol{W}\boldsymbol{0} \\
=& \boldsymbol{0}
\end{aligned} \tag{17}$$

- For the asymptotic distribution, multiply both sides of equation 16 by $\sqrt{n}$:

$$\sqrt{n}(\widehat{\beta}-\beta)=\left(\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{X}}{n}\right)^{-1}\frac{\boldsymbol{X}'\boldsymbol{Z}}{n}\boldsymbol{W}_n\frac{\boldsymbol{Z}'\boldsymbol{e}}{\sqrt{n}} \tag{18}$$

- Assuming $\boldsymbol{W}_n\xrightarrow{p}\boldsymbol{W}$, $\boldsymbol{Q}_{\boldsymbol{ZX}'}=\mathrm{E}[\boldsymbol{ZX}']$ and $\boldsymbol{\Omega}=\mathrm{E}[\boldsymbol{gg}']=\mathrm{E}[\boldsymbol{ZZ}'e^2]$,

$$\begin{gathered}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\right)\xrightarrow{p}\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{ZX}'}\\ \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)\boldsymbol{W}_n\left(\frac{1}{\sqrt{n}}\boldsymbol{Z}'e\right)\xrightarrow{d}\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}N(\boldsymbol{0},\boldsymbol{\Omega}).\end{gathered} \tag{19}$$

- This can be used to establish that

$$\sqrt{n}(\widehat{\beta}_{\mathrm{GMM}}-\beta)\xrightarrow{d}N(\boldsymbol{0},\boldsymbol{V}_\beta) \tag{20}$$

with $\boldsymbol{V}_\beta=(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{XZ}'})^{-1}(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{ZX}'})(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{ZX}'})^{-1}$.

Distribution of GMM Estimator

- Using $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$ yields the GMM estimator

$$\widehat{\beta}_{\text{GMM}} = (\boldsymbol{X}'\mathbf{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{Z}\boldsymbol{\Omega}^{-1}\boldsymbol{Z}'\boldsymbol{Y} \tag{21}$$

with asymptotic distribution

$$\sqrt{n}(\widehat{\beta}_{\text{GMM}} - \beta)\xrightarrow{d}N(\mathbf{0}, (\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{\Omega}^{-1}\boldsymbol{Q}_{\boldsymbol{ZX}'})^{-1}) \tag{22}$$

- It can be shown that

$$(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{XZ}'})^{-1}(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{ZX}'})(\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{ZX}'})^{-1} - (\boldsymbol{Q}'_{\boldsymbol{ZX}'}\boldsymbol{\Omega}^{-1}\boldsymbol{Q}_{\boldsymbol{ZX}'})^{-1} \tag{23}$$

is positive semi-definite, so using $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$ results in the efficient GMM estimator.

- A consistent estimator of the unknown $\boldsymbol{\Omega}$ (and its inverse) has the same limiting distribution.
- The efficient weight matrix can be estimated using

$$\widehat{\boldsymbol{W}}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{g}}_i\widehat{\boldsymbol{g}}_i' - \overline{\boldsymbol{g}}\overline{\boldsymbol{g}}'\right)^{-1} \tag{24}$$

(see Hansen (2021) for other alternatives)

Estimation of IV/2SLS models as GMM

- The regression equation is

$$Y = \boldsymbol{X}'\beta + e, \ \text{but}\, \mathrm{E}[\boldsymbol{X}e] \neq 0, \tag{25}$$

  but we have a set of valid instruments $\boldsymbol{Z}$ such that $\mathrm{E}[\boldsymbol{Z}e] = 0$.

- As the moment condition

$$\mathrm{E}[\boldsymbol{Z}e] = \mathrm{E}[\boldsymbol{Z}(Y - \boldsymbol{X}'\beta)] = \mathrm{E}[\boldsymbol{g}(Y, \boldsymbol{X}, \boldsymbol{Z}, \beta)] = \boldsymbol{0} \tag{26}$$

  holds, β can be estimated using GMM .

- The GMM estimator using weight matrix $\boldsymbol{W}_n$ is

$$\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{Y}. \tag{27}$$

IV and 2SLS estimators

- When $l = k$, the GMM estimator simplifies to the *instrumental variables (IV)* estimator:

$$\begin{aligned}\widehat{\beta} &= (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{Y} \\ &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{W}_n^{-1}(\boldsymbol{X}'\boldsymbol{Z})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}_n\boldsymbol{Z}'\boldsymbol{Y} \\ &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}\end{aligned} \tag{28}$$

  Note that this does not depend on $\boldsymbol{W}_n$. (IV is a method-of-moments estimator, i.e., not GMM.)

- The *two-stage least squares (2SLS, TSLS)* estimator regresses Y on fitted values of $\boldsymbol{X}$ (2nd stage) from a regression of $\boldsymbol{X}$ on $\boldsymbol{Z}$ first stage:

$$\widehat{\beta} = (\widehat{\boldsymbol{X}}'\widehat{\boldsymbol{X}})^{-1}\widehat{\boldsymbol{X}}'\boldsymbol{Y} = (\boldsymbol{X}'\boldsymbol{P}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{P}\boldsymbol{Y} \tag{29}$$

  where $\widehat{\boldsymbol{X}} = \boldsymbol{P}\boldsymbol{X}$, with $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$.

- The 2SLS estimator is the GMM estimator using $\boldsymbol{W}_n = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}$ as the weight matrix:

$$\widehat{\beta} = (\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{Y} \tag{30}$$

  If $\Omega = \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}'e^2] = \sigma^2\mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']$ (i.e., we have homoscedasticity), this is also the efficient GMM estimator.

The variance of the 2SLS estimator

- The asymptotic variance of the 2SLS estimator is a special case of GMM with $\boldsymbol{W}_n = (\boldsymbol{Z}'\boldsymbol{Z})^{-1} \to \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']^{-1} = \boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'}^{-1} = \boldsymbol{W}$, or

$$
\begin{aligned}
\boldsymbol{V}_{\beta_{2SLS}} =& (\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})^{-1}(\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{W}\boldsymbol{\Omega}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})(\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{W}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})^{-1} \\
=& (\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'}^{-1}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})^{-1}(\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'}^{-1}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})(\boldsymbol{Q}'_{\boldsymbol{Z}\boldsymbol{X}'}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{Z}'}^{-1}\boldsymbol{Q}_{\boldsymbol{Z}\boldsymbol{X}'})^{-1}
\end{aligned} \tag{31}
$$

- This simplifies considerably under homoscedasticity, since in that case $\boldsymbol{\Omega} = \sigma^2\mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}']$, but is inferior to efficient GMM in general.

Estimation of the variance matrix

- To estimate the variance matrix of the estimator's asymptotic distribution, each of the expectations $\boldsymbol{Q}_{\boldsymbol{ZX}'}$, $\boldsymbol{Q}_{\boldsymbol{ZZ}'}$ are replaced by their estimators, i.e., the corresponding sample averages.
- To estimate $\boldsymbol{\Omega}$, let $\widetilde{e}_i = Y_i - \boldsymbol{X}_i' \widetilde{\beta}_{2SLS}$ and $\widetilde{\boldsymbol{g}}_i = \boldsymbol{g}(\widetilde{\beta}_{2SLS}) = \boldsymbol{Z}_i \widetilde{e}_i$ and $\overline{\boldsymbol{g}}_n = \frac{1}{n} \sum_{i=1}^n \widetilde{\boldsymbol{g}}_i$
- As set up, $\mathrm{E}[\boldsymbol{g}(\beta)] = 0$ so this estimator works:

$$\widehat{\boldsymbol{\Omega}} = \frac{1}{n} \sum_{i=1}^n \widetilde{\boldsymbol{g}}_i \widetilde{\boldsymbol{g}}_i' \tag{32}$$

- A robust alternative is

$$\widehat{\boldsymbol{\Omega}^*} = \frac{1}{n} \sum_{i=1}^n (\widetilde{\boldsymbol{g}}_i - \overline{\boldsymbol{g}}_n)(\widetilde{\boldsymbol{g}}_i - \overline{\boldsymbol{g}}_n)'. \tag{33}$$

- Setting $\boldsymbol{W}_n = \widehat{\boldsymbol{\Omega}}^{-1}$ or $\boldsymbol{W}_n = \widehat{\boldsymbol{\Omega}^*}^{-1}$ leads to the efficient GMM estimator.
- Note the use of $\widetilde{\beta}_{2SLS}$ to estimate the residuals. which makes this a two-step asymptotically efficient GMM estimator.

Estimation of the variance matrix with clustered dependence

- Finally, with clustered dependence, estimation of the variance matrix is slighly more involved. In particular, suppose we have for a cluster g the structural equation
$$\boldsymbol{Y}_g = \boldsymbol{X}_g\beta + \boldsymbol{e}_g. \tag{34}$$
- The centered estimator can be written as
$$\widehat{\beta}_{GMM} - \beta = (\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{W}\left(\sum_{g=1}^{G}\boldsymbol{Z}_g\boldsymbol{e}_g\right). \tag{35}$$
- The cluster-robust covariance estimators relies on
$$\widehat{\boldsymbol{S}} = \sum_{g=1}^{G}\boldsymbol{Z}_g'\widehat{\boldsymbol{e}}_g\widehat{\boldsymbol{e}}_g'\boldsymbol{Z}_g \tag{36}$$
with clustered residuals
$$\widehat{\boldsymbol{e}}_g = \boldsymbol{Y}_g - \boldsymbol{X}_g\widehat{\beta}_{GMM}. \tag{37}$$

Over-Identification Test

- Recall eq. 5 in which $\mathrm{E}[\boldsymbol{X}_1 e] = \mathbf{0}$ *and* $\mathrm{E}[\boldsymbol{X}_2 e] = \mathbf{0}$.
- If $\beta_2 = \mathbf{0}$ we have 6, $Y = \boldsymbol{X}_1'\beta_1 + e$ so

$$\mathrm{E}[\boldsymbol{X}_1(Y - \boldsymbol{X}_1'\beta_1)] = \mathbf{0} \text{ and } \mathrm{E}[\boldsymbol{X}_2(Y - \boldsymbol{X}_1'\beta_1)] = \mathbf{0}. \tag{38}$$

- But if $\beta_2 \neq \mathbf{0}$, both moment conditions in eq. 38 can not hold, so $\mathrm{E}[\boldsymbol{g}] \neq \mathbf{0}$.
- As $J_n(\widehat{\beta})$ (cf eq 13) is a quadratic form in $\overline{\boldsymbol{g}}$, a natural test statistic for the overidentifying restrictions being valid is

$$J_n = n\overline{\boldsymbol{g}}'\boldsymbol{W}_n\overline{\boldsymbol{g}} \xrightarrow{d} \chi^2_{l-k} \tag{39}$$

Hypothesis Testing: The Distance Statistic

- Suppose there is a set of restrictions $\boldsymbol{h} : \mathbb{R}^k \longrightarrow \mathbb{R}^r$ so

$$\mathbb{H}_0 : \boldsymbol{h}(\beta) = \mathbf{0}. \tag{40}$$

- The unrestricted and restricted estimators are

$$\begin{aligned} \widehat{\beta} &= \arg\min_{\beta} J_n(\beta) \\ \widetilde{\beta} &= \arg\min_{\boldsymbol{h}(\beta)=0} J_n(\beta) \end{aligned} \tag{41}$$

  where $J()$ is the value of the criterion function (cf eq 13) in the restricted and unrestricted cases

- It turns out that under $\mathbb{H}_0$, the *distance statistic*

$$D_n = J_n(\widetilde{\beta}) - J_n(\widehat{\beta}) \xrightarrow{d} \chi_r^2 \tag{42}$$

  can be used to test $\mathbb{H}_0$ against $\mathbb{H}_1 : \boldsymbol{h}(\beta) \neq \mathbf{0}$.

GMM: The General Case

- In general, with k unknown parameters in β and $l \geq k$ moment conditions in $\boldsymbol{g}$ (which may be non-linear)

$$\mathrm{E}[\boldsymbol{g}(\beta)] = \mathbf{0}, \tag{43}$$

and a sample estimate $\overline{\boldsymbol{g}}$, the GMM estimator is

$$\widehat{\beta}_{\mathrm{GMM}} = \arg\min_{\beta} J_n(\beta) = \arg\min_{\beta} n(\overline{\boldsymbol{g}}(\beta))' \boldsymbol{W}_n (\overline{\boldsymbol{g}}(\beta)) \tag{44}$$

- Setting $\boldsymbol{W}_n$ to $(n^{-1}\sum_{i=1}^{n} \widehat{\boldsymbol{g}}_i \widehat{\boldsymbol{g}}_i' - \overline{\boldsymbol{g}\boldsymbol{g}}')^{-1}$, and letting $\boldsymbol{G} = \mathrm{E}[(\partial \boldsymbol{g}/\partial \beta')]$ results in

$$\sqrt{n}(\widehat{\beta}_{\mathrm{GMM}} - \beta) \xrightarrow{d} N(\mathbf{0}, (\boldsymbol{G}'\boldsymbol{\Omega}^{-1}\boldsymbol{G})^{-1}) \tag{45}$$

- Replacing $\boldsymbol{G}$ and $\boldsymbol{\Omega}$ with their sample analogs results in the same limiting distribution.

References

Hansen, Bruce E (2021). *Econometrics*. Madison, WI: University of Wisconsin.