Generalized Method of Moments
Introductory Notes

Laura Magazzini

Sant'Anna School of Advanced Studies
laura.magazzini@santannapisa.it

The OLS estimator

$y_i = x_i'\beta + u_i;\ i = 1, \ldots, N$

- The OLS estimator is unbiased and consistent if all variables in x_i are uncorrelated with the error term

$$cov(regressors, error\ term) = 0$$

or

$$E(x_i u_i) = 0$$

- For simplicity, let us assume that x_i only includes two regressors:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + u_i$$

so we have:

$$E(x_{i1} u_i) = 0$$
$$E(x_{i2} u_i) = 0$$

- An orthogonality conditions on the mixed moments of x_i and u_i

The OLS estimator

$E(x_{i1}u_i) = 0$ and $E(x_{i2}u_i) = 0$

- We can also write this conditions as:

$$E[x_{i1}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2)] = 0$$
$$E[x_{i2}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2)] = 0$$

- The OLS estimator can be obtained as the solution of the system of two equations where the *population moment conditions* above are replaced by the corresponding *sample moments*

$$\frac{1}{N}\sum_{i=1}^{N} x_{i1}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2) = 0$$
$$\frac{1}{N}\sum_{i=1}^{N} x_{i2}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2) = 0$$

  - System of two equations in two unknowns (β_1, β_2)
  - These two equations are equivalent to the first order conditions derived from the minimization of the sum of squared residuals
  - The solution is the OLS estimator $\hat{\beta}_1$ and $\hat{\beta}_2$

The IV estimator

- Now suppose that $E(x_{i1}u_i) \neq 0$, while we retain the assumption $E(x_{i2}u_i) = 0$
- The OLS estimator is no longer consistent
- To get a consistent estimator, we search for an instrument (or instrumental variable)
- Suppose z_i is available that is correlated with the endogenous regressor, x_{1i} but not with the error term u_i
- The following orthogonality conditions (population moments) can therefore be written:

$$E[z_i(y_i - x_{i1}\beta_1 - x_{i2}\beta_2)] = 0$$
$$E[x_{i2}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2)] = 0$$

The IV estimator

- The IV estimator can be obtained as a solution to the system of equation that replaces the population moment equations above with sample analogues:

$$\frac{1}{N}\sum_{i=1}^{N} z_i(y_i - x_{i1}\beta_1 - x_{i2}\beta_2) = 0$$
$$\frac{1}{N}\sum_{i=1}^{N} x_{i2}(y_i - x_{i1}\beta_1 - x_{i2}\beta_2) = 0$$

  - ▷ System of two equations in two unknowns (β_1, β_2)
  - ▷ The solution to the system is the IV estimator of β_1 and β_2

- The general solution can be written as

$$\begin{aligned}\hat{\beta}_{IV} &= (Z'X)^{-1}Z'y \\ &= \left(\frac{1}{N}\sum_i z_i x_i'\right)^{-1}\left(\frac{1}{N}\sum_i z_i y_i\right) \\ &= S_{XZ}^{-1}S_{ZY}\end{aligned}$$

The GMM estimator

The general case

- The model of interest is

$$y_i = \mathrm{x}_i'\beta + u_i$$

  with β of dimension $K \times 1$

- Let us assume that there are R instruments available in the vector z_i, which correspond to the following R moment conditions (in the population)

$$E(\mathrm{z}_i u_i) = 0$$

  or

$$E[\mathrm{z}_i(y_i - \mathrm{x}_i'\beta)] = 0$$

- Let us now consider the R corresponding sample moments

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N}[\mathrm{z}_i(y_i - \mathrm{x}_i'\beta)] = 0$$

Identification

System of R equations in K unknowns

- If $R < K$ the model is *not* identified
  - We need additional moment conditions!
- If $R = K$, the model is *exactly identified*
  - We are back to the previous situation, and the IV estimator can be obtained from the sample moment conditions by solving the system of K equations in K unknowns $g_N(\beta) = 0$
- If $R > K$ the model is *overidentified*
  - In this case it is not possible to solve for an estimate of β by solving the system defined by the sample moments $g_N(\beta) = 0$
  - The reason is that there are more equations than unknowns
  - Instead of dropping instruments one can choose β in such a way that the R sample moments $(g_N(\beta) = 0)$ are as close as possible to zero

If $R > K$...

- We choose β in such a way that the R sample moments ($g_N(\beta) = 0$) are as close as possible to zero
- This is done by minimizing the following quadratic form

$$\begin{aligned} \min_{\beta} Q_N(\beta, W_N) &= g_N(\beta)' W_N g_N(\beta) \\ &= \left(\frac{1}{N} \sum_{i=1}^{N} [\mathrm{z}_i (y_i - \mathrm{x}_i' \beta)] \right)' W_N \left(\frac{1}{N} \sum_{i=1}^{N} [\mathrm{z}_i (y_i - \mathrm{x}_i' \beta)] \right) \end{aligned}$$

where W_N is a positive definite symmetric matrix, known as *weighting matrix*, such that

$$W_N \xrightarrow{p} W$$

where W is itself a positive definite symmetric matrix

- To find the min we take first derivative wrt β and equate to 0
- In the case of linear moment conditions a closed form solution can be found:

$$\hat{\beta}_{GMM} = (X'ZW_NZ'X)^{-1}X'ZW_NZ'y$$

  - The solution exists if the matrix $Z'X$ is of rank K
  - In general, the GMM estimator, $\hat{\beta}_{GMM}$, is a function of the weighting matrix, W_N

→ Details

Properties of $\hat{\beta}_{GMM}$

It can be proven that

(a) The GMM estimator converges in probability to the true parameter vector β

$$\hat{\beta}_{GMM} \xrightarrow{p} \beta$$

(b) The GMM estimator is asymptotically distributed as

$$\sqrt{N}(\hat{\beta}_{GMM} - \beta) \xrightarrow{d} N(0, V_{GMM}(W))$$

with

$$V_{GMM}(W) = (\Sigma_{XZ} W \Sigma_{ZX})^{-1} \Sigma_{XZ} W \Sigma W \Sigma_{ZX} (\Sigma_{XZ} W \Sigma_{ZX})^{-1}$$

- $\Sigma = E(u_i^2 z_i z_i')$ is the covariance matrix of the sample moments
- Σ_{XZ} the probability limit of $S_{XZ} = X'Z$

Properties of $\hat{\beta}_{GMM}$

(c) In finite samples, it holds approximately that

$$\hat{\beta}_{GMM} \overset{a}{\sim} N\left[\beta, \frac{(\Sigma_{XZ} W \Sigma_{ZX})^{-1} \Sigma_{XZ} W \Sigma W \Sigma_{ZX} (\Sigma_{XZ} W \Sigma_{ZX})^{-1}}{N}\right]$$

(d) A consistent estimator for the covariance matrix is given by

$$\widehat{avar(\hat{\beta}_{GMM})} = \frac{(S_{XZ} W_N S_{ZX})^{-1} S_{XZ} W_N S W_N S_{ZX} (S_{XZ} W_N S_{ZX})^{-1}}{N}$$

with S consistent estimate of Σ: $S = \sum_i (\hat{u}_i^2 z_i z_i')/N$

- This formula is robust to heteroskedasticity
- In case of heterosk & correlationa different formula can be applied

Efficient choice of W_N

$\hat{\beta}_{GMM} = (X'ZW_NZ'X)^{-1}X'ZW_NZ'y$

- Different weighting matrices lead to different consistent GMM estimators with generally different asymptotic covariance matrices
  - ▷ It only affects efficiency of the estimator, not its consistency
- Recall the asy. var-cov matrix of GMM

$$\text{avar}(\hat{\beta}_{GMM}) = (\Sigma_{XZ}W\Sigma_{ZX})^{-1}\Sigma_{XZ}W\Sigma W\Sigma_{ZX}(\Sigma_{XZ}W\Sigma_{ZX})^{-1}/N$$

- What is the **optimal choice** of W_N?
- It can be shown that the optimal weighting matrix is proportional to the inverse of the covariance matrix of the sample moments

$$\Sigma = E(u_i^2 z_i z_i')$$

- When W_N is chosen optimally, the asy. var-cov matrix simplifies to

$$\text{avar}(\hat{\beta}_{GMM}) = (\Sigma_{XZ}W\Sigma_{ZX})^{-1}/N = (\Sigma_{XZ}\Sigma^{-1}\Sigma_{ZX})^{-1}/N$$

Efficient choice of W_N

- The optimal choice is proportional to $\Sigma = E(u_i^2 z_i z_i')$
  - ▷ In general, this will be a function of β
- A consistent estimator of Σ is

$$S = \frac{1}{N} \sum_i (\hat{u}_i^2 z_i z_i')$$

  with $\hat{u}_i$ estimated on the basis of a consistent $\hat{\beta}$
- The efficient choice therefore is $W_N = S^{-1}$

Efficient choice of W_N

- When W_N is chosen optimally ($W_N = S^{-1}$):

$$\hat{\beta}^*_{GMM} = (X'ZS^{-1}Z'X)^{-1}X'ZS^{-1}Z'y = (S_{XZ}S^{-1}S_{ZX})^{-1}S_{XZ}S^{-1}s_{Zy}$$

- It can be proven that
  (a*) The efficient GMM estimator is consistent
  (b*) $\sqrt{N}(\hat{\beta}^*_{GMM} - \beta) \xrightarrow{d} N[0, (\Sigma_{XZ}W\Sigma_{ZX})^{-1}]$
  (c*) In finite samples, it holds approximately that

$$\hat{\beta}^*_{GMM} \overset{a}{\sim} N\left[\beta, \frac{(\Sigma_{XZ}W\Sigma_{ZX})^{-1}}{N}\right]$$

  (d*) A consistent estimator for the covariance matrix of $\hat{\beta}^*_{GMM}$ is given by

$$\widehat{avar(\hat{\beta}^*_{GMM})} = \frac{(S_{XZ}W_NS_{ZX})^{-1}}{N}$$

One-step and two-step estimation

- The choice of W_N only affects efficiency of the estimator, not its consistency
- Computation of optimal W_N requires an estimate of u_i
- Estimation can proceed in two steps
- In the first step, any W_N can be employed in estimation, e.g. the identity matrix I – To be more general, let us denote this matrix by $W_N^{(1s)}$
- The one-step estimator is obtained by minimizing $Q_N\left(\beta, W_N^{(1s)}\right)$
- The one-step estimator, $\hat{\beta}^{1s}_{GMM}$, is consistent

One-step and two-step estimation

- Two-step estimation
  - Use $\hat{\beta}^{1s}_{GMM}$ to obtain a consistent estimate of $S = \frac{1}{N}\sum_i(\hat{u}_i^2 z_i z_i')$
  - Set (estimate) the optimal weighting matrix to $S^{-1} = W_N^{(2s)}$
  - The two-step estimator is obtained by minimizing $Q_N(\beta, W_N^{(2s)})$
  - The two-step estimator, $\hat{\beta}^{2s}_{GMM}$, is consistent and efficient in the case of non-spherical disturbances
- Under the more general case (heterosk. or correlation), an alternative strategy can be employed
  - Consider one-step estimation (consistent)
  - Use the formula for the variance-covariance matrix of $\hat{\beta}^{1s}_{GMM}$ that is robust to heteroskedasticity and correlation – formula in (d) rather than (d*)

Special case

Spherical disturbances

- Under the assumption of homoskedasticity and uncorrelation of the error term, the first step estimation is also efficient
- In this case, the optimal weighting matrix does not depend on β

$$\Sigma = E(u_i^2 \mathrm{z}_i \mathrm{z}_i') = \sigma_u^2 E(\mathrm{z}_i \mathrm{z}_i')$$

- The optimal weighting matrix is therefore proportional to $E(\mathrm{z}_i \mathrm{z}_i')$
- It can be estimated by $S = \sum_i (\mathrm{z}_i \mathrm{z}_i')/N$
- Therefore, the weighting matrix can be chosen "optimally" in the first stage:

$$W_N = S^{-1}$$

The Sargan/Hansen test

- Consistency of the GMM estimator relies on the validity of moment conditions
  - ▷ We evaluate if the length of the vector $z'u$ is near to 0
- The Sargan/Hansen test allows you to check

$$H_0 : E(z_i u_i) = 0$$

- Let $\hat{u}_i = y_i - x_i' \hat{\beta}^*_{GMM}$
- Consider the test statistics

$$\begin{aligned}
\xi &= N \left[\frac{1}{N} \sum_i z_i \hat{u}_i \right]' \left[\frac{1}{N} \sum_i (\hat{u}_i^2 z_i z_i') \right]^{-1} \left[\frac{1}{N} \sum_i z_i \hat{u}_i \right] \\
&= \left[\sum_i z_i \hat{u}_i \right]' \left[\sum_i (\hat{u}_i^2 z_i z_i') \right]^{-1} \left[\sum_i z_i \hat{u}_i \right]
\end{aligned}$$

The Sargan/Hansen test

- In the case $R = K$, $\xi = 0$
- In this case, GMM reduces to IV: $\hat{\beta}_{GMM} = (z'x)^{-1}z'y$
- The test statistic is therefore given by:

$$z'\hat{u} = z'(y - x\hat{\beta}) = z'(y - x(z'x)^{-1}z'y) = 0$$

- When $R > K$, ξ will in general be positive
- Under $H0$: the instruments are valid, it should be small as

$$p\lim\left(\frac{1}{N}\sum_i z_i\hat{u}_i\right) = 0$$

  - ▷ One can expect all elements in $\left(\frac{1}{N}\sum_i z_i\hat{u}_i\right)$ to be close to zero
- It can be proven that, under $H0$

$$\xi \xrightarrow{d} \chi^2_{R-K}$$

- The associated test is commonly known as the *Sargan-Hansen test* or, alternatively, as the *overidentifying restrictions* test

Appendix

Solution to min Q

- Using matrix notation for convenience we can rewrite our objective function as

$$
\begin{aligned}
Q_N(\beta, W_N) &= \left(\frac{1}{N} \sum_{i=1}^{N} [\mathrm{z}_i (y_i - \mathrm{x}_i' \beta)] \right)' W_N \left(\frac{1}{N} \sum_{i=1}^{N} [\mathrm{z}_i (y_i - \mathrm{x}_i' \beta)] \right) \\
&= \left(\frac{1}{N} Z'(y - X\beta) \right)' W_N \left(\frac{1}{N} Z'(y - X\beta) \right) \\
&= \frac{1}{N^2} \big(y'ZW_N Z'y - y'ZW_N ZX\beta \\
&\qquad - \beta' X' Z W_N Z' y + \beta' X' Z W_N Z' X \beta \big) \\
&= \frac{1}{N^2} \left(y'ZW_N Z'y - 2y'ZW_N ZX\beta + \beta' X' Z W_N Z' X \beta \right)
\end{aligned}
$$

- The solution to the minimization problem gives us the GMM estimator for β

Solution to min Q (cont.d)

$$\frac{\partial Q_N(\beta, W_N)}{\partial \beta} = -2X'ZW_NZ'y + 2X'ZW_NZ'X\hat{\beta}_{GMM} = 0$$

- This in turn imply:

$$X'ZW_NZ'y = X'ZW_NZ'X\hat{\beta}_{GMM}$$

- This is a system with K equations and K unknowns
- Provided that the matrix $Z'X$ is of rank K, the solution is given by

$$\hat{\beta}_{GMM} = (X'ZW_NZ'X)^{-1}X'ZW_NZ'y$$

- In general, the GMM estimator, $\hat{\beta}_{GMM}$, is a function of the weighting matrix, W_N

Back