Lecture 3: Generalized Method of Moments

Chris Conlon

February 16, 2021

NYU Stern

Review Definitions

- Sample average moment conditions: $g_N(\theta) \in \mathbb{R}^q$, where $g_N(\theta)$ is a q-dimensional vector of moment conditions.

$$\mathbb{E}[g(w_i,\theta)] \approx \frac{1}{N}\sum_{i=1}^{N} g(w_i,\theta) \equiv g_N(\theta)$$

- At the truth θ_0: $\mathbb{E}[g(w_i,\theta_0)] = \mathbf{0}_q$
- Choose $\widehat{\theta}_{gmm}$ to minimize

$$Q_N(\theta) = g_N(\theta)' \cdot W_N \cdot g_N(\theta)$$

- Have to choose a weighting matrix W_n.
- Jacobian: $D(\theta) \equiv \mathbb{E}[\frac{\partial g(w_i,\theta)}{\partial \theta}]$, which is a $q \times k$ matrix.
- Evaluated at the optimum, $\frac{1}{\sqrt{N}}\sum_{i=1}^{N} g(w_i,\theta_0) \xrightarrow{d} N(0,S)$ where $S = E[g(w_i,\theta_0)g(w_i,\theta_0)']$ is a $q \times q$ matrix.

GMM: Linear IV

For the linear IV problem this becomes:

$$
\begin{aligned}
g_N(\theta)' W_N g_N(\theta) &= \frac{1}{N^2} \cdot (\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\beta))' W_N (\mathbf{Z}'(\mathbf{Y} - \mathbf{X}\beta)) \\
&= \frac{1}{N^2} \cdot [Y'ZW_N Z'Y - 2\beta X'ZW_N Z'Y + \beta' X'ZW_N Z'X\beta]
\end{aligned}
$$

We can ignore the $\frac{1}{N^2}$ and take the first-order condition:

$$
\begin{aligned}
2X'ZW_N Z'Y &= 2X'ZW_N Z'X\beta \\
\hat{\beta}_{GMM} &= (X'ZW_N Z'X)^{-1} X'ZW_N Z'Y
\end{aligned}
$$

Hopefully this looks familiar

GMM: OLS

- Suppose that we do not have any excluded instruments so that $Z = X$ (and thus $q = k$).
- Also suppose that $W_N = \mathbf{I}_q$ (the identity matrix).
- Then we can see that:

$$
\begin{aligned}
\hat{\beta}_{GMM} &= (X'X\mathbf{I}_q X'X)^{-1} X'X\mathbf{I}_q X'Y \\
&= (X'XX'X)^{-1} X'XX'Y \\
&= (X'X)^{-1}(X'X)^{-1}(X'X)X'Y = (X'X)^{-1}X'Y = \hat{\beta}_{OLS}
\end{aligned}
$$

- In other words, OLS is a special case of the GMM estimator.
- Also, the identification condition $D = \frac{\partial g(w_i,\theta)}{\partial \theta} = \frac{1}{N}\sum_{i=1}^{N} z_i' x_i = \frac{1}{N}\sum_{i=1}^{N} x_i' x_i$ becomes that $rank(X'X) = k$ the well-known OLS rank condition.

2SLS Estimator

Suppose that we do have excluded instruments so that $dim(Z) = q > dim(X) = k$ and that $W_N = (Z'Z)^{-1}$. It immediately follows that:

$$\hat{\beta}_{GMM} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y = \hat{\beta}_{2SLS}$$

If $dim(Z) = q = dim(X) = k$ then $(X'Z)$ is square (and invertible). This expression further simplifies:

$$(X'Z(Z'Z)^{-1}Z'X)^{-1} = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}$$
$$\rightarrow \hat{\beta}_{GMM} = (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'Y = (Z'X)^{-1}Z'Y = \hat{\beta}_{IV}$$

Efficient GMM

An important question remains how one should choose the weighting matrix W_N. We've already seen two options:

1. The identity matrix $\mathbf{I}_q$ equally penalizes violations of all q moments
2. the TSLS weighting matrix $(Z'Z)^{-1}$ which can be thought about as the inverse of the covariance of the instruments.
3. The choice of weighting matrix only matters in the overidentified case $q > k$. Why?

We are interested in efficient GMM which is the GMM estimator with the lowest variance.

Efficient GMM

In order to find the W_N which minimizes the variance of $\hat{\theta}_{GMM}$ we recall the asymptotic variance of the GMM estimator:

$$V_\theta = (DWD')^{-1}(DWSW'D')(DWD')^{-1}$$

It turns out that the best choice of $W_N = S^{-1}$ (which sets *filling* = *bread*). This is easy to see, because W_N is positive semi-definite.

$$(DS^{-1}D')^{-1}(DS^{-1}SS^{-1'}D')(DS^{-1}D')^{-1} = (DS^{-1}D')^{-1}(DS^{-1}D')(DS^{-1}D')^{-1} = (DS^{-1}D')^{-1}$$

Efficient GMM: Discussion

This gives us some insight into what we are looking for from moment conditions.

- We want S to be small (we want the sampling variation/noise of our moments to be as small as possible).
- We also want D (the Jacobian of the moments) to be large.
    - This means that small violations in moment conditions lead to large changes in the objective function.
    - In practical terms, the problem is well identified when the objective function is steep around θ_0.
    - When the problem becomes flat, it becomes hard to distinguish one θ in favor of another.
- The problem is that $S = E[g(w_i, \theta_0)g(w_i, \theta_0)']$ is not something that we readily observe from our data. In fact, the asymptotic covariance evaluated at θ_0 is infeasible.

Efficient GMM: Feasible Weight Matrix

The best we can hope for is to use some sample analogue $W_N = \hat{S}^{-1}$ in its place. One way to compute that is the covariance of the moments estimated at some $\hat{\theta}$ for an initial guess of W:

$$\hat{W} = \hat{S}^{-1} = \left(\frac{1}{N}\sum_{i=1}^{N}(g(w_i,\hat{\theta}) - g_N(\hat{\theta}))\,(g(w_i,\hat{\theta}) - g_N(\hat{\theta}))'\right)^{-1}$$

Because $E[g(w_i,\theta_0)] = 0$ at θ_0 there is a tendency to use $\left(\frac{1}{N}\sum_{i=1}^{N} g(w_i,\hat{\theta})\, g(w_i,\hat{\theta})'\right)^{-1}$ (without de-meaning the moments). In theory this would work fine, but in practice it is nearly always a bad idea.

Efficient GMM: Review

The overall procedure works as follows:

1. Pick some initial weighting matrix W_0: often $\mathbf{I}_q$ or $(Z'Z)^{-1}$.
2. Solve $\hat{\theta} = \arg\min_\theta g_N(\theta)' W_0 g_N(\theta)$.
3. Update $\hat{W} = \left(\frac{1}{N}\sum_{i=1}^{N}(g(w_i,\hat{\theta}) - g_N(\hat{\theta}))\,(g(w_i,\hat{\theta}) - g_N(\hat{\theta}))'\right)^{-1}$
4. Solve $\hat{\theta}_{GMM} = \arg\min_\theta\, g_N(\theta)'\,\hat{W}\,g_N(\theta)$.
5. Compute $D(\hat{\theta}_{GMM})$ and $S(\hat{\theta}_{GMM})$ and compute standard errors.

Estimating the Variance Matrix

For the linear IV estimator when i is independent then $g(w_i, \theta) = z_i \epsilon_i$ and $E[z_i \epsilon_i] = 0$

$$\hat{S} = \frac{1}{N} \sum_{i=1}^{N} z_i z_i' \epsilon_i^2$$

- When there is homoskedastic variance $E[\epsilon_i^2 | z_i] = \sigma^2$ and the covariance of the moments becomes $\frac{\sigma^2}{N} \sum_{i=1}^{N} z_i z_i'$.
- Because scaling weighting matrix by a constant has no effect on the maximum this is equivalent to the 2SLS weight matrix: $\sum_{i=1}^{N} z_i z_i'$ or $Z'Z$
  - 2SLS is only the efficient estimator under homoskedasticity.
  - Likewise, if all regressors are exogenous then $X = Z$ and we are left with the GMM formula coincides with the covariance for heteroskedasticity robust standard errors.
  - Similarly, when appropriate we can consider extensions such as clustered standard errors which are robust to weaker forms of independence.

Estimating the Variance Matrix

As a practical matter, we should always use the sandwich form when calculating the GMM standard errors, rather than the simpler bread version which is only correct at θ_0 under asymptotic optimality conditions.

Example: Gravity Equation

- An important set of models in international trade talk about Gravity Equations
- They are called gravity because trade declines with distance (or distance^2).

$$T_{ij} = \alpha_0 Y_i^{\alpha_1} Y_j^{\alpha_2} D_{ij}^{\alpha_3} \eta_{ij}$$

Take Logs

$$\ln T_{ij} = \ln \alpha_0 + \alpha_1 \ln Y_i + \alpha_2 \ln Y_j + \alpha_3 \ln D_{ij} + \ln \eta_{ij}$$

- T_{ij} (Exports from i to j)
- (Y_i, Y_j) GDP of each country
- D_{ij} distance between two countries

Example: Gravity Equation

If the moment condition holds then everything is good:

$$E[ln(\eta_{ij})|Y_i, Y_j D_{ij}] = 0$$

Some problems

- Lots of Zeros in T_{ij} (so we can't take logs).
- If $(Y_i, Y_j D_{ij}, \eta_{ij})$ has heteroskedasticity then moment condition is violated.
- Why? Expectation is linear operator but $\log(\cdot)$ not so much.

Gravity: Part 2

Rearrange things so that:

$$T_{ij} = \exp\left(\beta_0 + \alpha_1 \log\left(Y_i\right) + \alpha_2 \log\left(Y_j\right) + \alpha_3 \log\left(D_{ij}\right)\right)\eta_{ij}$$
$$T_{ij} = \exp\left(x_i\beta\right)\eta_{ij}$$

This gives us our moment condition:

$$E\left[T_{ij} - \exp\left(x_i\beta\right)|x_i\right] = 0$$

This works as long as we are okay with proportional variance

Thanks!