Introduction to Generalised Method of Moments

Econometrics (2000) Hayashi, Chapter 3

陳 捷 jessekelighine.com

September 7-th, 2021

Roadmap

Motivation

Consider the normal linear regression form:

$$y_i = \begin{bmatrix} \mathbf{z}_{i1} & \cdots & \mathbf{z}_{iL} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_L \end{bmatrix} + \varepsilon_i \quad \rightsquigarrow \quad y_i = \mathbf{z}_i' \boldsymbol{\delta} + \varepsilon_i \qquad \text{(Model)}$$

where y_i is the dependent variable, $\mathbf{z}_i$ is the independent variable vector, $\boldsymbol{\delta}$ is the parameter vector. We are interested in the parameters $\boldsymbol{\delta}$.

- If $\mathbf{z}_i$ is exogenous, then we can simply use the OLS estimator.
- If $\mathbf{z}_i$ is *not* exogenous, then **instrument variables** can be used to estimate $\boldsymbol{\delta}$.

Let $\mathbf{x}_i$ be a vector of valid instruments (relevant and exogenous) with dimension $K \times 1$:

$$
\begin{aligned}
y_i = \mathbf{z}_i' \boldsymbol{\delta} + \varepsilon_i &\Longrightarrow \mathbf{x}_i y_i = \mathbf{x}_i \mathbf{z}_i' \boldsymbol{\delta} + \mathbf{x}_i \varepsilon_i \\
&\Longrightarrow \mathbb{E}[\mathbf{x}_i y_i] = \mathbb{E}\left[\mathbf{x}_i \mathbf{z}_i' \boldsymbol{\delta} + \mathbf{x}_i \varepsilon_i\right] \\
(\mathbf{x}_i \text{ is exogenous}) &\Longrightarrow \mathbb{E}[\mathbf{x}_i y_i] = \mathbb{E}\left[\mathbf{x}_i \mathbf{z}_i'\right] \boldsymbol{\delta} \\
(\mathbf{x}_i \text{ is relevant}) &\Longrightarrow \mathbb{E}\left[\mathbf{x}_i \mathbf{z}_i'\right]^{-1} \mathbb{E}[\mathbf{x}_i y_i] = \boldsymbol{\delta} \qquad \text{(IV)}
\end{aligned}
$$

If we replace the moments above with the corresponding estimators, then we obtain the familiar **instrument variable estimator** estimator $\hat{\boldsymbol{\delta}}_{\mathsf{IV}} = \mathbf{S}_{\mathbf{xz}}^{-1} \mathbf{s}_{\mathbf{x}y} \xrightarrow{p} \mathbb{E}[\mathbf{x}_i \mathbf{z}_i']^{-1} \mathbb{E}[\mathbf{x}_i y_i]$ where

$$
\mathbf{S}_{\mathbf{xz}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{z}_i' \quad \text{and} \quad \mathbf{s}_{\mathbf{x}y} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i y_i
$$

in which n is the number of samples.

However, what if $\mathbb{E}[\mathbf{x}_i\mathbf{z}_i']$ is not invertible?

It could be that...

1. $\mathbb{E}[\mathbf{x}_i\mathbf{z}_i']$ is not full rank. (invalid instruments)
   $\implies$ There really isn't remedy.
2. $\mathbb{E}[\mathbf{x}_i\mathbf{z}_i']$ is not square. (more instruments than endogenous variables, $K > L$)
   $\implies$ In general, a solution to $\boldsymbol{\delta}$ does not necessary exist.
   $\implies$ By the orthogonal condition, it is implied that it exists.
   $\implies$ For for solution to be unique, we have the following:

Definition (Identification)

The $K \times L$ matrix $\mathbb{E}[\mathbf{x}_i\mathbf{z}_i']$ is said to satisfy identification condition if it is of full *column* rank (rank $= L$). Denote this matrix by $\boldsymbol{\Sigma}_{\mathbf{xz}}$.

To be more precise, we can have the following three cases:

1. $(K > L)$ **under-identified**: Go find more instruments.
2. $(K = L)$ **just-identified**: IV-estimator.
3. $(K < L)$ **over-identified**: The solution to $\boldsymbol{\delta}$ cannot be obtained by IV-estimator, but we know a solution exists and is unique if the identification condition is met.

Therefore, the GMM (Generalised Method of Moments) is introduced to find a solution in the **over-identified** case.

It is apparent now that this methods builds on moment conditions that we've seen while deriving the IV-estimator.

The Generalised Method of Moments

Let's say the instrument variables are valid, we want to find the solution to the following equation:

$$\mathbb{E}[\mathbf{x}_i y_i] - \boldsymbol{\Sigma}_{\mathbf{xz}}\boldsymbol{\delta} = \mathbf{0} \qquad \text{(moment condition)}$$

We do not actually need to invert $\boldsymbol{\Sigma}_{\mathbf{xz}}$ to get a solution, we just need to find a $\boldsymbol{\delta}$ that makes the left-hand side zero. Similar to the IV case, we replace the moments with their estimators and let it be denoted by $\mathbf{g}_n(\tilde{\boldsymbol{\delta}})$:

$$\mathbf{g}_n(\tilde{\boldsymbol{\delta}}) := \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i(y_i - \mathbf{z}_i'\tilde{\boldsymbol{\delta}}) = \mathbf{s}_{\mathbf{x}y} - \mathbf{S}_{\mathbf{xz}}\tilde{\boldsymbol{\delta}} \overset{\text{let}}{=} \mathbf{0}$$

However, since we are now consider *samples*, it is not necessary that an "exact" solution exists. So we want to find a $\boldsymbol{\delta}$ that makes the equation as *close* to $\mathbf{0}$ as possible.

What we mean by *close to* $\mathbf{0}$? Consider the *quadratic form* as norm:

$$\|\tilde{\boldsymbol{\delta}}\| := \mathbf{g}_n(\tilde{\boldsymbol{\delta}})'\hat{\mathbf{W}}\mathbf{g}_n(\tilde{\boldsymbol{\delta}})$$

where $\hat{\mathbf{W}}$ is a matrix that converges to a symmetric and positive definite matrix $\mathbf{W}$ in probability as $n \to \infty$.

Therefore, our goal simplifies to minimising the above equation.

Note: In fact, We could have chosen any norm in $\mathbb{R}^K$. However, since the quadratic form is the most studied, understood, and mathematically tractable, it is chosen to be used as the norm.

Defining the Method

Definition (GMM Estimator)

Let $\hat{\mathbf{W}}$ be a $K \times K$ symmetric positive definite matrix (possibly dependent on the sample) s.t. $\hat{\mathbf{W}} \xrightarrow{p} \mathbf{W}$ also symmetric positive definite as the sample size $n \to \infty$. The <u>GMM estimator</u> of $\boldsymbol{\delta}$, denoted $\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})$, is

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = \arg\min_{\tilde{\boldsymbol{\delta}}} \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}}) \quad \text{where} \quad \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}}) := n \cdot \mathbf{g}_n(\tilde{\boldsymbol{\delta}})' \hat{\mathbf{W}} \mathbf{g}_n(\tilde{\boldsymbol{\delta}})$$

Assumptions

1. Linear model.
2. Ergodic stationary. (weaker assumption than iid)
3. Orthogonal condition. (exogeneity of instruments)
4. Rank condition for identification. (relevance of instruments)
5. $\mathbf{g}_i$ is a martingale difference sequence. (to ensure the asymptotic distribution of $\mathbf{g}_n(\boldsymbol{\delta})$ is normal)

Explicit Form

Assume that $\mathcal{J}$ is continuously differentiable, then we can obtain the explicit form by checking the first order condition:

$$\frac{\partial \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}})}{\partial \tilde{\boldsymbol{\delta}}} \equiv \begin{bmatrix} \frac{\partial \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}})}{\partial \tilde{\boldsymbol{\delta}}_1} \\ \frac{\partial \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}})}{\partial \tilde{\boldsymbol{\delta}}_2} \\ \vdots \\ \frac{\partial \mathcal{J}(\tilde{\boldsymbol{\delta}}, \hat{\mathbf{W}})}{\partial \tilde{\boldsymbol{\delta}}_K} \end{bmatrix} = 2n \cdot \mathbf{S}'_{\mathbf{xz}} \hat{\mathbf{W}} (\mathbf{s}_{\mathbf{x}y} - \mathbf{S}_{\mathbf{xz}} \tilde{\boldsymbol{\delta}}) \overset{\text{let}}{=} \mathbf{0}$$

by rearranging we have

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) = (\mathbf{S}'_{\mathbf{xz}} \hat{\mathbf{W}} \mathbf{S}_{\mathbf{xz}})^{-1} \mathbf{S}'_{\mathbf{xz}} \hat{\mathbf{W}} \mathbf{s}_{\mathbf{x}y}.$$

Sampling Error

The sampling error can be obtained by multiplying the true model by $\mathbf{x}_i$ on both sides and taking the average:

$$\begin{aligned} y_i &= \mathbf{z}_i'\boldsymbol{\delta} + \varepsilon_i \\ \Longrightarrow \mathbf{s}_{\mathbf{x}y} &= \mathbf{S}_{\mathbf{xz}}\boldsymbol{\delta} + \bar{\mathbf{g}} \end{aligned} \quad \text{where} \quad \bar{\mathbf{g}} := \mathbf{g}_n(\boldsymbol{\delta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\varepsilon_i$$

and then substitute $\mathbf{s}_{\mathbf{x}y}$ it into explicit form of GMM Estimator:

$$\begin{aligned} \hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) &= (\mathbf{S}_{\mathbf{xz}}'\hat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}_{\mathbf{xz}}'\hat{\mathbf{W}}(\mathbf{S}_{\mathbf{xz}}\boldsymbol{\delta} + \bar{\mathbf{g}}) \\ \Longrightarrow \hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}) - \boldsymbol{\delta} &= (\mathbf{S}_{\mathbf{xz}}'\hat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}_{\mathbf{xz}}'\hat{\mathbf{W}}\bar{\mathbf{g}} \xrightarrow{p} 0 \end{aligned}$$

The consistency of GMM immediately follows from the above.

Asymptotic Distribution

Consider the explicit form of GMM Estimation and multiply both sides by $\sqrt{n}$:

$$\underbrace{(\mathbf{S}'_{\mathbf{xz}}\hat{\mathbf{W}}\mathbf{S}_{\mathbf{xz}})^{-1}\mathbf{S}'_{\mathbf{xz}}\hat{\mathbf{W}}\left(\sqrt{n}\bar{\mathbf{g}}\right)}_{\sqrt{n}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})-\boldsymbol{\delta})} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \operatorname{Avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))\right)$$

where $\operatorname{Avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$ is the asymptotic covariance matrix. It can be obtained simply by the definition of covariance matrix and has the form

$$\operatorname{Avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}})) = (\boldsymbol{\Sigma}'_{\mathbf{xz}}\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1}\boldsymbol{\Sigma}'_{\mathbf{xz}}\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{xz}}(\boldsymbol{\Sigma}'_{\mathbf{xz}}\mathbf{W}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1}.$$

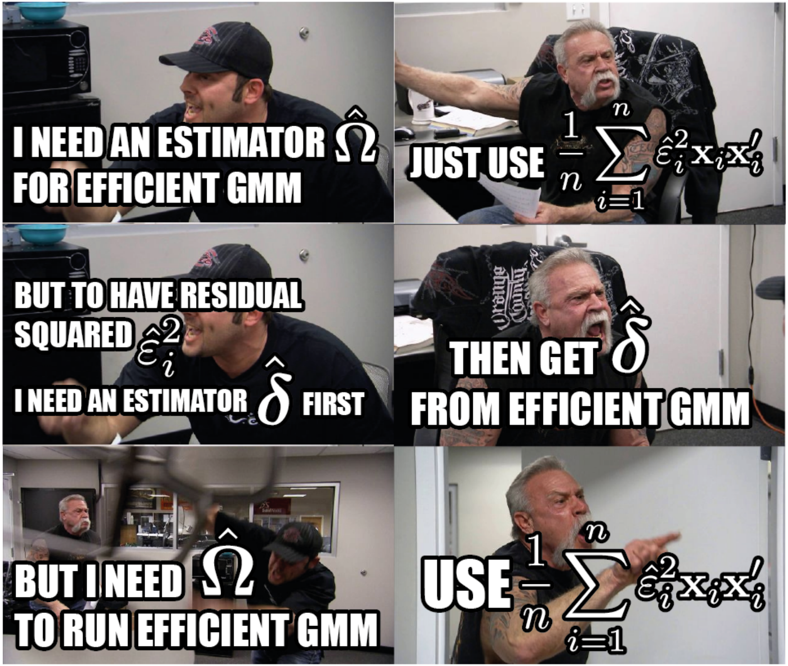
where $\boldsymbol{\Omega} := \mathbb{E}\left[\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right]$. With the asymptotic distribution, we can perform hypothesis testing as usual.

Efficient GMM

- The choice of $\hat{\mathbf{W}}$ will not effect the asymptotic distribution or consistency of GMM, but it will effect the **variance** of the estimator.
- Thus, a natural question is whether we can find some optimal $\hat{\mathbf{W}}$ that "minimises" the asymptotic variance $\mathrm{Avar}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{W}}))$.
- It can be shown that the by choosing $\mathbf{W} = \boldsymbol{\Omega}^{-1}$, we can achieve a lower bound on the asymptotic covariance matrix.

$$\begin{aligned}\mathrm{Avar}(\hat{\boldsymbol{\delta}}(\boldsymbol{\Omega}^{-1})) &= (\boldsymbol{\Sigma}_{\mathbf{xz}}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1}\,\cancel{\boldsymbol{\Sigma}_{\mathbf{xz}}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}}}\,\cancel{(\boldsymbol{\Sigma}_{\mathbf{xz}}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1}} \\ &= (\boldsymbol{\Sigma}_{\mathbf{xz}}'\boldsymbol{\Omega}^{-1}\boldsymbol{\Sigma}_{\mathbf{xz}})^{-1} \rightsquigarrow \text{lower bound}\end{aligned}$$

- However, how do we estimate $\boldsymbol{\Omega} = \mathbb{E}\left[\varepsilon_i^2\mathbf{x}_i\mathbf{x}_i'\right]$? $\rightsquigarrow$ We need to have an estimate of the residual $\hat{\varepsilon}_i$.


I NEED AN ESTIMATOR $\hat{\Omega}$ FOR EFFICIENT GMM
JUST USE $\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \mathbf{x}_i\mathbf{x}_i'$
BUT TO HAVE RESIDUAL SQUARED $\hat{\varepsilon}_i^2$
I NEED AN ESTIMATOR $\hat{\delta}$ FIRST
THEN GET $\hat{\delta}$ FROM EFFICIENT GMM
BUT I NEED $\hat{\Omega}$ TO RUN EFFICIENT GMM
USE $\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 \mathbf{x}_i\mathbf{x}_i'$

- However, to obtain a consistent estimator $\hat{\mathbf{\Omega}}$ for $\mathbf{\Omega} = \mathbb{E}\left[\varepsilon_i^2 \mathbf{x}_i \mathbf{x}_i'\right]$, we need to first obtain a consistent estimator for $\boldsymbol{\delta}$ since $\hat{\mathbf{\Omega}}$ is of the form

$$\hat{\mathbf{\Omega}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \mathbf{x}_i \mathbf{x}_i' \quad \text{where} \quad \hat{\varepsilon}_i^2 := y_i - \mathbf{z}_i' \hat{\boldsymbol{\delta}}$$

- Therefore, in practice, there are a number of methods that overcomes this problem, either by doing a initial estimation of $\boldsymbol{\delta}$ first or update $\hat{\mathbf{W}}$ iteratively.
  - **Two-step feasible GMM**: Use $\hat{\mathbf{W}} = I$ or $\mathbf{S}_{\mathbf{xx}}^{-1}$ for the first stage, then use the first-stage estimator $\hat{\boldsymbol{\delta}}$ to obtain $\hat{\mathbf{\Omega}}^{-1}$.
  - **Iterated GMM**: Repeat till it converges.
  - **Continuously updating GMM**: Plug $\mathbf{W}(\boldsymbol{\delta}) = \left(\frac{1}{n} \sum_{i=1}^{n} (y_i - \mathbf{z}_i' \boldsymbol{\delta})\right)^{-1}$ into $\mathcal{J}$ and minimise it.

J-statistic

Checking for instrument validity is very hard. Consistent with the GMM idea, the J-test is proposed:

Theorem (Hansen's test of overidentifying restrictions)

If all assumptions for GMM hold, and there is a consistent estimator for $\mathbf{\Omega}^{-1}$, then

$$\mathcal{J}(\hat{\boldsymbol{\delta}}(\hat{\mathbf{\Omega}}^{-1}), \hat{\mathbf{\Omega}}^{-1}) \xrightarrow{d} \chi^2_{K-L}$$

Notice that this test is for *all* assumptions. That is, if the test fails, any assumption could be false. Furthermore, it is an necessary condition for the validity of the assumptions, i.e., even if the test holds, it is not necessary that the assumptions hold.

Homoskedasticity $\implies$ 2SLS

If we impose homoskedasticity, we have

$$\mathbb{E}\left[\varepsilon_i^2 \,|\, \mathbf{x}_i\right] = \sigma^2 \implies \mathbf{\Omega} = \mathbb{E}\left[\varepsilon_i^2 \mathbf{x}_i \mathbf{x}_i'\right] = \sigma^2 \mathbf{\Sigma}_{\mathbf{xx}}$$
$$\implies \hat{\mathbf{\Omega}} = \frac{\hat{\sigma}^2}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' = \hat{\sigma}^2 \mathbf{S}_{\mathbf{xx}}$$

Therefore, the efficient GMM collapses to

$$\hat{\boldsymbol{\delta}}(\hat{\mathbf{\Omega}}^{-1}) = (\mathbf{S}_{\mathbf{xz}}'(\hat{\sigma}^2 \mathbf{S}_{\mathbf{xx}})^{-1} \mathbf{S}_{\mathbf{xz}})^{-1} \mathbf{S}_{\mathbf{xz}}'(\hat{\sigma}^2 \mathbf{S}_{\mathbf{xx}})^{-1} \mathbf{s}_{\mathbf{x}y}$$
$$= (\mathbf{S}_{\mathbf{xz}}' \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{S}_{\mathbf{xz}})^{-1} \mathbf{S}_{\mathbf{xz}}' \mathbf{S}_{\mathbf{xx}}^{-1} \mathbf{s}_{\mathbf{x}y} = \hat{\boldsymbol{\delta}}(\mathbf{S}_{\mathbf{xx}}^{-1}),$$

which is equivalent to solving the two stage equation

$$Z = X\beta + \eta \implies \hat{Z} = X(X'X)^{-1}X'Z \equiv PZ \quad (1)$$
$$Y = \hat{Z}\boldsymbol{\delta} + \varepsilon \implies \hat{\boldsymbol{\delta}}_{\text{2SLS}} = (Z'P'PZ)^{-1}Z'P'Y \equiv \hat{\boldsymbol{\delta}}(\mathbf{S}_{\mathbf{xx}}^{-1}). \quad (2)$$

Conclusion

- The motivation of GMM comes from generalising the moment conditions we use in OLS or IV.
- Most often in econometric applications, GMM refers to solving regressions with more instrument variables than endogenous regressors.
- GMM is often more computationally intensive, since the minimisation process is not very straight forward compared to OLS or IV. Nevertheless, it is much more computationally friendly than maximum likelihood.
- 2SLS is implied by homoskedasticity.

References

- Econometrics (2000) Fumio Hayashi, Chapter 3.
- Wikipedia on GMM: https://en.wikipedia.org/wiki/Generalized_method_of_moments. (largely based on Hayashi)
- Wikipedia on Covariance Matrix: https://en.wikipedia.org/wiki/Covariance_matrix. (definiteness properties of covariance matrix)