Econometría (II / Práctica)

Magíster en Economía

Tema 6: Método Generalizado de Momentos (GMM)

Prof. Luis Chancí
www.luischanci.com

Apuntes Econometría
Prof. Luis Chancí

$E(Y_i|X_i = x) = \int z\, f(z|X_i = x)dz$

Introduction

The **Method of Moments (MM)** and the **Generalized Method of Moments (GMM)**

- In MLE we assumed we know the p.d.f., which involves strong assumptions.
- What if we use fewer moments to obtain the estimates of interest?
- **MM** is a technique that involves equating sample moments to their theoretical counterparts (under a given model).
- **GMM** (Lars Hansen, 1982) extends the MM by allowing more equations than unknown parameters (overidentification) and incorporating general nonlinear functions of observations and parameters to obtain the estimates.

Thus, the idea is that we can come up with 'moment conditions' such that

$$\mathbb{E}(g(\boldsymbol{\theta}_0, \boldsymbol{w}_i)) = 0$$

where $\boldsymbol{w}_i$ denotes all the observables (e.g., both dependent and independent variables) for $\mathcal{Y}_N = (w_1, \ldots, w_N)$.

The Method of Moments

Let $g(\boldsymbol{w}_i, \boldsymbol{\theta})$ be a known $r \times 1$ function of the i^{th} observation and a $k \times 1$ parameter $\boldsymbol{\theta}$. In this section, we will first consider the just-identified where $r = k$.

Basic Principle: For a model with parameters θ, equate the sample moments to theoretical moments. To illustrate,

- Theoretical moment: $\mathbb{E}(X|\theta)$
- Sample moment: $\bar{X}_n = \frac{1}{n}\sum_i X_i$
- The equation g would be: $g_i = X_i - \mathbb{E}(X|\theta)$, and, therefore, $\bar{X}_n = \mathbb{E}(X|\theta))$.

In other words, the estimation procedure is to solve the moment equation for θ to obtain the **Method of Moments** estimator.

Let's check the following example.

The Method of Moments

Example: Estimating ν for a t-Student Distribution: Given $y_i \sim t - \text{student}(\nu)$ with probability density function (pdf):

$$f(y|\nu) = \frac{\Gamma((\nu+1)/2)}{(\pi y)^{1/2}\Gamma(\nu/2)}(1 + (y^2/\nu))^{-(\nu+1)/2}$$

for a sample $\{y_i\}$, obtain the MM estimator of ν.

The Method of Moments (cont.)

Answer. Under the assumption $\nu > 2$, the t-Student distribution has $\mathbb{E}(y) = 0$ and $\mathbb{E}(y^2) = \nu/(\nu - 2)$. Also, as $\nu \to \infty$, $Var \to 1$, and the t-Student distribution converges to a normal distribution $f(\cdot) \to N(0, 1)$.

Estimation Process:

- Calculate the sample moment $\hat{\mu}_2 = (1/N) \sum_i y_i^2$, which converges in probability to the theoretical moment:

$$\hat{\mu}_2 \to_p \mathbb{E}(y^2)$$

- A consistent estimator for ν is then derived from the equality:

$$\frac{\hat{\nu}}{\hat{\nu} - 2} = \hat{\mu}_2$$

- Rearranging, the estimator for ν is:

$$\hat{\nu} = \frac{2\hat{\mu}_2}{\hat{\mu}_2 - 1}$$

This is defined for $\hat{\mu}_2 > 1$ and is known as the classical method of moments estimator.

The Method of Moments (cont.)

Example: OLS as a Special Case of MM

- **Case 1: OLS.** In a linear regression context, OLS can be viewed as a particular case of MM where the moments are covariances between the independent variables and the residuals.
  - Set $\boldsymbol{w}_i = (Y_i, \boldsymbol{X}_i)$ and $g(\boldsymbol{\beta}; \boldsymbol{w}_i) = g_i(\boldsymbol{\beta}) = X_i u_i = X_i(Y_i - X_i'\boldsymbol{\beta})$. Thus, $g = N^{-1} \sum X_i(Y_i - X_i'\boldsymbol{\beta})$.
  - Then, the MM estimator is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$
- **Case 2: OLS and Variance.**
  - Set

$$g_i(\boldsymbol{\beta}, \sigma^2) = \begin{pmatrix} X_i(Y_i - X_i'\boldsymbol{\beta}) \\ (Y_i - X_i'\boldsymbol{\beta})^2 - \sigma^2 \end{pmatrix}$$

  - Thus, since the MME estimator is the parameter value which sets $(1/n)(\sum_i g_i(\boldsymbol{\beta}, \sigma^2)) = 0$, we have that $\hat{\boldsymbol{\beta}}_{MM} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{Y})$ and $\hat{\sigma}^2 = (1/n)\sum_i (Y_i - X_i'\hat{\boldsymbol{\beta}})^2$.

The Generalized Method of Moments (GMM)

As mentioned, **GMM** extends the MM by allowing more equations than unknown parameters (overidentification).

- Utilizing multiple moments can lead to more efficient and reliable estimators.
- GMM is particularly advantageous when dealing with complex distributions or when higher moments provide additional insights.
- GMM estimators are generally more efficient than classical MM estimators.

However, because we have more moments than parameters to estimate (more equations than unknowns), we can not solve all equations at once. That is, we still have $\mathbb{E}(g(\boldsymbol{w}_i, \boldsymbol{\beta})) = 0$, but we can no longer choose $\hat{\boldsymbol{\beta}}$ so that $(1/n)(\sum_i g_i(\boldsymbol{\beta}, \sigma^2)) = 0$.

Instead we choose $\boldsymbol{\beta}$ that minimizes the difference between the theoretical moments and their sample counterparts. In particular, we choose a (symmetric positive definite) weighting matrix $\boldsymbol{W}$ and minimize

$$Q(\boldsymbol{\beta}) = \left[\frac{1}{n}\sum_i g_i(\boldsymbol{\beta}, \sigma^2)\right]' \boldsymbol{W} \left[\frac{1}{n}\sum_i g_i(\boldsymbol{\beta}, \sigma^2)\right]$$

Definition: The Generalized Method of Moments (GMM) estimator is $\hat{\boldsymbol{\beta}}_{GMM} = \arg\min_{\boldsymbol{\beta}} Q(\boldsymbol{\beta})$.

GMM - example

Example: t-Student Distribution with Multiple Moments.

Previously, for the t-Student distribution, we used a single moment. But, we can use, for instance, both the second and fourth moments $\left(\mu_4 = 3\nu^2(\nu-2)^{-1}(\nu-4)^{-1}\right)$.

Thus, setting

$$g \equiv \begin{bmatrix} \hat{\mu}_2 - \nu/(\nu-2) \\ \hat{\mu}_4 - 3\nu^2/[(\nu-2)(\nu-4)] \end{bmatrix}$$

and, for ease of exposition, choosing $W = I_2$ (both moments hold equal relevance), we have that the objective function to minimize in GMM is:

$$Q(\nu) = \left[\hat{\mu}_2 - \frac{\nu}{\nu-2}\right]^2 + \left[\hat{\mu}_4 - \frac{3\nu^2}{(\nu-2)(\nu-4)}\right]^2$$

Here, $\hat{\mu}_2$ and $\hat{\mu}_4$ are the sample second and fourth moments.

The Generalized Method of Moments

Hansen's GMM Formulation: a systematic way to combine multiple moments and find the best parameter estimates.

1. Observables and Parameters: Let $\boldsymbol{w}_i$ be an $r \times 1$ vector of observed variables, and let $\boldsymbol{\theta}$ be $k \times 1$ vector of unknown parameters.
2. Moment Conditions: Define $h(\boldsymbol{\theta}, \boldsymbol{w}_i)$ as an $r \times 1$ vector of functions mapping $(\mathbb{R}^k \times \mathbb{R}^h) \to \mathbb{R}^r$. The true parameter vector $\boldsymbol{\theta}_0$ is characterized by the orthogonality conditions: $\mathbb{E}\left\{h(\boldsymbol{\theta}_0, \boldsymbol{w}_i)\right\} = 0$.
3. Sample Moments: Let $\mathcal{Y}_N = (w_1, \ldots, w_N)$ be a $Nh \times 1$ vector containing all observations (sample of size N), and define $g(\boldsymbol{\theta}; \mathcal{Y}_N)$ as a $r \times 1$ vector of averages of functions $h(\boldsymbol{\theta}, \boldsymbol{w}_i)$, where $g : \mathbb{R}^k \to \mathbb{R}^r$:

$$g(\boldsymbol{\theta}, \mathcal{Y}_N) = \frac{1}{N} \sum_{i=1}^{N} h(\boldsymbol{\theta}; \boldsymbol{w}_i)$$

4. **GMM Estimator:** The GMM estimator $\hat{\boldsymbol{\theta}}_N$ is the value of $\boldsymbol{\theta}$ that minimizes:

$$Q(\boldsymbol{\theta}, \mathcal{Y}_N) = g(\boldsymbol{\theta}, \mathcal{Y}_N)'\, W_N\, g(\boldsymbol{\theta}, \mathcal{Y}_N)$$

   where W_N is a positive definite weighting matrix, possibly data-dependent. As we will review later, the choice of this weighting matrix can influence the efficiency of the estimator.

The Generalized Method of Moments

A couple of notes on GMM, MM, and W:

- If the number of parameters (k) is the same as the number of orthogonality conditions (r), typically the objective function will be minimized by setting $g(\hat{\theta}, \mathcal{Y}_n) = 0$. Therefore, when $k = r$, the GMM estimator is the $\hat{\theta}_n$ that satisfies these r equations (same MM).

- If there are more orthogonality conditions than parameters ($r > k$), then $g(\hat{\boldsymbol{\theta}}, \mathcal{Y}_N) = 0$ will no hold exactly. How close the i-th element of $g()$ is to zero depends on how much weight the i-th orthogonality condition is given by the weighting matrix W_N.

Example 1 - Hansen's GMM formulation.

MM as a special case. For the t-Student distribution:

- $w_i = y_i$, $\theta = \nu$, $W_n = 1$, $h(\theta, w_i) = y_i^2 - \nu/(\nu - 2)$, and $g(\theta; Y_n) = n^{-1} \sum_i y_i^2 - \nu/(\nu - 2)$.

- here, $r = k = 1$ and the objective function becomes

$$Q(\theta; Y_n) = \left\{ \left(\frac{1}{n}\right) \sum_i y_i^2 - \frac{\nu}{\nu - 2} \right\}$$

Examples - Hansen's GMM Formulation

Example 2 - Hansen's GMM formulation.

And in the GMM case we covered for the t-Student distribution, the formulation would be:

- $r = 2$ and

$$h(\nu, y_i) = \begin{bmatrix} y_i^2 - \nu/(\nu - 2) \\ y_i^4 - 3\nu^2/[(\nu - 2)(\nu - 4)] \end{bmatrix}$$

$$g(\nu, \mathcal{Y}_N) = \left(\frac{1}{n}\right) \begin{bmatrix} \sum_i y_i^2 - \nu/(\nu - 2) \\ \sum_i y_i^4 - 3\nu^2/[(\nu - 2)(\nu - 4)] \end{bmatrix}$$

- and $\hat{\nu}$ is obtained from

$$\min_{\nu} \; [g(\nu, \mathcal{Y}_n)]' \, W_n \; [g(\nu, \mathcal{Y}_n)]$$

Linear Moment Models and the GMM Estimator

Other estimators can also be viewed as examples of GMM:

- OLS
- IV
- 2SLS
- Nonlinear simultaneous equations estimations
- Estimators for dynamic rational expectations models
- (many cases of) MLE

Linear Moment Models and the GMM Estimator (cont.)

In particular, let's focus on the (overidentified) IV model with moment equations

$$h(\boldsymbol{\beta}, w_i) = Z_i(Y_i - X_i'\boldsymbol{\beta})$$

GMM estimator: The GMM criterion can be written as

$$Q(\beta) = n(\boldsymbol{Z}'Y - \boldsymbol{Z}'\boldsymbol{X}\beta)'W(\boldsymbol{Z}'Y - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta})$$

The first order conditions are

$$0 = -2\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right)W\left(\frac{1}{n}\boldsymbol{Z}'(Y - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right)$$

Therefore, for the (overidentified) IV model:

$$\hat{\boldsymbol{\beta}}_{GMM} = \left(\boldsymbol{X}'\boldsymbol{Z}W\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}W\boldsymbol{Z}'Y\right)$$

Notes:

- For the just-identified model, $\boldsymbol{X}'\boldsymbol{Z}$ is $k \times k$, then $\hat{\boldsymbol{\beta}}_{GMM} = (X'Z)^{-1}W^{-1}(Z'X)^{-1}(X'Z)W(Z'Y) = (X'Z)^{-1}(Z'Y) = \hat{\boldsymbol{\beta}}_{IV}$.
- If $W = (Z'Z)^{-1}$ then $\hat{\boldsymbol{\beta}}_{GMM} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z(Z'Z)^{-1}Z'Y) = (X'P_ZX)^{-1}(X'P_ZY) = \hat{\boldsymbol{\beta}}_{2SLS}$.

Optimal Construction of the Weighting Matrix in GMM

In GMM, the weighting matrix W_N plays a crucial role in determining the efficiency of the estimator. As mentioned, this matrix is used in the objective function to give different weights to various moment conditions.

Theory (assumptions). Suppose $\{h(\theta_0, w_i)\}$ has zero mean and the variance-covariance matrix is given by $\Omega_\tau = \mathbb{E}\{(h(\theta_0, w_i))(h(\theta_0, w_j))'\}$. For time series, covariances are *absolutely summable* , leading to $\mathbb{S} = \sum_\tau \Omega_\tau$, where $\mathbb{S}$ is the asymptotic variance of the sample mean in $h(\theta_0, w_i)$,

$$\mathbb{S} = \lim_{N\to\infty} N\,\mathbb{E}\{(g(\theta_0; X_N))(g(\theta_0; X_N))'\}$$

Optimal Weighting Matrix.

- Theoretically, the optimal W_N is given by $\mathbb{S}^{-1}$.
- $\mathbb{S}$ depends on θ, implying that:

$$\hat{\mathbb{S}}_N \equiv \left(\frac{1}{N}\right)\sum_i [h(\hat{\theta}_N, w_i)][h(\hat{\theta}_N, w_i)]' \underset{p}{\rightarrow} \mathbb{S}$$

valid for any consistent estimator of θ_0.

Circular Dependency Issue and Practical Iterative Approach

In short, the optimal weighting matrix minimizes the variance of the GMM estimator and it is often chosen as the inverse of the covariance matrix of the moment conditions.

Circularity in Estimation: To obtain W_N, an estimate of $\hat{\theta}$ is required. However, W_N is necessary to minimize the objective function in GMM and thus obtain the estimated parameters $\hat{\theta}$.

Practical iterative approach:

- **Initial Estimate:** Start with an initial estimate of $\hat{\theta}$. Alternatively, start with a guess for W, such as the identity matrix for W (that is, $W^{(0)} = I$).
- **Update:** Using the initial guess ($W^{(0)}$) in the GMM criterion Q, compute $\hat{\theta}^{(0)}$.
- **Re-Estimate $\hat{W}$:** Update the weighting matrix to $W^{(1)}$ using $\hat{\theta}^{(0)}$, by calculating $W^{(1)} = (\hat{\mathbb{S}}^{(0)})^{-1}$. This matrix can also be used to compute $\hat{\theta}^{(1)}$.
- **Iterate:** Repeat the process until convergence is achieved. For instance, $||\hat{\theta}^{(j)} - \hat{\theta}^{(j+1)}|| < \epsilon$.

Asymptotic Distribution in GMM

Let $\hat{\theta}_N$ be the value that minimizes $[g(\theta, X_N)]'\hat{\mathbb{S}}_N^{-1}[g(\theta, X_N)]$. The GMM estimator is a solution to the system:

$$\left[\left. \frac{\partial g(\theta, X_N)}{\partial \theta'} \right|_{\theta = \hat{\theta}} \right] \hat{\mathbb{S}}_N^{-1} g(\hat{\theta}_N, X_N) = 0$$

Central Limit Theorem (CLT) Application:

Given that $g(\boldsymbol{\theta}, \mathcal{Y}_N)$ is the sample mean of a process with a population mean of zero, under additional conditions (e.g., continuity of $h(\cdot)$), $g(\boldsymbol{\theta}, \mathcal{Y}_N)$ should satisfy the CLT. Therefore,

$$\sqrt{N}\, g(\theta_0, \mathcal{Y}_N) \xrightarrow{L} N(0, \mathbb{S})$$

Asymptotic Distribution in GMM (cont.)

Proposition for GMM Estimator:

Consider $g(\theta_0, \mathcal{Y}_N)$ to be differentiable, and let $\hat{\theta}_N$ be the GMM estimator (for $r \geq k$).

Assuming that:

- $\hat{\theta}_N \to_p \theta_0$
- $\sqrt{N} g(\theta_0, \mathcal{Y}_N) \to_d \mathcal{N}(0, \mathbb{S})$
- $\operatorname{plim} \left(\partial g(\cdot) / \partial \theta \right)_{\theta = \hat{\theta}_N} \equiv D'$, with columns linearly independent.

Then, under these conditions, the GMM estimator is asymptotically normal:

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \to_L N(0, V)$$

where $V = (D\,\mathbb{S}^{-1} D')^{-1}$.

Asymptotic Distribution in GMM - Linear Moment Model

For the overidentified model with **linear moment equations** $h(\boldsymbol{\beta}, w_i) = \boldsymbol{Z}_i(Y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$:

Let $Q_{ZX} = \mathbb{E}(ZX')$ and $\Omega = \mathbb{E}(ZZ'u^2)$, then

- $(X'Z/N)W(Z'X/N) \underset{p}{\rightarrow} Q'_{XZ}WQ_{ZX}$
- $(X'Z/N)W(Z'e/N) \underset{d}{\rightarrow} Q'_{XZ}W\mathcal{N}(0,\Omega)$

Asymptotic Distribution: Under the assumptions listed in the slides for IV, as $N \rightarrow \infty$

$$\sqrt{N}\left(\hat{\beta}_{GMM} - \beta\right) \underset{d}{\rightarrow} \mathcal{N}(0, V_\beta)$$

where $V_\beta = (Q'_{XZ}WQ_{ZX})^{-1}(Q'_{XZ}W\Omega WQ_{ZX})(Q_{XZ}WQ_{ZX})^{-1}$.

Testing the Overidentifying Restrictions

Sargan (1958) introduced an overidentification test for the 2SLS estimator under the assumption of homoskedasticity.

Hansen (1982) generalized the test to cover the GMM estimator allowing for general heteroskedasticity. The idea is to test whether all the sample moments $g()$ are close to zero as one would be expected if the corresponding population moments $\mathbb{E}(h(\theta_0; w_i))$ were trully zero.

Overidentified models are special in the sense that there may not be a parameter value such that the moment condition $H_0 : (E)\{g(\boldsymbol{\theta}, \mathcal{Y}_N)\} = 0$ holds. Thus, the overidentifying restrictions are testable.

Since $\sqrt{N}\, g(\boldsymbol{\theta}_0, \mathcal{Y}_N) \rightarrow \mathcal{N}(0, \mathbb{S})$ and $g(\hat{\boldsymbol{\theta}}, \mathcal{Y}_N)$ contains ($r - k$) nondegenerate random variables, a test of the overidentifying restrictions is

$$\left(\sqrt{N}\, g(\hat{\theta}, \mathcal{Y}_N)\right)' \hat{\mathbb{S}}^{-1} \left(\sqrt{N}\, g(\hat{\theta}, \mathcal{Y}_N)\right) \underset{d}{\rightarrow} \chi^2_{(r-k)}$$

If reject H_0, then the estimator (GMM) is inconsistent for θ.

Example of GMM using R

```r
# Let's first simulate some data:
library(gmm)
set.seed(123)
N <- 100 ; X <- rnorm(N, mean=3, sd=1.2) ; b_0 <- 1.2 ; b_1 <- 2.5 ; eps <- rnorm(N) ; Y <- b_0 + b_1*X + eps
```

```r
# Case 1: using X'u for beta1:
#         (regresión lineal)

gmm_moments1 <- function(theta,yx) {
  y  <- yx[, 1]
  x  <- yx[,-1]
  u  <- y - (theta[1] + theta[2]*x)
  h  <- cbind(u, x*u)                # Moments: u and x*u
  return(h)
}

gmm_model1 <- gmm(gmm_moments1, x=as.matrix(cbind(Y,X)), t0=c(0.1,0.1))

mi_tabla(gmm_model1)              # mi_tabla: own function for html
```

Variable	Coeff.	S.error	t.stat.	p-value
Theta[1]	0.553	0.252	2.19	0.0285
Theta[2]	**2.651**	**0.077**	**34.46**	**<0.001**

```r
# Case 2: using X'u and X^2'u for beta1
#         (base on projection onto a linear subspace)

gmm_moments2 <- function(theta,yx) {
  y  <- yx[, 1]
  x  <- yx[,-1]
  u  <- y - (theta[1] + theta[2]*x)
  h  <- cbind(u, x*u, x^2*u)            # Moments: u, x*u, and x^2*u
  return(h)
}

gmm_model2 <- gmm(gmm_moments2, x=as.matrix(cbind(Y,X)), t0=c(0.1,0.1))

mi_tabla(gmm_model2)
```

Variable	Coeff.	S.error	t.stat.	p-value
Theta[1]	1.235	0.233	5.31	<0.001
Theta[2]	**2.464**	**0.068**	**36.19**	**<0.001**

Cierre

¿Preguntas?

O vía E-mail: lchanci1@binghamton.edu