

Método Generalizado de Momentos (MGM/GMM)

Econometría Aplicada y Ciencia de Datos - CIDE Fall 2026

Carlos Brito

$\hat{\beta}_{v1} = (Z'X)^{-1}(Z'y)$

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

$\left\{ \begin{array}{l} 1^{st} \text{ Stage : get } \hat{X} \\ 2^{nd} \text{ Stage : } \hat{\beta} = (\hat{X}'\hat{X})^{-1}(\hat{X}'y) \end{array} \right.$

$\text{When } n=q$

- 1 Introducción al MGM
- 2 Método de Momentos: Caso Exactamente Identificado ($r = q$)
- 3 Método de Momentos: Caso Sobreidentificado ($r > q$)
- 4 Matriz de Varianzas Óptima
- 5 MGM para Variables Instrumentales (VI-MGM)
- 6 VI-MGM ($r > q$) - Heterocedasticidad
- 7 VI-MGM ($r > q$) - Homocedasticidad
- 8 VI-MGM: Caso Exactamente Identificado
- 9 Síntesis

¿Qué es el Método Generalizado de Momentos?

El MGM (*GMM*) es un marco que **generaliza** varios estimadores comunes en econometría:

- MCO, máxima verosimilitud (MV), variables instrumentales (VI), entre otros.

¿Qué es el Método Generalizado de Momentos?

El MGM (*GMM*) es un marco que **generaliza** varios estimadores comunes en econometría:

- MCO, máxima verosimilitud (MV), variables instrumentales (VI), entre otros.

Empezamos con:

Condiciones de momentos poblacionales

Suponemos r condiciones de momentos, independientes entre sí, para q parámetros:

$$E[h(w_i, \theta_0)] = 0$$

- θ : vector de parámetros, $q \times 1$.
- $h(\cdot)$: función vectorial, $r \times 1$, con $r \geq q$.
- w_i : datos observados (variable dependiente, regresores exógenos y endógenos, instrumentos).

La elección de la función $h(\cdot)$ **equivale** a elegir el modelo de estimación:

$h(\cdot)$	Método de estimación
$E[x(y - x'\beta)] = 0$	MCO
$\frac{\partial \ell}{\partial \theta} = 0$	Máxima Verosimilitud
$E[z\mu] = 0$	Variables Instrumentales

Distintas elecciones de $h(\cdot)$ recuperan distintos estimadores ya conocidos como casos particulares de un mismo principio: hacer que un promedio muestral de h sea (aproximadamente) cero.

Método de Momentos ($r = q$)

Condiciones de momentos poblacionales

$$E[h(w_i, \theta_0)] = 0$$

Estimador de Método de Momentos

$\hat{\theta}_{MM}$ resuelve directamente el sistema muestral:

$$\frac{1}{N} \sum_{i=1}^N h(w_i, \hat{\theta}) = 0$$

que es lo mismo que minimizar:

$$Q_N(\theta) = \left(\frac{1}{N} \sum_{i=1}^N h(w_i, \theta) \right)' \left(\frac{1}{N} \sum_{i=1}^N h(w_i, \theta) \right)$$

La lógica: se reemplaza la esperanza poblacional por su análogo muestral (el promedio), y se despeja θ .

Ejemplo: la Media Poblacional

Sea $\{y_i\}$ una muestra i.i.d. con media poblacional μ . El momento poblacional es

$$E[y] = \mu \Rightarrow E[y] - \mu = 0 \Rightarrow E[y - \mu] = 0$$

↳ Exactly identified:

↳ parameter: μ

↳ moment: $E[y - \mu] = 0$

$$E(\underbrace{y - \mu}) = 0 \\ = R(w_i, \theta)$$

$$\frac{1}{N} \sum_{i=1}^N (y_i - \mu) = 0$$

$$\frac{1}{N} \sum_{i=1}^N y_i - \underbrace{\frac{1}{N} \sum_{i=1}^N (\mu)}_{\frac{1}{N} \cdot N \cdot \mu} = 0 \Rightarrow \hat{\mu}_{MM} = \frac{1}{N} \sum_{i=1}^N y_i$$

constant

$$\hat{\mu}_{MM} = \underbrace{\frac{1}{N} \sum_{i=1}^N y_i}_{\text{Sample mean}}$$

Método de Momentos $r > q$

Número de condiciones de momentos = número de parámetros: modelo **exactamente identificado**

Más condiciones de momentos que parámetros ($r > q$): modelo **sobreidentificado**. Ya no es posible resolver el sistema exactamente (no existe una solución única).

Estimador de MGM

$\hat{\theta}_{GMM}$ minimiza la forma cuadrática

$$Q_N(\theta) = \left(\frac{1}{N} \sum_i h(w_i, \theta) \right)' \underbrace{W_N}_{\text{Weighting}} \left(\frac{1}{N} \sum_i h(w_i, \theta) \right)$$

Método de Momentos $r > q$

Número de condiciones de momentos = número de parámetros: modelo **exactamente identificado**

Más condiciones de momentos que parámetros ($r > q$): modelo **sobreidentificado**. Ya no es posible resolver el sistema exactamente (no existe una solución única).

Estimador de MGM

$\hat{\theta}_{GMM}$ minimiza la forma cuadrática

$$Q_N(\theta) = \left(\frac{1}{N} \sum_i h(w_i, \theta) \right)' W_N \left(\frac{1}{N} \sum_i h(w_i, \theta) \right)$$

- W_N : matriz simétrica, definida positiva, que **no** depende de θ (matriz identidad, por ejemplo).
- Distintas elecciones de W_N generan distintos estimadores de MGM (todos consistentes, pero con distinta eficiencia).

Condiciones de Primer Orden

Derivando $Q_N(\theta)$ respecto a θ e igualando a cero:

$$\frac{\partial Q_N(\theta)}{\partial \theta} = \left(\frac{1}{N} \sum_{i=1}^N \frac{\partial h_i(\hat{\theta})'}{\partial \theta} \right) W_N \left(\frac{1}{N} \sum_{i=1}^N h_i(\hat{\theta}) \right) = 0$$

(usamos $h_i(\theta) \equiv h(w_i, \theta)$ para aligerar la notación)

- Es un sistema de ecuaciones, en general **no lineal**.
- Se resuelve con métodos numéricos para obtener $\hat{\theta}_{MGM}$.
- A partir de aquí se construye la teoría asintótica del estimador.

Bajo condiciones de regularidad (LGN, TLC)

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta_0) \overset{a}{\sim} \mathcal{N}\left(0, (G_0' W_0 G_0)^{-1} (G_0' W_0 S_0 W_0 G_0) (G_0' W_0 G_0)^{-1}\right)$$

donde W_0 es finita, simétrica y definida positiva, y

$$G_0 = \text{plim} \frac{1}{N} \sum_{i=1}^N \frac{\partial h_i}{\partial \theta'} \Big|_{\theta_0}, \quad S_0 = \text{plim} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (h_i h_j) \Big|_{\theta_0}$$

La derivación formal es más elaborada que la de los estimadores extremos estándar; véase Cameron & Trivedi (2005), Sec. 6.3.9.

La Elección Óptima de W_N

Para implementar MGM debemos elegir tanto las condiciones de momentos como la matriz de ponderación W_N .

for example:

$$v_2 = 3 \left\{ \begin{array}{l} (1) E[\quad] = 0 \\ (2) E[\quad] = 0 \\ (3) E[\quad] = 0 \end{array} \right. \quad \underbrace{\text{Var}((1)) > \text{Var}((2)) > \text{Var}((3))}_{\leftarrow \text{receive more weight}}$$

$$\hookrightarrow W_N \text{ óptima} = S^{-1}$$

S = variance matrix of the population moments

La Elección Óptima de W_N

Para implementar MGM debemos elegir tanto las condiciones de momentos como la matriz de ponderación W_N .

Resultado clave

Si S_0 fuera conocida, el estimador de MGM **más eficiente** (menor varianza asintótica) se obtiene con

$$W_N = S_0^{-1}, \quad S_0 = \text{plim} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (h_i h_j) |_{\theta_0}$$

La Elección Óptima de W_N

Para implementar MGM debemos elegir tanto las condiciones de momentos como la matriz de ponderación W_N .

Resultado clave

Si S_0 fuera conocida, el estimador de MGM **más eficiente** (menor varianza asintótica) se obtiene con

$$W_N = S_0^{-1}, \quad S_0 = \text{plim} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (h_i h_j)' |_{\theta_0}$$

Con esta elección, la varianza asintótica se simplifica considerablemente:

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta_0) \overset{a}{\sim} \mathcal{N}\left(0, (G_0' S_0^{-1} G_0)^{-1}\right)$$

En la práctica S_0 es desconocida, así que se reemplaza por un estimador consistente $\hat{S}$. Una mala elección de $h(\cdot)$ podría dar lugar a un estimador muy ineficiente.

¿Por Qué es Óptima Esta Elección?

No es solo una simplificación algebraica

Que $W_N = S_0^{-1}$ sea la elección óptima **no** se debe a que la expresión resultante sea más corta.

Es óptima porque se puede demostrar que, para *cualquier* matriz W_0 definida positiva,

$$(G_0' W_0 G_0)^{-1} (G_0' W_0 S_0 W_0 G_0) (G_0' W_0 G_0)^{-1} \geq (G_0' S_0^{-1} G_0)^{-1}$$

en el sentido de matrices semidefinidas positivas: la varianza con cualquier otra W_0 es (débilmente) mayor.

Una demostración de este resultado puede consultarse en Hayashi (2000), p. 245.

MGM Óptimo Factible: Procedimiento en Dos Etapas

Como S_0 es desconocida, se estima siguiendo un procedimiento de dos etapas:

MGM Óptimo Factible: Procedimiento en Dos Etapas

Como S_0 es desconocida, se estima siguiendo un procedimiento de dos etapas:

- 1 **Primera etapa:** obtener un estimador de MGM con una matriz subóptima, típicamente $W_N = I_r$, y usar esos coeficientes para construir

$$S_0 = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \sum_i \sum_j h_i h_j' \Big|_{\theta = \theta_0}$$
$$\hat{S} = \frac{1}{N} \sum_i h_i(\hat{\theta}) h_i(\hat{\theta})'$$

$\hat{\theta}_{GMM}$ using $W_N = I_n$

MGM Óptimo Factible: Procedimiento en Dos Etapas

Como S_0 es desconocida, se estima siguiendo un procedimiento de dos etapas:

- 1 **Primera etapa:** obtener un estimador de MGM con una matriz subóptima, típicamente $W_N = I_r$, y usar esos coeficientes para construir

$$\hat{S} = \frac{1}{N} \sum_i h_i(\hat{\theta}) h_i(\hat{\theta})'$$

- 2 **Segunda etapa:** obtener el **estimador de MGM de dos etapas óptimo** $\hat{\theta}_{MGM,O}$ minimizando

$$\hat{\theta}_{MGM,O} = \min: Q_N(\theta) = \left(\frac{1}{N} \sum_i h(\theta) \right)' \hat{S}^{-1} \left(\frac{1}{N} \sum_i h(\theta) \right)$$

La varianza estimada de $\hat{\theta}_{MGM,O}$ es $\hat{V}(\hat{\theta}_{MGM,O}) = N^{-1}(\hat{G}' \hat{S}^{-1} \hat{G})^{-1}$, evaluada en $\hat{\theta}_{MGM,O}$.

MGM para Variables Instrumentales

Consideremos el modelo lineal

$$\underline{y_i = x_i' \beta + u_i}$$

con algún componente de x_i endógeno. Recurrimos a un instrumento z_i que cumple

$$E(u_i | z_i) = 0$$

La restricción de exclusión implica la condición de momentos

$$E[\underline{z_i(y_i - x_i' \beta)}] = 0$$

Esta es exactamente la forma $h(\cdot) = z(y - x'\beta)$ de la tabla inicial — el caso de VI dentro del marco de MGM.

$$\begin{array}{c} h(\bar{w}_i, \theta) \\ \downarrow \quad \downarrow \\ y, x, z \quad \beta \end{array}$$

Estimador Lineal de VI-MGM: Solución Cerrada

El estimador de MGM minimiza

$$Q_N(\beta) = \left(\frac{1}{N} (y - X\beta)' z \right) W_N \left(\frac{1}{N} z' (y - X\beta) \right)$$

Condiciones de primer orden:

$$\frac{\partial Q_N(\beta)}{\partial \beta} = -2 \left(\frac{1}{N} X' z \right) W_N \left(\frac{1}{N} z' (y - X\beta) \right) = 0$$

isolating β

Estimador lineal de VI-MGM

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

Sus propiedades asintóticas pueden derivarse de forma análoga al estimador de MCO, o apelando directamente a la teoría general de MGM.

Antes de $\text{Var}(\hat{\beta}_{GMM})$: Repaso LLN y CLT

$$X \sim (\mu, \sigma^2) \quad E[X] = \mu, \text{Var}(X) = \sigma^2$$

* LLN: $\bar{X}_n \xrightarrow{p} \mu$

* CLT: $\bar{X}_n \xrightarrow{d} N(\mu, \sigma^2/n)$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

$$\frac{1}{\sqrt{n}} Z'_\mu = \frac{\Gamma_n Z'_\mu}{n}$$

$$= \underbrace{\Gamma_n \frac{Z'_\mu}{n}}_{\text{by CLT}} - 0$$

$$\xrightarrow{d} \underline{N(0, \text{Var}(\widehat{Z'_\mu}))}$$

Varianza de $\hat{\beta}_{GMM}$: Derivación

Siguiendo el mismo procedimiento que con MCO: sustituimos $y = X\beta + u$,

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

Reescalando por $\sqrt{N}$:

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

Si se cumple un TLC para $\frac{1}{\sqrt{N}}Z'u$: $\xrightarrow{d} N(0, S)$

$$S = \lim \frac{1}{N} \sum_{i=1}^N E(\mu_i^2 z_i z_i')$$

$\hat{\beta}_{GMM}$ es asintóticamente normal, centrado en β , con varianza asintótica estimada:

$$\hat{V}(\hat{\beta}_{GMM}) = N(X'ZW_NZ'X)^{-1}(X'ZW_N\hat{S}W_NZ'X)(X'ZW_NZ'X)^{-1}$$

donde $\hat{S}$ es un estimador consistente de S .

Estimador Óptimo de MGM - Heterocedasticidad

Para el estimador **óptimo**, elegimos $W = \hat{S}^{-1}$:

$$\hat{\beta}_{GMM,O} = (X'Z\hat{S}^{-1}Z'X)^{-1}X'Z\hat{S}^{-1}Z'y$$

Con **heterocedasticidad**:

$$\hat{S} = \frac{1}{N} \sum_i \hat{u}_i^2 z_i z_i' = \frac{1}{N} Z' D Z, \quad D = \text{diag} [\hat{u}_1^2, \hat{u}_2^2, \dots, \hat{u}_N^2]$$

- Se obtiene primero $\hat{\beta}$ con una matriz subóptima para construir $\hat{S}$, y luego se recalcula $\hat{\beta}_{GMM,O}$ usando $\hat{S}^{-1}$ como matriz de pesos.

$$\hat{V}(\hat{\beta}_{GMM,O}) = N(X'Z\hat{S}^{-1}Z'X)^{-1}X'Z\hat{S}^{-1}\hat{S}\hat{S}^{-1}Z'X(X'Z\hat{S}^{-1}Z'X)^{-1} = N(X'Z\hat{S}^{-1}Z'X)^{-1}$$

- Para estimar $\hat{S}$ se usan los residuos $\hat{u}_i = y_i - x_i' \hat{\beta}_{GMM,O}$.

Estimador Óptimo de MGM - Homocedasticidad

Con **homocedasticidad**:

can ignore in a weighting matrix

$$\hat{S}^{-1} = \left(\frac{1}{N} s^2 Z'Z \right)^{-1}, \text{ y elegimos } W = \left(\frac{1}{N} Z'Z \right)^{-1}$$

El estimador de MGM se reduce a:

$$\hat{\beta}_{MC2E} = (X' Z(Z'Z)^{-1} Z' X)^{-1} X' Z(Z'Z)^{-1} Z' y = (X' P_Z X)^{-1} X' P_Z y$$

con $P_Z = Z(Z'Z)^{-1} Z'$ la **matriz de proyección**. Este es el **estimador de MC2E (2SLS)** — también llamado **estimador de variables instrumentales generalizado**, y a veces **estimador de MGM de una etapa**.

$$\hat{V}(\hat{\beta}_{MC2E}) = s^2 (X' P_Z X)^{-1}$$

$r = q$: MGM = Estimador de VI

Cuando $r = q$ (tantos instrumentos como regresores endógenos + exógenos), $X'Z$ es **cuadrada e invertible**, así que

$$(X'ZW_NZ'X)^{-1} = (Z'X)^{-1}W_N^{-1}(X'Z)^{-1}$$

Sustituyendo en el estimador general de MGM:

$$\hat{\beta}_{GMM} = (X'ZW_NZ'X)^{-1}X'ZW_NZ'y = (Z'X)^{-1} \cancel{W_N^{-1}(X'Z)^{-1}X'ZW_NZ'} y = (Z'X)^{-1}Z'y = \hat{\beta}_{VI} = \hat{\beta}_{2SLS}$$

W_N don't matter

Resultado clave

En el modelo exactamente identificado, el estimador de MGM **no depende** de la elección de W_N — siempre coincide con el estimador simple de VI.

$$r = q: \hat{\beta}_{GMM} = \hat{\beta}_{VI} = \hat{\beta}_{2SLS}$$

$$\beta_{OLS} = (X'X)^{-1}(X'y)$$

$$* \beta_{IV} = (Z'X)^{-1}(Z'y)$$

$$* \hat{\beta}_{2SLS} = (\hat{X}'\hat{X})^{-1}(\hat{X}'y)$$

$$= (X'Z(Z'Z)^{-1}Z'X)(X'Z(Z'Z)^{-1}Z'y)$$

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

$$* \hat{\beta}_{GMM,0} = (X'Z\hat{S}Z'X)^{-1}(X'Z\hat{S}Z'y)$$

$$W_N = \hat{S}^{-1}, \hat{S} = \frac{1}{N} \sum \hat{u}_i^2 z_i z_i'$$

$$* \text{homocedasticity: } \hat{\beta}_{GMM,0} = \hat{\beta}_{MC2E} = (X'P_ZX)^{-1}(X'P_Zy)$$

$$P_Z = Z(Z'Z)^{-1}Z'$$

$$\otimes n = q: \hat{\beta}_{IV} = \hat{\beta}_{GMM} = \hat{\beta}_{2SLS}$$

~~$$\hat{\beta}_{GMM,0} \hat{\beta}_{MC2E}$$~~

$$\otimes n > q:$$

$$\hookrightarrow \text{homoc.}: \hat{\beta}_{2SLS} \text{ efficient}$$

$$\hat{\beta}_{2SLS} = \hat{\beta}_{MC2E}$$

$$\hookrightarrow \text{hetero}: \hat{\beta}_{GMM,0} \text{ efficient}$$

$\Leftrightarrow$

$$(X'P_ZX)^{-1}(X'P_Zy)$$

$$P_Z = Z(Z'Z)^{-1}Z'$$

Siguiendo las convenciones de Cameron & Trivedi (2005):

- **Estimador de MGM:** el caso general de método de momentos, para *cualquier* forma de los momentos y *cualquier* W_N .
- **Estimador óptimo de MGM:** elige $W = \hat{S}^{-1}$; es la versión más eficiente. Se usa en el caso más general — modelos de VI sobreidentificados con heterocedasticidad.
- **Estimador de variables instrumentales generalizado (MC2E):** el estimador óptimo de MGM bajo el supuesto adicional de **homocedasticidad** en un modelo sobreidentificado. Se llama “generalizado” porque generaliza VI al caso sobreidentificado.
- **Estimador de variables instrumentales:** surge en el modelo **exactamente identificado** — coincide con MGM para cualquier elección de W_N .