Variables Instrumentales en la Práctica

Econometría Aplicada y Ciencia de Datos - CIDE Fall 2026

Carlos Brito

Índice

Repaso: VI-GMM, Optimal VI-GMM, 2SLS, VI

Table 6.2. *GMM Estimators in Linear IV Model and Their Asymptotic Variance*[a]

Estimator	Definition and Asymptotic Variance
GMM	$\widehat{\boldsymbol{\beta}}_{\text{GMM}} = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{y}$
(general $\mathbf{W}_N$)	$\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = N[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\widehat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}][\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$
Optimal GMM	$\widehat{\boldsymbol{\beta}}_{\text{OGMM}} = [\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}$
($\mathbf{W}_N = \widehat{\mathbf{S}}^{-1}$)	$\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = N[\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$
2SLS	$\widehat{\boldsymbol{\beta}}_{\text{2SLS}} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$
($\mathbf{W}_N = [N^{-1}\mathbf{Z}'\mathbf{Z}]^{-1}$)	$\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = N[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\widehat{\mathbf{S}}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]$ $\times [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$
	$\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = s^2[\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}$ if homoskedastic errors
IV	$\widehat{\boldsymbol{\beta}}_{\text{IV}} = [\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{Z}'\mathbf{y}$
(just-identified)	$\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}] = N(\mathbf{Z}'\mathbf{X})^{-1}\widehat{\mathbf{S}}(\mathbf{X}'\mathbf{Z})^{-1}$

[a] Equations are based on a linear regression model with dependent variable $\mathbf{y}$, regressors $\mathbf{X}$, and instruments $\mathbf{Z}$. $\widehat{\mathbf{S}}$ is defined in (6.40) and s^2 is defined after (6.41). All variance matrix estimates assume errors that are independent across observations and heteroskedastic, aside from the simplification for homoskedastic errors given for the 2SLS estimator. Optimal GMM uses the optimal weighting matrix.

Repaso: de VI-GMM para Optimal VI-GMM

Partimos de la forma general del estimador de MGM y de su varianza:

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

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

Con la elección óptima

Repaso: de VI-GMM para Optimal VI-GMM

Partimos de la forma general del estimador de MGM y de su varianza:

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

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

Con la elección óptima $W = \hat{S}^{-1}$, $\quad S = \lim \frac{1}{N} \sum_i E(u_i^2 z_i z_i')$:

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

Los Datos

Usamos los datos de Card (1995) sobre retornos a la educación, con la proximidad a una universidad como instrumento de los años de escolaridad acumulados.

R

```r
data.ingresos <- read_csv("ingresos_iv.csv",
                          locale = locale(encoding = "latin1"))
```

Modelo con cinco regresores más la constante: `educ`, `exper`, `expersq`, `black`, `south`. El instrumento excluido es `nearc4` (universidad cercana).

Modelo Exactamente Identificado: Referencia con ivreg (IV)

R

```
iv_ei <- ivreg(lwage ~ educ + exper + expersq + black + south |
                . - educ + nearc4, data = data.ingresos)
```

	Coeficiente
educ	0.2214
	(0.0409)
exper	0.1439
	(0.0187)
N	3,010

Replicando con Álgebra de Matrices (IV)

Construimos X (regresores), Y (dependiente) y Z (instrumentos, reemplazando `educ` por `nearc4`) y aplicamos directamente la fórmula del estimador exactamente identificado:

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

R

```
X <- data.matrix(select(data.ingresos, constant, educ, exper,
                        expersq, black, south))
Y <- data.matrix(select(data.ingresos, lwage))
Z <- data.matrix(select(data.ingresos, constant, nearc4, exper,
                        expersq, black, south))

N <- nrow(X)
k <- ncol(X) # incluyendo la constante

b <- solve(t(Z) %*% X) %*% t(Z) %*% Y
```

El resultado reproduce exactamente el coeficiente de `educ` $= 0.2214$ obtenido con `ivreg`.

2SLS Matriz de Varianzas (Homocedástica)

$$\widehat{\boldsymbol{\beta}}_{\text{2SLS}} = \left[\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{X}\right]^{-1}\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{y}$$

$$\widehat{V}[\widehat{\boldsymbol{\beta}}] = s^2\left[\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{X}\right]^{-1} \quad \text{if homoskedastic errors}$$

Con la matriz de proyección $P_Z = Z(Z'Z)^{-1}Z'$ y $\hat{\sigma}^2 = N^{-1}\hat{u}'\hat{u}$:

$$\hat{V}(\hat{\beta}) = \hat{\sigma}^2(X'P_ZX)^{-1} \cdot \frac{N}{N-k}$$

(el factor $N/(N-k)$ es el ajuste por grados de libertad que usa R por defecto)

R

```
u_hat <- Y - X %*% b
sigma2 <- as.numeric((1/N) * t(u_hat) %*% u_hat)
P <- Z %*% solve(t(Z) %*% Z) %*% t(Z)

V <- sigma2 * solve(t(X) %*% P %*% X) * (N / (N - k))
sqrt(diag(V))
```

Los errores estándar replicados coinciden con los reportados por `ivreg` (0.0409 para `educ`).

2SLS Matriz de Varianzas (Heterocedasticidad)

$$\widehat{\beta}_{2SLS} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'y$$

$$\widehat{V}[\widehat{\beta}] = N\left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}\left[X'Z(Z'Z)^{-1}\widehat{S}(Z'Z)^{-1}Z'X\right] \times \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}$$

$$\hat{S} = \frac{1}{N}Z'DZ, \qquad D = \text{diag}\left[\hat{u}_i^2\right]$$

R

```
D <- diag(as.vector((Y - X %*% b)^2))
S_hat <- (1/N) * t(Z) %*% D %*% Z
Vr <- N * solve(t(X) %*% Z %*% solve(t(Z) %*% Z) %*% t(Z) %*% X) %*%
  (t(X) %*% Z %*% solve(t(Z) %*% Z) %*% S_hat %*% solve(t(Z) %*% Z) %*% t(Z) %*% X) %*%
  solve(t(X) %*% Z %*% solve(t(Z) %*% Z) %*% t(Z) %*% X)

sqrt(diag(Vr))
```

	Homoced.	Heteroced.	Heteroced. (corrección pequeñas muestras)
educ	0.2214 (0.0409)	0.2214 (0.0403)	0.2214 (0.0404)
exper	0.1439 (0.0187)	0.1439 (0.0185)	0.1439 (0.0186)

la corrección de muestra finita: $N/(N-k)$.

VI es el Estimador de MGM para Cualquier W

En el modelo **exactamente identificado**, comparar la matriz óptima ($W = \hat{S}^{-1}$) contra la matriz identidad ($W = I$) da **exactamente el mismo resultado**:

	Matriz óptima	Identidad
educ	0.2214 (0.0409)	0.2214 (0.0409)
exper	0.1439 (0.0187)	0.1439 (0.0187)

Esto confirma, empíricamente, el resultado teórico visto antes: cuando $r = q$, la elección de W_N es irrelevante — el estimador de MGM colapsa siempre al estimador simple de VI.

Modelo Sobreidentificado ($r > q$): Dos Instrumentos ivreg (2SLS)

$$\widehat{\boldsymbol{\beta}}_{2\text{SLS}} = \left[\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{X}\right]^{-1}\mathrm{X}'\mathrm{Z}(\mathrm{Z}'\mathrm{Z})^{-1}\mathrm{Z}'\mathrm{y}$$

Agregamos un segundo instrumento excluido, `nearc2` (universidad de 2 años cercana):

R

```
iv_si <- ivreg(lwage ~ educ + exper + expersq + black + south |
                . - educ + nearc4 + nearc2, data = data.ingresos)
```

	Clásicos
educ	0.2403 (0.0406)
exper	0.1517 (0.0188)

Ahora usamos la fórmula general de 2SLS: $\hat{\beta} = (X'P_Z X)^{-1}X'P_Z Y$, con Z ampliada a incluir ambos instrumentos — el resultado replica exactamente los coeficientes de `ivreg`.

Estimador Óptimo de MGM con el Paquete gmm

R

```r
gmm_opt <- gmm(lwage ~ educ + exper + expersq + black + south,
               ~ nearc4 + nearc2 + exper + expersq + black + south,
               vcov = "HAC", wmatrix = "optimal",
               type = "twoStep", data = data.ingresos)
```

Replicando con matrices (procedimiento en dos etapas):

1. Primera etapa con $W = I$ para obtener $\hat{\beta}_1$ y construir $\hat{S}$.
2. Segunda etapa con $W = \hat{S}^{-1}$ para obtener el estimador óptimo $\hat{\beta}_{GMM,O}$.

El coeficiente de educ converge a ≈ 0.239, cercano — pero no idéntico — al 2SLS (0.240), reflejando que ambos son válidos pero difieren en eficiencia bajo heterocedasticidad.

La Prueba de Hausman: ¿Importa Usar Variables Instrumentales?

También llamada prueba de Hausman, Wu–Hausman o Durbin–Wu–Hausman: compara dos estimadores $\tilde{\theta}$ y $\hat{\theta}$ que comparten el mismo límite en probabilidad bajo H_0, pero difieren bajo H_a:

$$H_0 : \ \text{plim}(\tilde{\theta} - \hat{\theta}) = 0 \qquad \text{vs.} \qquad H_a : \ \text{plim}(\tilde{\theta} - \hat{\theta}) \neq 0$$

La Prueba de Hausman: ¿Importa Usar Variables Instrumentales?

También llamada prueba de Hausman, Wu–Hausman o Durbin–Wu–Hausman: compara dos estimadores $\tilde{\theta}$ y $\hat{\theta}$ que comparten el mismo límite en probabilidad bajo H_0, pero difieren bajo H_a:

$$H_0 : \ \text{plim}(\tilde{\theta} - \hat{\theta}) = 0 \qquad \text{vs.} \qquad H_a : \ \text{plim}(\tilde{\theta} - \hat{\theta}) \neq 0$$

Estadístico de prueba

$$H = (\tilde{\theta} - \hat{\theta})' \left[\hat{V}(\tilde{\theta} - \hat{\theta}) \right]^{-1} (\tilde{\theta} - \hat{\theta}) \ \overset{a}{\sim} \ \chi^2(q)$$

Se rechaza H_0 si $H > \chi^2_\alpha(q)$.

La Prueba de Hausman: ¿Importa Usar Variables Instrumentales?

También llamada prueba de Hausman, Wu–Hausman o Durbin–Wu–Hausman: compara dos estimadores $\tilde{\theta}$ y $\hat{\theta}$ que comparten el mismo límite en probabilidad bajo H_0, pero difieren bajo H_a:

$$H_0 : \ \text{plim}(\tilde{\theta} - \hat{\theta}) = 0 \qquad \text{vs.} \qquad H_a : \ \text{plim}(\tilde{\theta} - \hat{\theta}) \neq 0$$

Estadístico de prueba

$$H = (\tilde{\theta} - \hat{\theta})' \left[\hat{V}(\tilde{\theta} - \hat{\theta}) \right]^{-1} (\tilde{\theta} - \hat{\theta}) \ \overset{a}{\sim} \ \chi^2(q)$$

Se rechaza H_0 si $H > \chi^2_\alpha(q)$.

La implementación se complica porque $\hat{V}(\tilde{\theta} - \hat{\theta}) = \hat{V}(\tilde{\theta}) - \hat{V}(\hat{\theta}) - 2\,\widehat{\text{Cov}}(\tilde{\theta}, \hat{\theta})$.

Aplicación para caso con Instrumento:

- Bajo H_0 (x es exógeno):
- Bajo H_a (x es endógeno):

La Prueba de Hausman: Versión Simplificada

Con errores homocedásticos, MCO es eficiente, y puede mostrarse que el término de covarianza se cancela:

lo cual es fácil de calcular en cualquier software.

- Sin homocedasticidad, hay que estimar $\widehat{\text{Cov}}(\tilde{\theta}, \hat{\theta})$ explícitamente — la mayoría de los paquetes estadísticos ya lo implementan.
- Esta prueba (simplificada) sirve para comparar cualquier par de estimadores donde uno es más eficiente que el otro (no solo MCO vs. VI).

Prueba de Hausman con Regresión Auxiliar

Consideremos $y = x_1\beta_1 + x_2\beta_2 + u$, con x_1 endógena y x_2 exógena. La **regresión auxiliar** es

$$y = x_1\gamma_1 + x_2\gamma_2 + \hat{v}\,\gamma_3 + \varepsilon, \qquad \hat{v} = x_1 - \hat{x}_1$$

donde $\hat{x}_1$ son los valores ajustados de la primera etapa $x_1 = z\pi_1 + x_2\pi_2 + v$.

- Si x_1 está correlacionada con u, entonces v también lo está: $u = v\gamma_3 + \varepsilon$.
- El test de Hausman equivale a contrastar $H_0 : \gamma_3 = 0$.
- Rechazar $H_0 \Rightarrow$ evidencia de correlación entre x_1 y u (es decir, de endogeneidad).

Prueba de Sobreidentificación (Hansen/Sargan)

También conocida como prueba de Hansen (forma general) o de Sargan (forma particular para VI lineal): evalúa qué tan cerca está de cumplirse $E[h(w, \theta_0)] = 0$.

Estadístico J de Hansen (1982)

$$J = \left(\frac{1}{N} \sum_i \hat{h}_i \right)' \hat{S}^{-1} \left(\frac{1}{N} \sum_i \hat{h}_i \right) \overset{a}{\sim} \chi^2(r - q)$$

- J es simplemente la función objetivo de MGM evaluada en $\hat{\theta}_{MGM}$.
- Un valor grande de $J \Rightarrow$ se rechaza que las condiciones de momentos poblacionales se cumplan $\Rightarrow$ el estimador de MGM sería inconsistente.
- Modelos exactamente identificados:

Estadístico J para Variables Instrumentales

En el caso lineal de VI, el estadístico toma la forma específica

$$J = \hat{u}' Z \hat{S}^{-1} Z' \hat{u}, \qquad \hat{u} = y - X\hat{\beta}_{MGM}$$

Interpretación

Rechazar H_0 sugiere que los instrumentos son endógenos — aunque también podría deberse a una mala especificación del modelo. Rechazar H_0 indica que **algo** está mal, pero no dice exactamente **qué**.

MCO Insesgado vs. MC2E Sesgado

Discusión intuitiva en Angrist & Pischke, *Mostly Harmless Econometrics* (2009).

- MCO es tanto **consistente** como **insesgado**: en una muestra de cualquier tamaño, la distribución de $\hat{\beta}_{OLS}$ está centrada exactamente en β.
- 2SLS/MC2E es **consistente pero sesgado** en muestras finitas.
- En muestras grandes el estimador de 2SLS está *cerca* del valor poblacional, pero el sesgo en muestras finitas tiene consecuencias importantes para la estimación e inferencia.

Sesgo del Estimador de 2SLS: Derivación

Modelo simple con un regresor endógeno $y = \beta x + \mu$, primera etapa $x = Z\pi + \xi$. El estimador de 2SLS es

$$\hat{\beta}_{MC2E} = \beta + (x' P_Z x)^{-1} x' P_Z \mu$$

Sustituyendo $x = Z\pi + \xi$:

Como la esperanza no es lineal en general, Angrist & Pischke (2009) aproximan el sesgo:

$$E(\hat{\beta}_{MC2E} - \beta) \approx \left[E(x' P_Z x)\right]^{-1} E(\pi' Z' \mu) + \left[E(x' P_Z x)\right]^{-1} E(\xi' P_Z \mu)$$

Sesgo del Estimador de 2SLS: la Regla del $F > 10$

La expresión del sesgo puede reescribirse como

$$E(\hat{\beta}_{MC2E} - \beta) \approx \frac{\sigma_{\mu\xi}}{\sigma_{\xi}^2} \cdot \frac{1}{F+1}$$

donde $\sigma_{\mu\xi}/\sigma_{\xi}^2$ es precisamente el **sesgo de MCO**.

- Si $\pi = 0$ (instrumento totalmente irrelevante):
- Cuando F es pequeña (instrumento débil), el sesgo de 2SLS se acerca al de MCO.
- Staiger & Stock (1997) mostraron por simulación que con $F > 10$, el sesgo máximo de 2SLS es de apenas el 10% del sesgo de MCO.

Origen de la regla práctica

De aquí surge la conocida regla de dedo $F > 10$ para descartar instrumentos débiles.

Recomendaciones Prácticas

1. **Reportar la primera etapa** y verificar que los coeficientes tengan sentido económico.
2. **Reportar el estadístico** F de la primera etapa para los instrumentos excluidos.
3. **Reportar también el modelo exactamente identificado**, usando el *mejor* instrumento disponible.
4. **Prestar atención a la forma reducida**: es proporcional al efecto causal de interés.

> *"Si no puedes ver la relación causal de interés en la forma reducida, es porque probablemente no haya nada ahí."*
>
> — Angrist & Krueger (2001)

5. **Más recientemente, también reportar**:
   - **Estadístico F de Wald de Kleibergen-Paap**: prueba de instrumentos débiles robusta a heterocedasticidad y *clustering* ($F > 10$).
   - **Estadístico LM de Kleibergen-Paap**: prueba de subidentificación robusta a heterocedasticidad y *clustering* (buscar un valor-p bajo).

Síntesis

- El álgebra de matrices detrás de VI/2SLS/MGM óptimo se puede replicar directamente y coincide exactamente con lo que reportan paquetes como `ivreg` y `gmm`.
- En el modelo exactamente identificado, **la elección de W_N no importa**: siempre se obtiene el estimador de VI.
- La prueba de **Hausman** contrasta si vale la pena usar VI en lugar de MCO (evidencia de endogeneidad); la prueba de **Hansen/Sargan** contrasta la validez de las restricciones de sobreidentificación.
- Con **instrumentos débiles** (F pequeña), 2SLS puede estar tan sesgado como MCO — de ahí la regla práctica $F > 10$.