NOTA TÉCNICA 44

Preparada por

Wilson Pérez Oviedo

UNA APROXIMACION AL METODO GENERALIZADO DE LOS MOMENTOS Y SUS LIMITACIONES

1. Introducción

El método generalizado de los momentos (GMM) es un poderoso instrumento de estimación de parámetros estadísticos. En la actualidad se conocen las propiedades asintóticas de los estimadores obtenidos por este método, los cuales, bajo supuestos no muy restrictivos, son consistentes y con funciones de distribución fácilmente calculables. Herramientas ampliamente divulgadas y utilizadas en econometría, tales como los mínimos cuadrados ordinarios, mínimos cuadrados generalizados, estimación en dos etapas e incluso (bajo algunos supuestos adicionales) máxima verosimilitud, pueden ser considerados casos particulares de GMM.

Otra de las características del método es que no requiere la especificación de una forma particular de distribución de las variables aleatorias involucradas en el modelo que se estudia. Sin embargo, esta generalidad también puede significar un uso no eficiente de la información disponible en la muestra (Hamilton,1994, p. 409). Además, recientes desarrollos demuestran que en muestras pequeñas los estimadores pueden estar bastante lejos del valor real del parámetro (Chumacero 1997a), por lo que no es aconsejable su uso cuando estén disponibles solamente un reducido número de observaciones, situación tan frecuente en econometría.

En el presente documento, primero se hace una breve descripción del método generalizado de los momentos (sección 2), luego se estudian como casos particulares de éste a algunas herramientas de uso común en la profesión (sección 3). Finalmente, en la sección 4, se hace una aplicación de GMM a un modelo con un agente representativo con expectativas racionales, evaluando sus propiedades en muestras pequeñas mediante simulaciones de Montecarlo.

2. El método generalizado de los momentos

Sea g_i función de $\boldsymbol{Xt}$**,** vector aleatorio (fila) en el tiempo t, y de $\boldsymbol{q}$, un vector de coeficientes desconocidos que se pretende estimar, es decir:

$$g_i(X_t,\theta) \quad i=1..k$$

Si, suponiendo que $\boldsymbol{q_0}$ es el verdadero valor de $\boldsymbol{q}$, se puede asegurar que se cumplen las igualdades:(1)

$$E\left[g_i(X_t,\theta_0)\right]=0 \quad i=1..k$$

denominadas "condiciones de ortogonalidad", en las que (1) se sustenta el método de los momentos generalizados. Este tipo de condiciones pueden originarse en proposiciones justificadas en los datos, como por ejemplo $E(\boldsymbol{X_t}\, e_t)=\boldsymbol{0}$, para el modelo lineal clásico $Y_t=\boldsymbol{X_t B}+e_t$; o también en construcciones *ad-hoc* sobre los datos, como en el uso de variables instrumentales $\boldsymbol{Z_t}$: $E(\boldsymbol{Z_t}\, e_t)=0$, para el mismo modelo lineal mencionado. Pero, posiblemente, las condiciones de ortogonalidad de mayor interés son aquellas que surgen de las ecuaciones de Euler que se obtienen de problemas de optimización intertemporal, típicos de la microeconomía y de la macroeconomía con bases micro.

Por supuesto, las condiciones establecidas en (1) deben estar sustentadas teóricamente. Pero, en la práctica la estimación tiene que realizarse usando un conjunto finito de observaciones $\boldsymbol{X_T}=\{X_t\}$ $(t=1..T)$.

Es decir, que si las igualdades en (1) son verdaderas, la estimación del valor de $\boldsymbol{q_0}$ debe basarse en la media muestral:(2) $$f_i(X_T,\theta) = \frac{1}{T}\sum_{t=1}^{T} g_i(X_t,\theta)$$

Si el número de condiciones de ortogonalidad es igual al número de parámetros a estimar (es decir k=l, donde l es la dimensión de $\boldsymbol{q}$), la estimación de q mediante GMM corresponde a la solución del sistema de ecuaciones:

(3) $$f_i(X_T,\theta) = 0$$

Por cierto, el caso más frecuente no es éste; lo es cuando el número de parámetros es menor que el número de condiciones de ortogonalidad. En esta situación es imposible que todas las igualdades (3) se cumplan, y corresponde entonces buscar el valor de $\boldsymbol{q}$ que "acerque más" a cero al vector:

(4) $$f(X_T,\theta) = \begin{bmatrix} f_1(X_T,\theta) \\ f_2(X_T,\theta) \\ \cdot \\ f_k(X_T,\theta) \end{bmatrix}$$

Establecer "qué tan cerca" está el vector de cero requiere de la definición de una métrica. Una expresión muy general de una métrica es:

$$\left[f(X_T,\theta)\right]' M\left[f(X_T,\theta)\right]$$

donde $\boldsymbol{M}$ es una matriz definida positiva y $\boldsymbol{f'}$ es el vector $\boldsymbol{f}$ traspuesto. El problema es, por lo tanto, la definición de la matriz $\boldsymbol{M}$. Hansen (1982) demostró que la mejor elección de $\boldsymbol{M}$ es la matriz de varianza-covarianza del vector $\boldsymbol{f}$, aquí notada por **S**.

Según Hamilton (1994, p.413), si la serie multidimensional de tiempo $\{f(X_T,q_0)\}$ $T = -\infty \ldots \infty$ es serialmente no correlacionada, la matriz $\boldsymbol{S}$ puede ser estimada de manera consistente de la siguiente forma:

(5) $$\hat{S}_T = \frac{1}{T}\sum_{t=1}^{T}\left[f(X_t,\hat{\theta}_0)\right]\left[f(X_t,\hat{\theta}_0)\right]'$$

Sin embargo, se debe notar que para estimar **S** es necesario contar con la estimación de $\boldsymbol{q_0}$, y viceversa. Para salir de esta trampa circular, Chumacero (1997,b) sugiere el siguiente proceso iterativo:

$\hat{S}_T^0 = I_k$ (matriz identidad de orden k)

$$\hat{\theta}_0^1 = \arg\min_{\theta}\left\{\left[f(X_T,\theta)\right]'\left[\hat{S}_T^0\right]^{-1}\left[f(X_T,\theta)\right]\right\}$$

y de ahí en adelante:

(6) $$\hat{\theta}_0^k = \arg\min_{\theta}\left\{\left[f(X_T,\theta)\right]'\left[\hat{S}_T^{k-1}\right]^{-1}\left[f(X_T,\theta)\right]\right\}$$

(7) $$\hat{S}_T^k = \frac{1}{T}\sum_{t=1}^{T}\left[f(X_t,\hat{\theta}_0^k)\right]\left[f(X_t,\hat{\theta}_0^k)\right]'$$

El proceso continuaría hasta que $\hat{\theta}_0^{k-1} \approx \hat{\theta}_0^k$, de acuerdo a alguna medida de distancia de vectores que se haya elegido.

Las propiedades asintóticas de los estimadores de GMM son muy buenas. Tal como lo establece Hamilton (1994), el método general de los momentos ofrece estimadores consistentes, bajo supuestos generales de estacionariedad, continuidad de las funciones y las condiciones establecidas para los momentos. El mismo autor, en base al trabajo de Hansen (1982), afirma que se puede tratar al estimador de ***q***, aproximadamente como:

$$\hat{\theta}_T \approx N(\theta_0, \hat{V}_T / T)$$

donde:

$$V_T = \left[\hat{D}_T\left[\hat{S}_T\right]^{-1}\hat{D}_T'\right]^{-1}$$ y $$\hat{D}_T^I = \left[\left.\frac{\partial f_i}{\partial \theta_j}\right|_{\theta=\hat{\theta}_T}\right]_{i=1..k}^{j=1..l}$$

Estas excelentes propiedades de GMM, lastimosamente, no siempre están presentes en la aplicación de este método a muestras pequeñas. Chumacero (1997a) sintetiza así los problemas que se pueden presentar, y que han sido ya ampliamente estudiados: utilizar GMM para contrastar hipótesis puede llevar a rechazar hipótesis ciertas; en algunos casos se presentan desviaciones importantes en las estimaciones de los parámetros; y, los resultados mejoran lentamente ante incrementos del tamaño de la muestra.

3. Algunos casos particulares de GMM

Algunos de los métodos econométricos de más amplia difusión pueden ser planteados como casos especiales del método generalizado de los momentos. Si se plantea un modelo lineal :

(8) $Yt=X_tB+e_t$

se debe cumplir que

(9) $E(X_t\, e_t)=0$

como requisito para que **B** estimado mediante mínimos cuadrados ordinarios sea insesgado. Esta condición es el resultado de una especificación correcta del modelo, y permite escribir la condición de ortogonalidad (usando la ecuación 8):

$$E\left[X_t(y_t - X_tB)\right] = 0 \quad t=1..T$$

Expresando el concepto en términos de la media muestral, podríamos obtener el vector:

$$\frac{1}{T}X'(y-XB)$$

Nótese que se trata de un sistema lineal con igual número de incógnitas y de ecuaciones, entonces, la estimación de ***B*** por GMM correspondería a:

$$\frac{1}{T}X'(y-X\hat{B})=0$$

Simples transformaciones matriciales llevan a:

(10) $\hat{B}=(X'X)^{-1}X'y$

que corresponde exactamente al ***B*** estimado por el método de mínimos cuadrados ordinarios. Si, por otro lado, se utilizan variables instrumentales ***Z***, tal que **Z** está correlacionada con ***X*** pero no correlacionada con ***e***, la condición de ortogonalidad es:

(11) $E(Z_t\, e_t)=0$

siguiendo el mismo procedimiento anterior, es fácil comprobar que GMM conduce en este caso al estimador de ***B*** del método de las variables instrumentales, que por tanto también puede ser considerado un caso particular de GMM:

(12) $\hat{B}=(Z'X)^{-1}Z'y$

Sin embargo, nada asegura que el número de variables instrumentales usadas (columnas de ***Z***) sea igual al número de variables originales (columnas de ***X***). Si son iguales, la matriz ***(Z'X)*** es cuadrada y su inversa puede existir. Si hay más variables instrumentales que originales, sobran restricciones y la matriz ***(Z'X)*** ni siquiera es cuadrada. Una solución no óptima, pues significa perder información, es eliminar algunas restricciones. La otra, recordar que para estos casos GMM propone retomar el vector de condiciones de normalidad y, de acuerdo a cierta métrica, minimizarlo. De (8) las condiciones de ortogonalidad aplicadas a la muestra son:

$$\frac{1}{T}\left[Z'(y-XB)\right]$$

El estimador de ***B*** correspondería a:

$$\hat{B}=\arg\min_{B}\left\{\frac{1}{T^2}\left[Z'(y-XB)\right]'[\hat{S}]^{-1}\left[Z'(y-XB)\right]\right\}$$

Tal como se plantea en (7). Derivado la ecuación matricial, se obtienen las condiciones de primer orden:

$$(X'Z)[\hat{S}]^{-1}(Z'y-Z'X\hat{B})=0$$

de donde se puede despejar la estimación de **B**:

(13) $\hat{B}=\left[X'Z[\hat{S}]^{-1}Z'X\right]^{-1}X'Z[\hat{S}]^{-1}Z'y$

Únicamente resta calcular la estimación de la varianza de *(1/T)(**Z'**e)*, para la cual Hamilton(1994) sugiere, asumiendo errores independientes e idénticamente distribuidos:

$$\hat{S} = \frac{\hat{\sigma}^2}{T^2} Z'Z$$

que reemplazando en (13) conduce al estimador de mínimos cuadrados en dos etapas:

$$\hat{B} = \left[X'Z[Z'Z]^{-1}Z'X \right]^{-1} X'Z[Z'Z]^{-1}Z'y \tag{14}$$

También el método de máxima verosimilitud puede ser obtenido desde la perspectiva del GMM. Para ello, es necesario considerar una función de distribución:

$$h(\theta; y_t)$$

donde y_t es una variable aleatoria en el tiempo *t*. Por supuesto, cualquier función de densidad que se respete debe cumplir con la condición:

$$\int_{-\infty}^{+\infty} h(\theta, y_t) dy_t = 1$$

Derivando la igualdad:

$$\int_{-\infty}^{+\infty} \frac{\partial h(\theta, y_t)}{\partial \theta} dy_t = 0 \tag{15}$$

Esta última igualdad puede ser reescrita de la siguiente forma:

$$\int_{-\infty}^{+\infty} \frac{\partial \ln h(\theta, y_t)}{\partial \theta} h(\theta, y_t) dy_t = 0 \tag{16}$$

Es decir, definiendo:

$$g(\theta, Y_t) = \frac{\partial \ln h(\theta; y_t)}{\partial \theta}$$

la ecuación (15) no es otra cosa que:

$$E\left[g(\theta, Y_t)\right] = 0$$

Que es una condición de ortogonalidad, la cual en términos de muestra finita que va desde *t=1..T* pueden escribirse dentro del método de GMM, de la siguiente forma:

$$\sum_{t=1}^{T} \frac{\partial \ln h(\theta, y_t)}{\partial \theta} = 0 \tag{17}$$

Se observa que este es precisamente el sistema de ecuaciones (casi siempre no lineales) que permite obtener el estimador de máxima verosimilitud.

4. Aplicación de GMM a un modelo dinámico de expectativas racionales: propiedades en muestras pequeñas

En esta sección se intenta responder a la siguiente pregunta: suponiendo que la economía funciona como un modelo simple dinámico de agentes racionales ¿cuantos datos son suficientes para obtener buenas estimaciones de los parámetros del modelo usando el método generalizado de los momentos? Se tratará de dar una respuesta en base a simulaciones de Montecarlo.

4.1 El modelo teórico

La economía a estudiar está compuesta por individuos representativos idénticos, que tienen una función de utilidad con coeficiente de aversión al riesgo relativamente constante:

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma} \qquad \gamma \neq 1$$

El agente representativo busca maximizar:

$$U_0 = \mathrm{E}_0\left[\sum_{t=0}^{\infty} \beta^t u(c_t)\right]$$

Donde *b* es el factor de descuento intertemporal, que se supone menor que uno y mayor que cero. E_0 es el operador esperanza condicional a la información disponible en el tiempo 0. El agente recibe una dotación aleatoria y_t y se sujeta a la siguiente restricción presupuestaria:

$$c_t + b_t = y_t + (1 + r_t)b_{t-1}$$

donde b_{t-1} es una cantidad de activos de precio invariable que rinden una tasa de interés variable r_t , fijada en el período *t-1*. El agente debe decidir cuánto ahorrar en estos activos. Si se resuelve el problema de optimización dinámica, la condición de Euler es:

(18) $$\frac{1}{1+r_{t+1}} = \mathrm{E}_0\left[\beta\left(\frac{y_{t+1}}{y_t}\right)^{-\gamma}\right]$$

que fácilmente puede escribirse como una condición de ortogonalidad:

(19) $$\mathrm{E}_0\left[\beta\left(\frac{y_{t+1}}{y_t}\right)^{-\gamma} - \frac{1}{1+r_{t+1}}\right] = 0$$

Para simular numéricamente la economía modelada, y así obtener series de r_t e y_t, se requiere hacer explícita la ley que gobierna el proceso aleatorio de y_t, del cual suponemos que su tasa de crecimiento es estacionaria y sigue un proceso AR(1):

(20) $$\ln\left(\frac{y_{t+1}}{y_t}\right) = \alpha_0 + \alpha_1\left(\frac{y_{t1}}{y_{t-1}}\right) + \varepsilon_t \qquad \varepsilon_t \rightarrow N(0,\sigma^2)$$

Tomando dos valores iniciales de y_t ($y_{1,}$ y_2) la estimación de r_3 se basa en la expresión 18, para lo cual se debe calcular el miembro derecho de la igualdad. Este valor se lo estima generando L valores aleatorios de e_t , que dan lugar a otros tantos valores de ln($y_{3/}$ y_2). Luego de las correspondientes operaciones de exponenciación y producto, se tiene una muestra aleatoria:

$$\left\{ \beta \left(\frac{y_{3,j}}{y_2} \right)^{-r} \right\} \qquad j = 1..L \tag{21}$$

El promedio simple de estas cifras permite hallar una estimación del valor esperado que se busca. Usando 17 se puede calcular r_3 .

El procedimiento descrito se puede aplicar de manera secuencial hasta simular series de (y_t, r_t) de cualquier tamaño, dígase N. De esta forma, se pueden generar H simulaciones. Es decir se cuenta con las siguientes series:

$$\left\{ (y_t, r_t) \,/\, t = 1..N \right\}_l \qquad l = 1...H \tag{22}$$

Recuérdese que la pregunta es ¿cuál es el tamaño mínimo de muestra que "asegura" una buena estimación de b y g, suponiendo que la economía se rige por el modelo expuesto? Con este fin para cada uno de los valores de T, se toma una muestra de ese tamaño en cada una las simulaciones. Así por ejemplo, para la j-ésima simulación se extraería la serie:

$$\left\{ (y_t^j, r_t^j) \,/\, t = 1..T \right\} \tag{23}$$

a estos datos se aplica el método generalizado de los momentos. Las estimaciones de b y g así conseguidas corresponden a una simulación y a un tamaño de muestra determinados. Hecho lo mismo para todas las simulaciones y todos los tamaños de muestra disponibles se consigue un conjunto de estimaciones:

$$\left\{ (\hat{\beta}_T^j, \hat{\gamma}_T^j) \,/\, j = 1..H; T = T_0..N \right\}$$

4.2 Los resultados numéricos

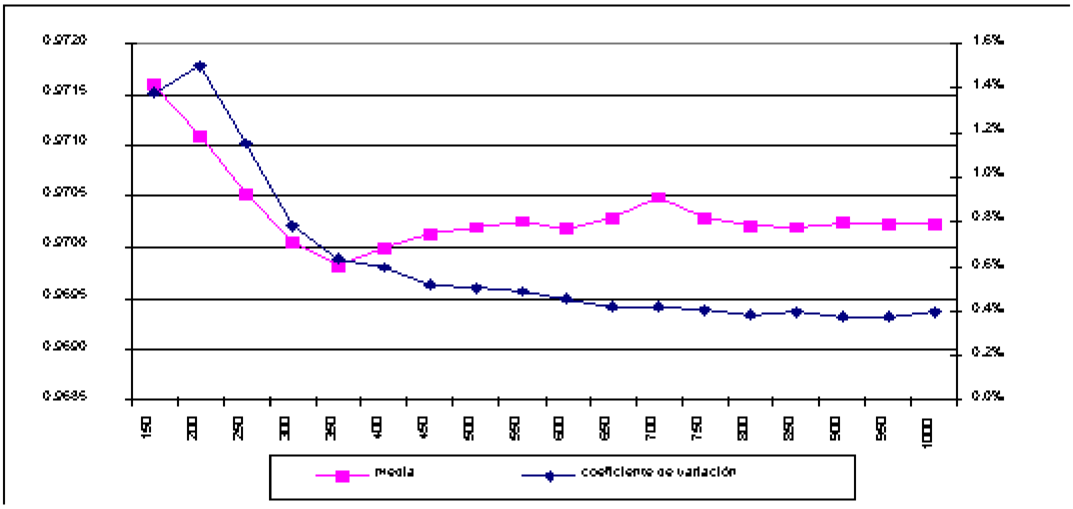
Se realizaron en total 250 simulaciones y se estimaron los parámetros para cada una de estas simulaciones, con un tamaño de muestra que va desde 150 a 1000 observaciones. Se utilizó un beta de 0.97 y un gama de 1.3. La ley que genera la tasa de crecimiento del producto es:

$$\ln\left(\frac{y_{t+1}}{y_t} \right) = 0.013 + 0.178 \left(\frac{y_{t1}}{y_{t-1}} \right) + \varepsilon_t \qquad \varepsilon_t \rightarrow N(0, 0.00144)$$

A continuación se exponen los resultados obtenidos. En el gráfico 1 se expone el promedio y el coeficiente de variación de las 250 estimaciones de beta para cada uno de los tamaños de muestra. Como era de esperarse, mientras más grande es la muestra, la media de las estimaciones se acerca más al valor verdadero de beta (0.97), mientras que el coeficiente de variación se reduce en forma continua. Más aún, la media de las estimaciones rápidamente se sitúa en valores muy cercanos al valor verdadero y la dispersión es muy baja. En efecto, con tan solo 150 puntos, los valores son muy cercanos: 0.9716 de promedio y 1.4% de coeficiente de variación, y con mil observaciones, los valores obtenidos por GMM son muy precisos, 0.9702 y 0.4%, respectivamente.

Gráfico 1

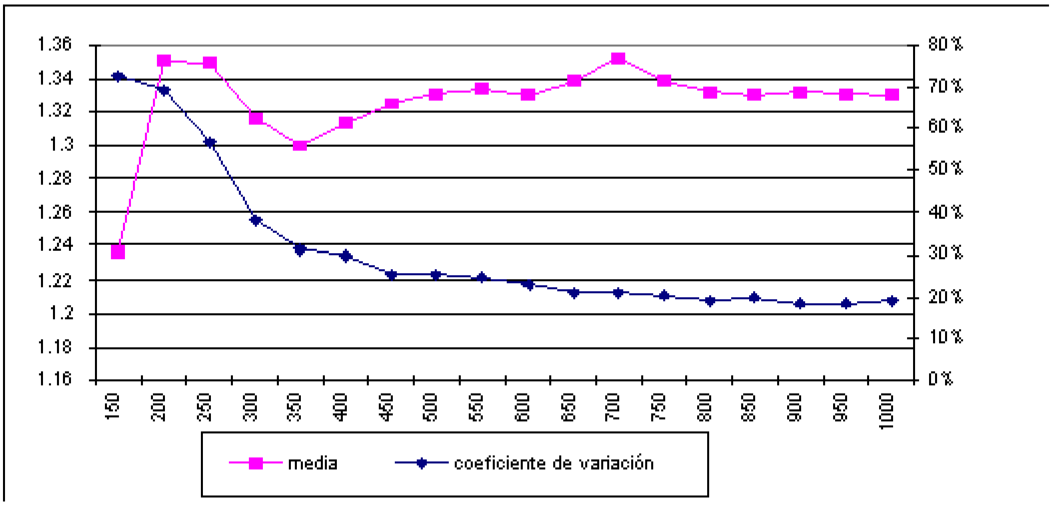
Promedio (izq.) y coeficiente de variación (der.) de las estimaciones de Beta, según tamaño de muestra.



Fuente y elaboración: autor.Para el coeficiente gama, la historia es muy distinta. El promedio de las estimaciones se mueve muy lentamente hacia el valor real (1.3), y la dispersión de los gama estimados es muy alta. En efecto, con 150 puntos, la media de las estimaciones es un cercano 1.23, pero la dispersión de las mismas es muy alta: 72%. Los resultados no mejoran, sino muy lentamente, e incluso con una muestra de tamaño mil, el coeficiente de variación tiene niveles altos (19%), a pesar de que la media tiene valores aceptables (1.329).

Gráfico 2

Promedio (izq.) y coeficiente de variación (der.) de las estimaciones de Gama, según tamaño de muestra.



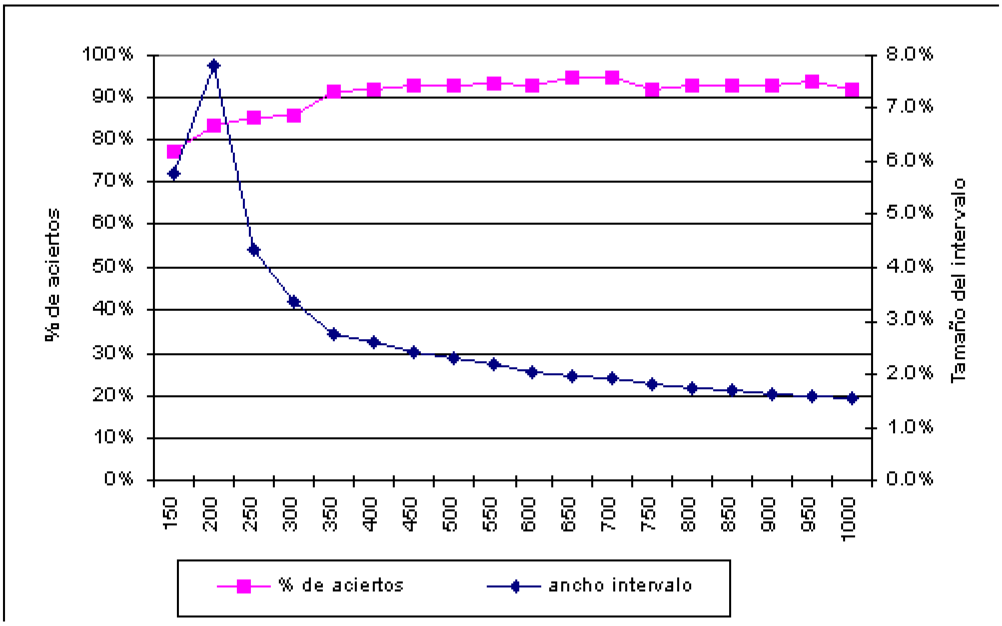
Fuente y elaboración: autor.

Otra forma de medir la bondad del método generalizado de los momentos, es realizar estimaciones de los intervalos de confianza, en lugar de puntuales. Por supuesto, intervalos excesivamente anchos no son muy útiles, y el tamaño promedio de aquellos puede ofrecer una prueba de la efectividad del método. Adicionalmente, es conveniente contabilizar cuántas veces el valor real del parámetro fue interior al intervalo de confianza calculado.

En el caso del coeficiente beta, la precisión de las estimaciones se confirma. De esta forma, el ancho del intervalo de confianza (al 95%), como proporción del valor real del parámetro, no es alto desde el inicio (un máximo de 8% con un tamaño de muestra de 200) y se reduce rápidamente, hasta llegar a un 1.6% con mil observaciones. La precisión de la estimación es impresionante, pero contrasta vivamente con los resultados que se obtienen para el coeficiente de aversión al riesgo (gama).

Gráfico 3

Ancho promedio del intervalo de confianza y porcentaje de aciertos de las estimaciones de beta, según tamaño de muestra



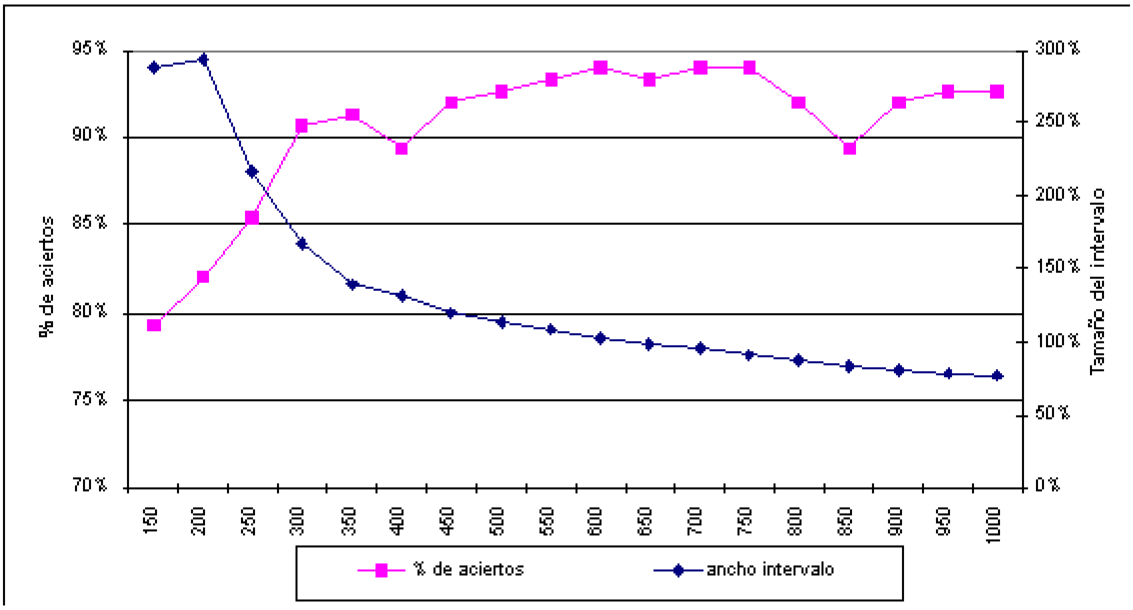
Fuente y elaboración: autor.

Para el coeficiente de aversión al riesgo, los resultados son desalentadores, aún con tamaños grandes de muestra. Así, para 150 observaciones, el porcentaje de veces que el intervalo calculado incluye al valor real de 1.3 es del 79%, pero el ancho del intervalo es el 288% del valor real. Es decir, la gran mayoría de los intervalos incluye a cero, y a números negativos, entre los valores posibles del coeficiente de aversión al riesgo, lo cual no solo implica diversos órdenes de magnitud, sino diferencias cualitativas importantes en el comportamiento de los agentes. Concretamente, no se podría ni siquiera definir si los agentes son amantes al riesgo (gama negativo), neutrales (gama cero) o adversos al riesgo (gama mayor que cero).

Para tener una idea clara de la actitud frente al riesgo por parte del agente representativo de esta economía, es necesario recurrir a muestras de mayor tamaño. En efecto, apenas desde las 650 observaciones el tamaño promedio del intervalo de confianza es menor al 100%, y con mil de ellas el ancho promedio del intervalo es de 77%.

Gráfico 4

Ancho promedio del intervalo de confianza y porcentaje de aciertos de las estimaciones de Gama, según tamaño de muestra



5. Conclusiones

Varios autores han demostrado que los estimadores del método generalizado de los momentos (GMM), bajo supuestos no muy restrictivos, son consistentes y sus funciones de distribución fácilmente calculables. Las propiedades asintóticas de los parámetros obtenidos por GMM son bien conocidas y también de fácil cálculo. Además, herramientas ampliamente divulgadas en econometría pueden entenderse como versiones particulares de GMM.

A pesar de estas cualidades, las propiedades de los estimadores obtenidos por el método generalizado de los momentos no son siempre buenas en muestra pequeñas. Más aún, para ciertos parámetros, las estimaciones mejoran muy lentamente conforme aumenta el tamaño de la muestra.

Así, dependiendo del parámetro a estimar, el GMM puede o no ofrecer resultados adecuados. Por supuesto, esto no significa que el método deba ser dejado de lado. Posiblemente un mecanismo para adquirir algún nivel de seguridad en los resultados sea modelar teóricamente el problema que se estudia, luego simularlo y estudiar las propiedades de GMM en diferentes tamaños de muestra de esas simulaciones. De esta manera, se conocería qué parámetros están siendo estimados con precisión por el método generalizado de momentos y cuáles no, que tamaño de muestra mínimo es necesario para tener intervalos de confianza adecuados y qué conclusiones se puede obtener de los resultados.

En definitiva, si se ha de aplicar el método generalizado de los momentos, el tamaño adecuado de la muestra depende del tipo de problema, del parámetro de interés y de la precisión que se requiera en las estimaciones. La mejor forma de determinar el número de observaciones necesarias, puede ser realizar simulaciones de Montecarlo, tal como se ha hecho en el presente documento.

6. Bibliografía

Blanchard,O. y Fischer S. (1989) "Lectures in macroeconomics", MIT Press.

Chumacero, Rómulo. (1997a) "Finite sample propierties of the efficient method of moments", mimeo, Universidad de Chile.

Chumacero, Rómulo (1997b) Notas de clase del curso de Macro-econometría, Banco Central del Ecuador.

Greene, William. (1993) "Econometric Analysis", 2da. Edición, MacMillan.

Johnston, J. Y Dinanrdo (1997) "Econometric methods", 4ta. Edición, MacGraw Hill.

Kocherlakota, Narayama (1990). "On test of representative consumer asset pricing models", Journal of Monetary Economics, N. 26.

Hansen, Lars. (1982) "Large sample properties of generalized method od moments estimators", Econometrica, Vol 50, N.4.

Hamilton, James. (1994) "Time series analysis", Pricenton University Press.

Tauchen, G. y Hussey R. (1991) "Quadrature-Based methods for obtaining aproximate solutions to nonlinear asset pricing models", Econometrica 59.

Turnovsky, Stephen. (1995) "Methods of macroeconomic dynamics", The MIT Press.