Econometría I

Multicolinealidad

Miguel Jerez y Sonia Sotoca

Universidad Complutense de Madrid

Noviembre, 2007

Índice

- Introducción
- Efectos de la colinealidad
- Casos en que suele haber problemas de colinealidad
- Criterios de diagnóstico
- Soluciones

Introducción

El término multicolinealidad (o colinealidad) en Econometría se refiere a una situación en la que dos o más variables explicativas están fuertemente interrelacionadas y, por tanto, resulta difícil medir sus efectos individuales sobre la variable endógena. Cabe distinguir dos casos:

- **Multicolinealidad exacta**, cuando $\left|\boldsymbol{X}^T\boldsymbol{X}\right| = 0$. En este caso existen infinitas soluciones para el sistema $\boldsymbol{X}^T\boldsymbol{X}\hat{\beta}_{MCO} = \boldsymbol{X}^T\boldsymbol{y}$
- **Multicolinealidad de grado**, en este caso $\left|\boldsymbol{X}^T\boldsymbol{X}\right| \simeq 0$ y, por tanto, existe una solución formalmente óptima al problema de mínima suma de cuadrados. Sin embargo, esta solución está *mal condicionada*, ya que la función objetivo es muy plana en el entorno del óptimo y, por tanto, existen infinitas soluciones casi tan buenas como la óptima.

Para presentar este tema, seguiremos el siguiente esquema:

- Efectos de la multicolinealidad.
- Casos en que suele presentarse un problema de multicolinealidad.
- Criterios para decidir cuándo la colinealidad de grado constituye un problema.
- Soluciones al problema.

Efectos de la colinealidad (I)

El efecto fundamental de la colinealidad exacta es que no existe una solución única del sistema de ecuaciones normales.

Cuando la colinealidad es de grado:

- Las estimaciones individuales de los parámetros están mal identificadas
- Se produce una inflación de la varianza de las estimaciones.
- Las estimaciones resultan muy sensibles a la muestra.

Mala identificación de las estimaciones. Por ejemplo, sea el modelo:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t \qquad (1)$$

en donde:

$$x_{t2} = \alpha_1 x_{t1} + u_t \qquad (2)$$

Sustituyendo (2) en (1) se obtiene:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2(\alpha_1 x_{t1} + u_t) + \varepsilon_t = \beta_0 + (\beta_1 + \beta_2\alpha_1) x_{t1} + \beta_2 u_t + \varepsilon_t$$

y, si la varianza de u_t es "pequeña", el parámetro de x_{t2} estará mal identificado, ya que esta variable aporta poca información que no esté ya contenida en x_{t1}. En el límite, si la varianza de u_t fuera nula, tendríamos un problema de colinealidad exacta.

Efectos de la colinealidad (II)

Inflación de la varianza de las estimaciones. Como:

$$cov(\hat{\beta}_{MCO}) = \sigma_\varepsilon^2 (\boldsymbol{X}^T \boldsymbol{X})^{-1} = \sigma_\varepsilon^2 \frac{1}{\left|\boldsymbol{X}^T \boldsymbol{X}\right|} adj\left(\boldsymbol{X}^T \boldsymbol{X}\right)^T$$

si $\left|\boldsymbol{X}^T \boldsymbol{X}\right| \simeq 0$ entonces las varianzas de los parámetros tenderán a ser mayores que en una situación bien condicionada. Por tanto, los contrastes de hipótesis serán menos precisos y, concretamente, puede ocurrir que se consideren no significativos parámetros que lo serían si la colinealidad fuera menor.

Estimaciones sensibles a la muestra. Puesto que la función objetivo (suma de cuadrados de residuos) es muy plana en el entorno del óptimo, pequeños cambios en los valores de $\boldsymbol{y}$ o de $\boldsymbol{X}$ pueden dar lugar a cambios importantes en las estimaciones.

Casos en que suele haber problemas de colinealidad

Resulta frecuente que surja un problema de colinealidad en los siguientes casos:

- **En modelos de series temporales, cuando se emplean variables explicativas con tendencia**.
- **En modelos de series temporales, cuando se incluyen como variables explicativas retardos sucesivos de la variable endógena o de alguna de las variables explicativas**. Esto provoca colinealidad porque los valores de una variable económica en distintos instantes de tiempo suelen estar correlados entre sí.
- **Cuando se consideran muchas variables explicativas**. Lógicamente, a medida que aumenta el número de variables explicativas, es más fácil que aparezca una relación entre ellas, que de lugar a un problema de colinealidad.
- **En modelos con variables cualitativas**. Por ejemplo, en el modelo:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t \ ; \ x_{t1} = \begin{cases} 1 \text{ si } t = 1,2,\ldots,n_1 \\ 0 \text{ en caso contrario} \end{cases} \quad ; \quad x_{t2} = 1 - x_{t1}$$

surge un problema de colinealidad exacta.

Criterios de diagnóstico

Para decidir si la colinealidad de grado constituye un problema debemos tener en cuenta los objetivos de nuestro análisis concreto. Por ejemplo, la colinealidad no nos preocupa demasiado si nuestro objetivo es predecir, pero es un problema muy grave si el análisis se centra en interpretar las estimaciones de los parámetros.

Para diagnosticar este problema estudiaremos dos métodos: a) los basados en la correlación entre variables explicativas, y b) los basados en el tamaño de $\left|\boldsymbol{X}^T\boldsymbol{X}\right|$

Métodos basados en la correlación entre variables explicativas. Si calculamos los coeficientes de correlación muestral entre cada par de variables, podemos decidir que existe un problema de colinealidad si algún coeficiente de correlación es mayor (en valor absoluto) que una tolerancia. Los problemas de este método son: a) sólo puede detectar correlación entre pares de variables explicativas y b) la tolerancia es arbitraria.

Métodos basados en el tamaño de $\left|\boldsymbol{X}^T\boldsymbol{X}\right|$. Como sabemos:

$$\left|\boldsymbol{X}^T\boldsymbol{X}\right| = \prod_{i=1}^{k} \lambda_i$$

siendo λ_i el *i*-ésimo autovalor de la matriz. Por tanto, podemos reducir el diagnóstico a comprobar si la matriz tiene algún autovalor próximo a cero. Para evitar el problema de unidades de medida, este análisis suele hacerse utilizando el *número de condición* de $\boldsymbol{X}^T\boldsymbol{X}$ que se puede definirse de varias maneras:

$$c_1 = \frac{\lambda_{\max}}{\lambda_{\min}};\ c_2 = \frac{1}{c_1} = \frac{\lambda_{\min}}{\lambda_{\max}};\ c_3 = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}; c_4 = \frac{1}{c_3} = \sqrt{\frac{\lambda_{\min}}{\lambda_{\max}}}$$

Soluciones

El problema de colinealidad consiste, esencialmente, en que la muestra no contiene suficiente información para estimar todos los parámetros que se desean. Por ello, resolver el problema requiere añadir nueva información (muestral o extramuestral) o cambiar la especificación. Algunas posibles soluciones en esta línea son:

Añadir nuevas observaciones. Aumentar el tamaño muestral puede reducir un problema de colinealidad de grado.

Restringir parámetros. Evidentemente, si la Teoría Económica o la experiencia empírica sugieren algunas restricciones sobre los parámetros del modelo más afectados por la colinealidad, imponerlas permitirá reducir el problema. El riesgo que se corre es, obviamente, imponer restricciones que no son ciertas.

Suprimir variables. Si se suprimen variables que están correladas con otras, la pérdida de capacidad explicativa será pequeña y la colinealidad se reducirá. Existe, sin embargo, el riesgo de eliminar variables que debieran mantenerse en el modelo ya que, como hemos visto, cuando hay colinealidad las varianzas de los parámetros están hinchadas y los parámetros pueden ser formalmente no significativos.

Transformar las variables del modelo. Si la colinealidad se debe a que se están relacionando series temporales con tendencia, puede ser conveniente transformar las variables para eliminar esta tendencia.