

ECONOMETRÍA

APLICADA UTILIZANDO R.

PAPIME PE302513 LIBRO ELECTRÓNICO Y COMPLEMENTOS DIDÁCTICOS EN MEDIOS COMPUTACIONALES, PARA EL FORTALECIMIENTO DE LA ENSEÑANZA DE LA ECONOMETRÍA

Capítulo 6

Multicolinealidad

Luis Quintana Romero



Objetivo

Identificar la multicolinealidad en un modelo econométrico, así como las pruebas de detección de la multicolinealidad y correcciones.



Introducción

La multicolinealidad debe considerarse como un problema de grado que se presenta de manera cotidiana en los modelos econométricos. Esto significa que el comportamiento de buena parte de las variables económicas guarda algún tipo de relación unas con otras y esa relación puede ser de menor o mayor grado. Solamente cuando dicha relación es de mayor grado podría ser un problema dentro de la modelación econométrica.

La relación exacta entre las variables se denomina multicolinealidad perfecta, lo cual significa que alguna o algunas de las variables que forman las columnas de la matriz de regresores sería una combinación lineal exacta del resto de columnas. Resulta más usual que se presente multicolinealidad imperfecta, lo cual intuitivamente implica que los regresores de la regresión se encuentran altamente correlacionadas, pero sin ser esos coeficientes del cien por ciento



La Multicolinealidad un Problema de Grado

- Multicolinealidad perfecta

Si suponemos que la matriz de regresores se compone de tres columnas con las variables x_1, x_2, x_3 se obtendría la siguiente relación lineal:

$$\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 = 0$$

Siendo las constantes λ_i simultáneamente diferentes de cero, esto es; $\lambda_i \neq 0 \forall i$. Lo cual permitiría expresar una variable en términos de las demás, por ejemplo al despejar x_1 :

$$x_1 = \frac{-\lambda_2 x_2 - \lambda_3 x_3}{\lambda_1}$$

Supuestos

- Violación del supuesto de rango completo de la matriz de regresores $[X]$.
- Tampoco se cumpliría para la matriz $[X'X]$, siendo singulares ambas matrices y sus determinantes iguales a cero.
- Daría lugar a la indeterminación de los estimadores de mínimos cuadrados ordinarios para los parámetros del modelo. Esta situación se explica debido a que no estaría definida la matriz inversa $[X'X]^{-1}$, que como sabemos es necesaria para obtener los estimadores de mínimos cuadrados ordinarios: $\hat{\beta} = [X'X]^{-1}[X'Y]$.

La Multicolinealidad un Problema de Grado

- Multicolinealidad Imperfecta

Si suponemos nuevamente que la matriz de regresores se compone de tres columnas con las variables x_1 , x_2 , x_3 se obtendría la siguiente relación lineal imperfecta entre ellas:

$$\lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3 + v = 0$$

Siendo las constantes λ_i simultáneamente diferentes de cero, como en el caso previo, pero ahora existe un término de error v . Debido a esto último, al despejar y expresar una variable en términos de las demás, por ejemplo al despejar x_1 , la combinación lineal que se obtiene ya no es exacta y, por ende, la multicolinealidad ya no es perfecta:

$$x_1 = \frac{-\lambda_2 x_2 - \lambda_3 x_3}{\lambda_1} + \frac{v}{\lambda_1} = \text{combinación lineal} + \text{error}$$

Supuestos

- El determinante de la matriz $[X]$ es cercano a cero, sin embargo ello no impide la obtención de los estimadores de mínimos cuadrados ordinarios, pero se mantiene el problema de identificación debido a que la variación de alguna de las X 's además de afectar a Y afectan a las demás variables impidiendo distinguir su efecto individual.

Pruebas para la Detección de Multicolinealidad

- Coeficientes t's no significativos y R^2 elevada
- Coeficientes de correlación: Elevados coeficientes de correlación entre pares de variables son un síntoma a favor de la multicolinealidad
- Regresiones auxiliares y efecto R^2 de Theil

Si el efecto de Theil fuera nulo no existiría multicolinealidad, entre mayor sea el efecto más grave es el problema

$$R^2_{\text{Theil}} = R^2 - \left[\sum_{i=1}^n (R^2 - R_i^2) \right]$$

- Regresiones auxiliares y regla de Klein

Si se tuvieran tres regresores x_1, x_2, x_3 en el modelo, las regresiones auxiliares serían las siguientes:

$$x_{1i} = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{3i} + \varepsilon_{1i}$$

$$x_{2i} = \alpha_1 + \alpha_2 x_{1i} + \alpha_3 x_{3i} + \varepsilon_{2i}$$

$$x_{3i} = \alpha_1 + \alpha_2 x_{2i} + \alpha_3 x_{1i} + \varepsilon_{3i}$$

En este caso tendremos tres coeficientes de determinación de las regresiones auxiliares R_1^2, R_2^2, R_3^2 si alguno de ellos es mayor a R^2 el problema de multicolinealidad se puede considerar grave.

Pruebas para la Detección de Multicolinealidad

- Índice de la condición de número: se deben obtener los valores característicos de la matriz $[X'X]$, a los cuales denominaremos λ_i y se divide el máximo valor característico entre el menor valor característico. una condición de número superior a 20 se considera síntoma de multicolinealidad problemática

$$\text{ICN} = \frac{\sqrt{\lambda_{\text{máximo}}}}{\sqrt{\lambda_{\text{mínimo}}}}$$

- Factor de inflación varianza: e utiliza como una medida del grado en que la varianza del estimador de mínimos cuadrados es incrementada por la colinealidad entre las variables

$$\text{VIF} = \frac{1}{1-R_i^2}$$

Por ejemplo, ante perfecta multicolinealidad $R_i^2 = 1$, lo cual hace que el VIF tienda a infinito, si la multicolinealidad es imperfecta y elevada, por ejemplo un $R_i^2 = 0.9$, el VIF será igual a 10



Soluciones al problema de Multicolinealidad

Existen métodos de corrección de multicolinealidad, si no es severo más vale no hacer nada, ya que los remediales generalmente pueden implicar problemas más fuertes que el que se buscaba corregir.

Si existe multicolinealidad los estimadores de mínimos cuadrados ordinarios siguen siendo insesgados, de modo que si el problema no es grave el modelo puede utilizarse sin que afecte en gran medida a la inferencia estadística.

Los remedios usuales son los siguientes:

- **Imponer restricciones al modelo**

si las variables x_2 y x_3 son altamente colineales es posible restringir el modelo utilizando información a priori o bien por estimaciones de corte transversal.

Suponga que nuestro modelo es:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

con $i=1,2,..n$

Soluciones al problema de Multicolinealidad

Al aplicar pruebas de detección de multicolinealidad se encontró que esta era grave y se debía a una elevada colinealidad entre x_2 y x_3

Ejemplo:

Suponga que la evidencia encontrada es que el coeficiente β_3 es un medio del coeficiente β_2 . Esto nos permite aplicar la siguiente restricción:

$$\beta_3 = 0.5\beta_2$$

Sustituyendo en el modelo original obtenemos la ecuación restringida:

$$y_i = \beta_1 + \beta_2 x_{2i} + 0.5\beta_2 x_{3i} + u_i$$

$$y_i = \beta_1 + \beta_2 (x_{2i} - 0.5x_{3i}) + u_i$$

$$y_i = \beta_1 + \beta_2 x_{2i}^* + u_i$$

Donde: $x_{2i}^* = x_{2i} - 0.5x_{3i}$

Una vez restringido el modelo la multicolinealidad se ha eliminado y al obtener el estimador de MCO $\hat{\beta}_2$ es posible obtener $\hat{\beta}_3$ si se sustituye el primero en la restricción $\beta_3 = 0.5\beta_2$

La principal limitante de este método es la carencia de antecedentes empíricos acerca de los coeficientes de interés en los modelos econométricos.

Soluciones al problema de Multicolinealidad

Otra alternativa que implica restringir el modelo original es la estimación de un modelo en corte transversal. Por ejemplo, para el caso que nos ocupa se podría estimar β_3 en un modelo de corte transversal y sustituir su valor estimado en el modelo de series de tiempo. Suponga que en la estimación de corte transversal se obtiene que:

$$\hat{\beta}_3 = 0.5$$

Se restringe el modelo sustituyendo ese valor en el modelo original:

$$y_i = \beta_1 + \beta_2 x_{2i} + 0.5 x_{3i} + u_i$$

$$y_i - 0.5 x_{3i} = \beta_1 + \beta_2 x_{2i} + u_i$$

$$y_i^* = \beta_1 + \beta_2 x_{2i} + u_i$$

Donde: $y_i^* = y_i - 0.5 x_{3i}$

La limitante de este procedimiento es que la interpretación de los parámetros de corte transversal y series de tiempo puede diferir ampliamente al calcularse sobre conjuntos de datos diferentes.

Soluciones al problema de Multicolinealidad

- **Componentes principales**

El método de componentes principales busca eliminar el problema de multicolinealidad a través de la obtención de un conjunto de variables a partir de las originales y sin implicar grandes pérdidas de información (Everitt y Hothorn, 2006). Las nuevas variables o componentes cumplen con la condición de ser ortogonales entre sí.

El método parte de una forma cuadrática $x'Ax$ que se minimiza sujeta a la condición de normalidad $x'x=1$:

$$x'Ax - \lambda(x'x - 1)$$

Donde A es una matriz simétrica.

Al derivar con respecto a x:

$$2Ax - 2\lambda x = 0$$

Al factorizar encontramos la ecuación característica:

$$(A - \lambda I)x = 0$$

Encontrar sus raíces nos permite obtener los valores característicos λ_i .

La principal limitante de este método es que las nuevas variables **Z** pueden no tener interpretación económica alguna

Soluciones al problema de Multicolinealidad

Eliminar variable

La eliminación de variables sospechosas de colinealidad puede ser otra opción para evitar el problema de multicolinealidad, sin embargo puede llevarnos a un problema más grave como el de variable relevante omitida. En nuestro ejemplo la eliminación de la variable x_{3i} deja el modelo como:

$$y_i = \beta_1 + \beta_2 x_{2i} + u_i$$

Sin embargo, si la variable omitida fuera relevante se genera un problema de sesgo en los estimadores de MCO.

Transformar variables

La transformación de variables con primeras diferencias o calculando porcentajes es otro remedio que busca diferenciar más las variables entre sí. Sin embargo, su principal limitante es que, por una lado la teoría relevante pudiera estar interesada únicamente en las variables de nivel y no en sus diferencias ni en sus porcentajes, por otro lado la variable dependiente pudiera estar relacionada con las demás en niveles pero no en porcentajes ni en diferencias.



Referencias

- L. R. Klein, An Introduction to Econometrics , Prentice-Hall, 1962;
- Theil, H, Principles of Econometrics, Wiley, 1971.
- Everitt,S. Brian y Torsten Hothorn, A handbook of statistical analysis using R, Chapman / Hall/CRC, 2006.
- Quintana Romero, Luis y Miguel Ángel Mendoza, Econometría básica, Plaza y Valdés, 2008.

