Tratamiento de la Multicolinealidad Aproximada Esencial usando Información a Priori

R. Salmerón (romansg@ugr.es) **C.B. García** (cbgarcia@ugr.es)
Dpto. de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

A. Rodríguez (arsanchez@ugr.es) **C. García** (garciaclaudia@ugr.es)
Programa de Doctorado en Ciencias Económicas y Empresariales
Universidad de Granada

XXXVIII Congreso Nacional de Estadística e Investigación Operativa
XII Jornadas de Estadística Pública

SEIO, Alcoi 3-6 de Septiembre de 2019

Índice

Introducción

Econometría

Econometría es una rama de la Economía que proporciona una base para refinar o refutar conocimiento teórico y conseguir signos y magnitudes de las relaciones de variables que se desean analizar.

Con tal objetivo se han de estimar los coeficientes del modelo lineal general:

$$\mathbf{Y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \cdots + \beta_i \cdot \mathbf{X}_i + \cdots + \beta_p \cdot \mathbf{X}_p + \mathbf{u} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \qquad (1)$$

Aplicando Mínimos Cuadrados Ordinarios (MCO) se llega al sistema de ecuaciones normales:

$$\left(\mathbf{X}'\mathbf{X}\right)\beta = \mathbf{X}'\mathbf{Y}.$$

Para que tenga solución única debe existir $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$:

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y}.$$

Si las variables que forman $\mathbf{X}$ son linealmente dependientes, entonces no existe $\left(\mathbf{X}'\mathbf{X}\right)^{-1}$ (multicolinealidad perfecta).

Introducción

¿Qué ocurre si son casi linealmente independientes? (multicolinealidad aproximada): resultados inestables y contradictorios.

- Pequeños cambios en los datos pueden suponer cambios sustanciales en las estimaciones de los coeficientes de los regresores.
- Tendencia a no rechazar que los coeficientes de los regresores son cero debido a desviaciones típicas estimadas de los coeficientes “infladas”.
- Coeficiente de determinación alto y, en consecuencia, tendencia a rechazar que todos los coeficientes son cero de forma simultánea.

Posibles soluciones:

- Mejora del diseño muestral, aumento del tamaño de la muestra o usar información a priori.
- Eliminar variables que se consideran problemáticas.
- Centrar variables.
- Usar métodos de estimación alternativos a MCO (como es la regresión cresta, alzada, LASSO, etc).
- Mínimos Cuadrados Restringidos (MCR).

Introducción

Ejemplo

Estimación por MCO del modelo sobre el crédito bancario en los Estados Unidos (variable dependiente: deuda pendiente de hipoteca):

Variable	MCO
Constante	5.469 (13.016)
Consumo personal	-4.252 (5.135)
Ingresos personales	3.1203 (2.035)
Crédito pendiente	0.0028 (0.0057)
σ^2	0.9325^2
R^2	0.9235
$F_{3,13}$ conjunta	52.3

Se observa que las estimaciones de los coeficientes del consumo, ingresos y crédito pendiente no son significativamente distintos de cero al mismo tiempo que se tiene que el modelo es conjuntamente válido (al 5 % de significación).

Esta contradicción apunta a que el grado de multicolinealidad aproximada existente en el modelo afecta al análisis estadístico del mismo.

Introducción

Propuesta

En el presente trabajo se propone mitigar este problema mediante el uso de información a priori incorporada a la estimación del modelo mediante la estimación del mismo por el método de Mínimos Cuadrados Restringidos (MCR). Puesto que este método proporciona estimaciones de las varianzas de los coeficientes estimados menores que el método de MCO, se estaría mitigando uno de los problemas producidos por una multicolinealidad aproximada grave.

Idea

La Idea del trabajo consiste en cómo obtener la restricción a considerar como hipótesis nula en los MCR dado un modelo de regresión lineal concreto de forma que no se rechace tal hipótesis y se puedan aplicar los MCR de forma correcta.

Mínimos Cuadrados Restringidos

Hay situaciones donde la Teoría Económica (o la teoría existente detrás del fenómeno en estudio) sugiere que los coeficientes de las variables independientes del modelo especificado cumplen ciertas condiciones. Este conocimiento a priori puede ser incorporado al modelo considerado con la idea de ganar eficiencia en el sentido de que se estaría usando toda la información disponible

Mínimos Cuadrados Restringidos: estimador

Si las restricciones comentadas se expresan con la hipótesis nula $H_0 : \mathbf{R}_{q\times k} \cdot \boldsymbol{\beta}_{k\times 1} = \mathbf{r}_{q\times 1}$, donde q es el número de restricciones especificadas, el estimador por Mínimos Cuadrados Restringidos (MCR) responde a la siguiente expresión:

$$\widehat{\boldsymbol{\beta}}_R = \widehat{\boldsymbol{\beta}} + \left(\mathbf{X}^t\mathbf{X}\right)^{-1}\mathbf{R}^t\left(\mathbf{R}\left(\mathbf{X}^t\mathbf{X}\right)\mathbf{R}^t\right)^{-1}\left(\mathbf{r} - \mathbf{R}\widehat{\boldsymbol{\beta}}\right). \tag{2}$$

A partir de esta expresión es inmediato comprobar que este estimador verifica las restricciones de H_0, es decir, que $\mathbf{R} \cdot \widehat{\boldsymbol{\beta}}_R = \mathbf{r}$.
Además, si el estimador por Mínimos Cuadrados Ordinarios (MCO), $\widehat{\boldsymbol{\beta}}$, verifica las restricciones dadas en H_0, esto es, $\mathbf{R} \cdot \widehat{\boldsymbol{\beta}} = \mathbf{r}$, entonces ambos estimadores coinciden. En tal caso, se tiene que el estimador por MCR es un estimador insesgado de $\boldsymbol{\beta}$.

Mínimos Cuadrados Restringidos

Mínimos Cuadrados Restringidos: matriz de varianzas-covarianzas

Por otro lado, la matriz de varianzas-covarianzas de los coeficientes estimados del modelo (1) por el método de los MCR viene dada por:

$$var\left(\widehat{\beta}_R\right) = var\left(\widehat{\beta}\right) - \sigma^2 \cdot \left(\mathbf{X}^t\mathbf{X}\right)^{-1} \mathbf{R}^t \left(\mathbf{R}\left(\mathbf{X}^t\mathbf{X}\right)\mathbf{R}^t\right)^{-1} \mathbf{R}\left(\mathbf{X}^t\mathbf{X}\right)^{-1}. \quad (3)$$

Puesto que el segundo sumando de la parte derecha de la expresión (3) es una matriz semidefinida positiva, se tiene que los elementos de la diagonal principal de la matriz $var\left(\widehat{\beta}_R\right)$ son más pequeños que los de $var\left(\widehat{\beta}\right)$.

Por tanto, si el investigador fuese capaz de especificar una hipótesis nula con restricciones de forma que no se rechacen, los MCR proporcionan un estimador del verdadero valor del parámetro insesgado y con menor varianza estimada que el dado por los MCO.

De esta forma, se podrían mitigar los posibles efectos negativos que la multicolinealdiad aproximada existente en el modelo puede tener sobre el análisis estadístico del modelo.

Tipos de Multicolinealidad

Las causas que producen multicolinealidad en un modelo son diversas.

Según Spanos y McGuirk (2002)

Multicolinealidad sistemática: debida a un problema estructural, es decir, a la alta correlación lineal de las variables exógenas consideradas.

Multicolinelidad errática: debido a un problema puramente numérico, es decir, a un mal condicionamiento de los datos considerados.

Mientras que Marquandt y Snee (1975)

Multicolinealidad no esencial: relación lineal de las variables exógenas con la constante (es sabido que se solventa centrando las variables).

Multicolinelidad esencial: relación lineal entre las variables exógenas (excluida la constante).

En este caso nos centraremos en la multicolinealidad aproximada del tipo esencial. Para la no esencial...

Detección de la Multicolinealidad: regresiones auxiliares

Dado un modelo como el dado en (1), para detectar el grado de multicolinealidad existente es habitual usar la regresión auxiliar:

$$\mathbf{X}_i = \delta_1 + \delta_2 \cdot \mathbf{X}_2 + \cdots + \delta_{i-1} \cdot \mathbf{X}_{i-1} + \delta_{i+1} \cdot \mathbf{X}_{i+1} + \cdots + \delta_p \cdot \mathbf{X}_p + \mathbf{v} = \mathbf{X}_{-i}\boldsymbol{\delta} + \mathbf{v}, \quad (4)$$

donde $i = 2, \dots, p$.
Si el coeficiente de determinación de esta regresión, R_i^2, es alto se tiene que $\mathbf{X}_i$ está altamente relacionada linealmente con el resto de variables independientes del modelo. Es decir, el grado de multicolinealidad aproximada existente es alto. Tradicionalmente, valores de R_i^2 superiores a 0.9 implican que la relación lineal de $\mathbf{X}_i$ con el resto de variables independientes del modelo es preocupante.

Ejemplo

$$\begin{aligned}
\widehat{\mathbf{C}} &= 2.478 + 0.331 \cdot \mathbf{I} + 0.000835 \cdot \mathbf{CP}, \quad R_1^2 = 0.998304, \\
\widehat{\mathbf{I}} &= -4.521 + 2.111 \cdot \mathbf{C} - 0.000735 \cdot \mathbf{CP}, \quad R_2^2 = 0.996452, \quad (5) \\
\widehat{\mathbf{CP}} &= -1940.43 + 663.216 \cdot \mathbf{C} - 91.751 \cdot \mathbf{I}, \quad R_3^2 = 0.994723. \quad (6)
\end{aligned}$$

Restricciones lineales sobre los coeficientes del modelo

¿Es posible usar las regresiones auxiliares para establcer restricciones lineales sobre los coeficientes del modelo de forma que al usarlas en los MCR no se rechace la hipótesis nula?

Ejemplo

$$\mathbf{Y} = \beta_1 + \beta_2 \cdot \mathbf{X}_2 + \beta_3 \mathbf{X}_3 + \mathbf{u}, \quad \mathbf{X}_2 = 1 + \mathbf{X}_3$$

$$\begin{aligned} \mathbf{Y} &= \beta_1 + \beta_2 \cdot (1 + \mathbf{X}_3) + \beta_3 \mathbf{X}_3 + \mathbf{u} \\ &= (\beta_1 + \beta_2) + (\beta_2 + \beta_3) \cdot \mathbf{X}_3 + \mathbf{u} = \gamma_1 + \gamma_2 \cdot \mathbf{X}_3 + \mathbf{u} \end{aligned}$$

$$\beta_1 + \beta_2 = \widehat{\gamma}_1, \quad \beta_2 + \beta_3 = \widehat{\gamma}_2$$

Restricciones lineales sobre los coeficientes del modelo

Ejemplo

Puesto que el mayor R_i^2 es el asociado a la regresión auxiliar donde la variable dependiente es el consumo personal, se considerará la estimación de esta regresión:

$$\widehat{\mathbf{C}} = 2.4786238142 + 0.3317662443 \cdot \mathbf{I} + 0.0008357644 \cdot \mathbf{CP}.$$

Sustituyendo esta relación en el modelo original, $\mathbf{D} = \beta_1 + \beta_2 \cdot \mathbf{C} + \beta_3 \cdot \mathbf{I} + \beta_4 \cdot \mathbf{CP} + \mathbf{u}$, se tiene que:

$$\begin{aligned} \mathbf{D} &= \beta_1 + \beta_2 \cdot (2.4786238142 + 0.3317662443 \cdot \mathbf{I} + 0.0008357644 \cdot \mathbf{CP}) \\ &\quad + \beta_3 \cdot \mathbf{I} + \beta_4 \cdot \mathbf{CP} + \mathbf{u} \\ &= (\beta_1 + 2.4786238142 \cdot \beta_2) + (\beta_3 + 0.3317662443 \cdot \beta_2) \cdot \mathbf{I} \\ &\quad + (\beta_4 + 0.0008357644 \cdot \beta_2) \cdot \mathbf{CP} + \mathbf{u}. \end{aligned}$$

Restricciones lineales sobre los coeficientes del modelo

Ejemplo

Si se estima el modelo $\mathbf{D} = \gamma_1 \cdot \mathbf{1} + \gamma_2 \cdot \mathbf{I} + \gamma_3 \cdot \mathbf{CP} + \boldsymbol{\eta}$, se tiene que:

$$\widehat{\gamma}_1 = -5.070908, \quad \widehat{\gamma}_2 = 1.709583, \quad \widehat{\gamma}_3 = -0.0006749105,$$

y entonces es factible establecer las relaciones lineales:

$$\beta_1 + 2.4786238142 \cdot \beta_2 = -5.070908, \quad \beta_3 + 0.3317662443 \cdot \beta_2 = 1.709583,$$

$$\beta_4 + 0.0008357644 \cdot \beta_2 = -0.0006749105.$$

Usando la regresión auxiliar (5) se llega a:

$$\beta_1 - 4.5215086279 \cdot \beta_3 = -8.6396297779, \quad \beta_2 + 2.110939188 \cdot \beta_3 = 2.3350180848,$$

$$\beta_4 - 0.0007357232 \cdot \beta_3 = 0.0005833713,$$

mientras que (6) conduce a las restricciones:

$$\beta_1 - 1940.42814 \cdot \beta_4 = -0.117458079, \quad \beta_2 + 663.21572 \cdot \beta_4 = -2.342952766,$$

$$\beta_3 - 91.75084 \cdot \beta_4 = 2.856233727.$$

Restricciones lineales sobre los coeficientes del modelo

Ejemplo

Otra opción, es la de combinar algunas de las anteriores, como por ejemplo:

$$\beta_3 + 0.3317662443 \cdot \beta_2 = 1.70958, \quad \beta_4 - 0.0007357232 \cdot \beta_3 = 0.000583371.$$

Variable	MCO	MCR 1	MCR 2	MCR 3	MCR 4
Constante	5.469 (13.016)	5.46926 (11.4728)	5.46817 (8.2963)	5.46826 (10.0825)	**5.46926** (2.20601)
Consumo personal	-4.252 (5.135)	-4.25243 (4.62868)	-4.25144 (3.8733)	-4.25243 (3.44609)	**-4.25243** (0.652718)
Ingresos personales	3.1203 (2.035)	3.1204 (1.53564)	3.12015 (1.83487)	**3.1204** (0.47674)	**3.1204** (0.21655)
Crédito pendiente	0.0028 (0.0057)	0.0028 (0.00386)	**0.0028** (0.00135)	0.0028 (0.00519)	**0.0028** (0.00016)
σ^2	0.9325^2	0.840519^2	0.840525^2	0.840519^2	0.868084^2
R^2	0.9235				
$F_{3,13}$ conjunta	52.3				
$F_{3,13}$ de H_0		$4.6439 \cdot 10^{-13}$	$6.0556 \cdot 10^{-5}$	$6.4095 \cdot 10^{-14}$	$3.1974 \cdot 10^{-16}$

Conclusiones y futuras líneas

Conclusiones y futuro

- Uno de los efectos que puede tener un alto grado de multicolinealidad aproximada en la estimación por MCO de un modelo de regresión lineal múltiple es la distorsión del análisis estadístico del mismo al provocar varianzas estimadas de los coeficientes elevadas (contrastes de significación individual).
- Puesto que los estimadores por MCR tienen menor varianza estimada de los coeficientes que los estimadores por MCO, estos estimadores pueden ser útiles para mitigar la distorsión sobre el análisis estadístico del modelo.
- Problema: para usar correctamente los MCR es imprescindible establecer restricciones lineales sobre los coeficientes del modelo. Si no hay una teoría que establezca dichas relaciones, ¿cómo plantearlas?
- Solución: usar las regresiones auxiliares para establecer restricciones lineales sobre los coeficientes con un alto grado de veracidad.
- ¿Cómo elegir el *mejor* conjunto de restricciones a considerar?
- ¿Cómo afecta la multicolinealidad a la estimación de γ y, por tanto, a las restricciones?

Bibliografía

Marquandt, D. W. and R. Snee (1975). "Ridge regression in practice". The American Statistician 29 (1), pp. 3–20.

Novales, A. (1993). *Econometría*. McGraw-Hill.

Spanos, A. y McGuirk, A. (2002). "The problem of near-multicollinearity revisited: erratic vs systematic volatily". Journal of Econometrics, 108 (2), pp. 365–393.

Tratamiento de la Multicolinealidad Aproximada Esencial usando Información a Priori

R. Salmerón (romansg@ugr.es) **C.B. García** (cbgarcia@ugr.es)
Dpto. de Métodos Cuantitativos para la Economía y la Empresa
Universidad de Granada

A. Rodríguez (arsanchez@ugr.es) **C. García** (garciaclaudia@ugr.es)
Programa de Doctorado en Ciencias Económicas y Empresariales
Universidad de Granada

XXXVIII Congreso Nacional de Estadística e Investigación Operativa
XII Jornadas de Estadística Pública

SEIO, Alcoi 3-6 de Septiembre de 2019