

Multicolinealidad

Econometría 2021-2022

GRADO INGENIERÍAS & ADE

Contenidos

Introducción

Síntomas de multicolinealidad aproximada

Tipos de multicolinealidad aproximada

Detección de multicolinealidad aproximada

Mitigación de multicolinealidad aproximada

Introducción

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

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i + \cdots + \beta_p X_p + u = X\beta + u. \quad (1)$$

Aplicando MCO se llega al sistema de ecuaciones normales:

$$(X'X) \beta = X'y.$$

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

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Si las variables que forman X son linealmente dependientes, entonces no existe $(X'X)^{-1}$ (multicolinealidad perfecta). En este caso no se puede obtener una estimación única de β .

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.
- ✖ 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, residualización, PCR, PLSR, LASSO, MCR, etc).
-

Efectos nocivos sobre el análisis estadístico del modelo



Wissel, J. (2009). *A new biased estimator for multivariate regression models with highly collinear variables*. Ph.D. thesis.

Variable	Estimación	Desviación típica
Constante	5.469264	13.016791
Consumo	-4.252429	5.135058
Ingresos	3.120395	2.035671
Crédito Pendiente	0.002879	0.005764
R^2		0.9235
$\hat{\sigma}^2$		0.8695563
$F_{3,13}$		52.3

Estimación por MCO del crédito en Estados Unidos

No se rechaza que $\beta_i = 0$, $i = 1, 2, 3$, y se rechaza que $\beta_2 = \beta_3 = \beta_4 = 0$.

Efectos nocivos sobre el análisis estadístico del modelo



Wooldridge, J.M. (2000). *Introductory Econometrics: A modern approach*. South-Western, Cengage Learning.

Variable	Estimación	Desviación típica
Constante	0.4404	0.09556
Tipos de interés a 3 meses	1.00569	0.01343
$R^2 = 0,9787$, $\hat{\sigma}^2 = 0,2025$, $F_{2,122} = 5611$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 meses

Variable	Estimación	Desviación típica
Constante	0.22471	0.0397
Tipos de interés a 3 meses	-0.62891	0.06582
Tipos de interés a 6 meses	1.59334	0.06394
$R^2 = 0,9965$, $\hat{\sigma}^2 = 0,03330625$, $F_{2,122} = 17371,66$		

Estimación por MCO de los tipos de interés a 12 meses en función de los tipos de interés a 3 y 6 meses

Correlación entre los tipos de interés a 3 y 12 meses: 0.9893021.

Efectos nocivos sobre el análisis numérico del modelo

Wissel

C	C.p	I	I.p	CP	CP.p	β	β_p
4.7703	4.838464	4.8786	4.992323	808.23	839.7796	5.46926428	-12.844487481
4.7784	4.910650	5.0510	5.136480	798.03	789.1984	-4.252429358	3.324325322
4.9348	5.038435	5.3620	5.385545	806.12	836.3163	3.120395253	0.316260872
5.0998	5.044642	5.5585	5.635094	865.65	871.7818	0.002879118	-0.002397538
5.2907	5.350734	5.8425	5.838306	997.30	1006.9633		
5.4335	5.501856	6.1523	6.256959	1140.70	1148.8869		
⋮	⋮	⋮	⋮	⋮	⋮		

Wooldridge

r3	r3.p	r6	r6.p	β	β_p
2.77	2.784788	3.02	3.027994	0.2247061	0.1845118
2.97	3.000442	3.43	3.463405	-0.6289143	-0.5209103
4.00	4.168316	4.32	4.423468	1.5933374	1.4875326
4.60	4.628050	4.68	4.721110		
4.16	4.148983	4.33	4.338835		
3.07	3.166463	3.50	3.588747		
⋮	⋮	⋮	⋮		

Tipos de multicolinealidad aproximada

Las causas que producen multicolinealidad en un modelo son diversas.
Según Spanos y McGuirk:

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

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

Mientras que Marquandt y Snee:

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

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

Tipos de multicolinealidad aproximada

Luego, se podrían distinguir los siguientes cuatro casos:

Multicolinealidad	Sistemática	Errática
No esencial	1	2
Esencial	3	4

¿Puede existir multicolinealidad aproximada preocupante en el modelo de regresión lineal simple $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + u$?

¿Y en el modelo $\text{Salario} = \beta_1 + \beta_2 \cdot \text{Experiencia} + \beta_3 \cdot \text{Edad} + u$?

- ✦ ¿Qué falla? ¿La definición de multicolinealidad, de variable independiente o de ambas?
 - ✦ A partir del tipo de modelo: multicolinealidad aproximada esencial y no esencial.
 - ✦ A partir del tipo de variables: multicolinealidad aproximada errática y sistemática.
 - ✦ ¿Cómo definir la variable Experiencia?
-

Factor de Inflación de la Varianza

Una de las medidas más usadas para detectar el grado de multicolinealidad existente en el MLG es el Factor de Inflación de la Varianza (FIV) dado por:

$$FIV(i) = \frac{1}{1 - R_i^2}, \quad i = 2, \dots, p, \quad (2)$$

donde R_i^2 es el coeficiente de determinación de la regresión de X_i sobre el resto de variables independientes, X_{-i} .

Si esta medida es superior a 10 se supone que el grado de multicolinealidad presente en el modelo es preocupante.

NOTA: El VIF no tiene en cuenta la relación de las variables exógenas del modelo, $X_2 \dots X_p$, con la constante, 1. Por tanto, no detecta la multicolinealidad no esencial.

Número de Condición

Otra medida muy extendida es el Número de Condición (NC), el cual viene dado por:

$$NC = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}, \quad (3)$$

donde λ_{max} y λ_{min} son, respectivamente, los autovalores máximo y mínimo de $\tilde{X}^t \tilde{X}$ donde:

$$\tilde{X} = [\tilde{1} \ \tilde{X}_2 \ \dots \ \tilde{X}_p],$$
$$\tilde{1} = \frac{1}{\sqrt{n}}, \quad \tilde{X}_i = \frac{X_i}{\sqrt{\sum_{j=1}^n X_{ji}^2}}, \quad i = 2, \dots, p.$$

Si esta medida es superior a 20 se supone que el grado de multicolinealidad presente en el modelo es moderado y si es superior a 30 preocupante.

NOTA: El NC tiene en cuenta la relación de las variables exógenas del modelo, $X_2 \dots X_p$, con la constante, 1.

Otras medidas

Otras medidas: índice de Stewart, CFIV, red indicator o matriz de correlaciones lineales de las variables independientes y su determinante.

Dada la matriz de correlaciones simples, R , de las variables independientes del modelo, X , ignora por completo a la constante y proporciona información de las relaciones dos a dos, por lo tanto, sólo es capaz de detectar multicolinealidad aproxima del tipo esencial dos a dos.

El determinante de la matriz de correlaciones, $\det(R)$, recoge estructuras más complejas aunque sigue ignorando la relación con la constante, por lo que detecta multicolinealidad aproxima del tipo esencial. En este caso, valores próximos a cero indica que este problema es grave.

García y otros muestran que valores del $\det(R)$ inferiores a $0,1013 + 0,00008626 \cdot n - 0,01384 \cdot k$ indican presencia de multicolinealidad aproximada esencial preocupante.

Mitigación de multicolinealidad aproximada

No todas las soluciones propuestas para mitigar la multicolinealidad aproximada tienen éxito ante todos los tipos de multicolinealidad aproximada.

$$\widehat{var}(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{n \cdot var(X_j) \cdot (1 - R_j^2)} \quad (4)$$

- ✦ Aumento de la muestra: quizás si la multicolinealidad aproximada es errática tenga éxito, pero ¿no con la sistemática? (la clave reside en si la nueva información es más de lo mismo o no).
 - ✦ Aumento de la muestra: puede mitigar los efectos sobre el análisis estadístico del modelo (ver expresión (4)), pero ¿qué pasa con los efectos sobre el análisis numérico del mismo?
 - ✦ Una buena especificación del modelo econométrico implica una SCR *pequeña* y, por tanto, un valor de $\hat{\sigma}^2$ también *pequeño*. Como consecuencia se reduciría el valor de $\widehat{var}(\hat{\beta}_j)$ (ver expresión (4)), mitigando los efectos sobre el análisis estadístico del modelo.
 - ✦ Centrar las variables solventa la multicolinealidad aproximada no esencial, pero no tiene efecto alguno sobre la esencial.
 - ✦ Uso de información a priori: MCR.
-