ECTRÍA – T7 | Multicolinealidad | Mod. No lineales

Recordar: $(X'X)\beta = X'Y$ ↔ ec's normales MCO

1_MULTICOLINEALIDAD: Concepto y consecuencias

La multicolinealidad aparece cuando las var. explicativas de un modelo econométrico están correlacionadas entre sí, y tiene implicaciones negativas en la estimación MCO.

Multicolinealidad exacta (perfecta): Una de las var. ~~exo~~ deterministas exógenas es c.l. de las demás.

La matriz $(X'X)$ es singular $\Rightarrow \nexists (X'X)^{-1} \Rightarrow$ no se puede estimar $\hat{\beta} = (X'X)^{-1}X'Y$.
$Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$

Multicolinealidad aproximada: Una de las var. explicativas es aprox c.l. de las demás.

La matriz $(X'X)$ no es singular, sólo aprox y $\hat{\beta}$ muy sensible.

↳ La soluc. está mal definida, existen β_1, β_2 ... soluc. aproximada ($\hat{\beta}$ indet)

↳ Pequeñas variaciones muestrales $\Rightarrow \hat{\beta}$ muy distinto

↳ $(X'X)^{-1}$ casi singular $\Rightarrow$ varianza muy grande. $\hat{\beta}$ poco preciso

Es prácticamente imposible encontrar un modelo sin multicolinealidad. La cuestión relevante en el trabajo empírico es discutir en qué medida existe multicolinealidad $\rightarrow$ grado.

Aumento de la correlación entre las var:

↳ $|X'X|$ ~~muy bajo~~ ↘
↳ $\hat{\beta}$ $Var(\hat{\beta}_{MCO})$ ↗
↳ Cov parcial ↗

| Mínimo cambio en $X'Y$ $\Rightarrow$ Cambio drástico en $\hat{\beta}$

Multic. Exacta $\Rightarrow$ Fácil detección (ordenador) $\Rightarrow$ Eliminar la causa

Multic. Aproximada $\Rightarrow$ el ordenador no lo detecta y ofrece soluc.

$Var(\hat{\beta})$ excesiva $\Rightarrow$ Intervalos de confianza demasiado grandes

Contrastes sesgados (aceptan H_0 + fácil)

Tendemos a creer que los coef. no son signif.

¡R^2 alto y muchos coef. no signif!

2_ DETECCIÓN de la multicolinealidad

- La ~~de~~ multicolinealidad es un problema de grado, no de clase.
- La multicolinealidad se refiere a las var. explicativas, que son deterministas, por lo que es una característica de la muestra y no de la población.

(1)

Detección

a) $\hat{\beta}$ sensible a pequeños cambios en los datos.
b) $\hat{\beta}_i$ con desv. típicas muy altas, individualmente poco significativos, paradójicamente R^2 global alto.

1_ Mt. basados en la correl. entre variables:

- Observar las correlaciones entre dos regresores → ¡cuidado! Puede haber colinealidad sin que sea alta. Aunque si es alta, hay colinealidad.
- Regresiones auxiliares de cada X_i sobre las restantes X. Calcular los coef. de determinación auxiliares de cada regresión, R_i^2. Regla práctica: $R_i^2 > R^2$ global ⇒ problema multicolinealidad

2_ Mt. basados en el tamaño de $(X'X)$.

$|X'X|$ bajo ⇒ multicolinealidad ¡Cuidado! Unid. de medida.
solución → Considerar valores propios de la matriz con los valores de x_i normalizados

λ_i pequeños ⇒ det $(X'X)$ pequeño

Nº condición de la matriz X $\xrightarrow{\text{Índice}}$ IC $= \sqrt{\dfrac{\lambda max}{\lambda min}}$; IC > 25 ⇒ Multic

3_ REMEDIOS contra la multicolinealidad

Antes de buscar remedios, unas observaciones:

Aún cuando la multicolinealidad es muy alta, los estimadores MCO siguen siendo lineales, insesgados y mínima varianza, por lo que no viola los supuestos básicos de la regresión.

El único problema está en que las desviaciones de los coef. de regresión estimados son muy grandes, igual que si tenemos pocas observaciones.

La solución que se tome debe depender de la finalidad del modelo.

Fines predictivos → se puede trabajar con multicolinealidad. El ajuste global es bueno ⇒ predic. buenas.

Fines estructurales → multicolinealidad problema grave. Interpretación difícil.

Remedios

a) Mediante estimación del modelo propuesto.

1. Preestimación. Excluir una var. y asociar al coef. una estimac. por otros proced. a ...
2. Regresión cresta. Al incrementar los elementos de la diagonal de $X'X$ en una cte se evita la singularidad. El estimador es sesgado, pero puede tener un ECM menor que MC. Problema: ∞'s soluciones, criterio de elección arbitrario.
3. Otras soluciones
   - ~~Tom~~ Añadir ec's con la relac. entre variables. → + complejo
   - Especificar la relac. entre retardos.
   - Utilizar componentes principales.

b) Mediante exclusión de variables.

Detectar la var. que depende de las demás y eliminarla. Problema de error de especificación y problema de significado estructural del modelo. → se omiten var's relevantes

c) Estimadores restringido y no restringido

Imponer las condiciones de exclusión de las var. que generan el problema de colinealidad.

Comparar $\hat{\beta}_{MCO}$ y $\hat{\beta}_{Restr.}$ y elegir el mejor de acuerdo a diferentes criterios.

4_ OBSERVACIONES INFLUYENTES.

En ocasiones, existen observaciones muestrales que pueden tener una importancia apreciable en la estimación MC.

En datos tipo sección cruzada, individuos muy influyentes que obviamente condicionen el resultado de la estimación.

En datos series temporales, instantes de tiempo relevantes.

Resulta interesante disponer de las estimaciones $\left\{ \begin{array}{c} \text{CON} \\ \text{Y} \\ \text{SIN} \end{array} \right\}$ estas observaciones influyentes y compararlas.

Idea: Examinar los residuos de una estimación inicial.
Si una observación es influyente, su residuo será muy alto en valor absoluto, atípico.

Antes de comparar los residuos hay que normalizarlos.

Habitualmente se utiliza un procedimiento más completo.

$\hat{\beta}(i/t) \equiv \hat{\beta}_{MCO}$ excluyendo la observación i-ésima (instante t)

$\downarrow$

$\hat{u}_i(i/t)$ y $Var(\hat{u}_{i/t}(i/t))$ $\longrightarrow$ Normalizar residuo MCO.

Si Residuo normalizado $> 2 \Rightarrow$ influyente.

5_ESPECIFICACIONES NO LINEALES

En el MLG suponíamos modelo lineal para simplificar los resultados, pero la realidad económica no tiene por qué ser lineal.

Podemos distinguir varios tipos de no linealidad, según esté en la var. endógena, las var. explicativas o los parámetros.

a) No linealidad en las var. exógenas (no en los coef.)

Por ejemplo, $Y_t = \beta_1 + \beta_2 e^{X_{2t}} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} X_{5t} + u_t$

Basta definir nuevas variables transformando las originales para conseguir la linealidad, y actuar como siempre.

$Z_{2t} = e^{X_{2t}}$; $Z_{3t} = \ln X_{3t}$; $Z_{4t} = X_{4t} X_{5t}$

$$Y_t = \beta_1 + \beta_2 Z_{2t} + \beta_3 Z_{3t} + \beta_4 Z_{4t} + u_t$$

b) No linealidad en la var. endógena.

Se hace imposible expresar Y_t de modo explícito en función de X_t y de β. La forma funcional es implícita.

c) No linealidad en los coeficientes, (no a las var.)

$Y_t = \beta_1 + (\ln \beta_2) X_t + u_t$ → Se estima $\ln \beta_2$ por los mét. tradic.

$\hat{\beta}_2 = \exp(\ln \beta_2)$

↳ $\hat{\beta}_2$ no hereda propiedades

$Y_t = \beta_1 + \beta_2 e^{\beta_3} X_t + u_t$

Se puede estimar por mt. tradic. el coef. de X_t, pero no se pueden recuperar estim. de β_2 y β_3, a no ser que se tenga información adicional.

En los modelos lineales, nº parám = nº var. explicativas.

En modelos NO lineales, nº parám ≠ nº var. explic.

6_ APROXIMACIÓN LINEAL al modelo no lineal

$$Y_t = f(X_t, \beta) + u_t \quad , t = 1 \dots T \quad \text{donde } f \text{ NO lineal}$$

Obtener la mejor aproximación lineal (mediante desarrollo en serie de Taylor) alrededor de una estimación inicial $\hat{\beta}$ y estimar el modelo lineal resultante por los mt. tradic.

Dicha aproximación es:

$$Y_t \simeq f(X_t, \hat{\beta}) + \left(\frac{\partial f(X_t, \hat{\beta})}{\partial \beta}\right)'_{\beta=\hat{\beta}} (\beta - \hat{\beta}) + u_t \quad , \quad t = 1 \dots T$$

$$\underbrace{Y_t - f(X_t, \hat{\beta}) + \left(\frac{\partial f(X_t, \hat{\beta})}{\partial \beta}\right)'_{\beta=\hat{\beta}} \hat{\beta}} = \left(\frac{\partial f(X_t, \hat{\beta})}{\partial \beta}\right)'_{\beta=\hat{\beta}} \cdot \beta + u_t \quad , t$$

$$Y^*_t = \left(\underset{K\times 1}{\frac{\partial f(X_t, \hat{\beta})}{\partial \beta}}\right)'_{\beta=\hat{\beta}} \cdot \underset{K\times 1}{\beta} + u_t \quad , \quad t = 1 \dots T$$

De este modo, dada una primera aproximación a ~~de~~ $\hat{\beta}$, se trata de construir Y^*_t, así como las k var. del gradiente de la función en el punto $\beta = \hat{\beta} \longrightarrow \frac{\partial f_t}{\partial \beta_i}$

Se estima por MC el modelo de Y^*_t a partir de $\frac{\partial f_t(\hat{\beta})}{\partial \beta}$ para obtener la nueva estimación de β.

El estimador MCO obtenido, $\tilde{\beta}$ es función de $\hat{\beta}$ y de $\hat{u}$, donde $\hat{u}$ es el residuo de la estim. inicial.

$$\hat{\sigma}^2_u = \frac{\tilde{u}'\tilde{u}}{T-K} \quad , \text{donde} \quad \tilde{u} = y - f(x, \tilde{\beta})$$ — residuo de la aprox. lineal

$\tilde{\beta}$ se distribuye como una normal

$\frac{(T-K)\hat{\sigma}^2_u}{\sigma^2_u} \longrightarrow \chi^2_{T-K}$

indep.

$\Rightarrow$ se puede utilizar mt. tradicionale

Resumen

$$Y_t = f(X_t, \beta) + u_t \text{ (no lineal)} \xrightarrow{\hat{\beta}\text{ inicial}} Y^*_t = X^*_t \cdot \beta + u_t \quad \text{(lineal)}$$

$\downarrow$ MCO

$\tilde{\beta}_{MCO}$ / $\hat{\sigma}^2_u = \frac{\tilde{u}'\tilde{u}}{T-K}$ / $\tilde{\beta}_{MCO} \rightarrow$ Normal

$\frac{(T-K)\hat{\sigma}^2_u}{\sigma^2_u} \rightarrow \chi^2_{T-K}$

ECTRÍA _ T7

7_ MINIMOS CUADRADOS NO LINEALES

Aplicar el procedimiento de minimización de la suma residual directamente al modelo no lineal.

La idea es la misma, pero la resolución analítica es + complicada.

$$\min_{\beta} SR(\hat{\beta}) = \sum \left(y_t - f(x_t, \hat{\beta}) \right)^2$$

~~Toma~~ C.N. Primera derivada $= 0$ → Sistema de k ec's normales

La solución al sistema de ec's normales es el estimador de mínimos cuadrados no lineales, MCNL.

Las ec's normales suelen ser no lineales ⇒ la obtención de $\hat{\beta}$ se complica.

Además, $\hat{\beta}$ suele ser una función no lineal de y.

Dificultades:

1_ $\hat{\beta}$ no suele ser insesgado.
las propiedades vienen de MV.

2_ las soluc. a un sistema de ec's no lineales no tienen por qué ser únicas ⇒ varios estimadores
⇒ no existe soluc.

Cuando el tamaño muestral crece,
$\hat{\beta} \longrightarrow$ Normal
lo que significa que se pueden realizar los habituales contrastes.

8_ Estimador MÁXIMA VEROSIMILITUD

Precisa de un supuesto sobre la distribución de u.

Sp. $u \rightarrow$ Normal

Max la f. verosimilitud es equivalente a max. su logaritmo.

Si σ_u^2 no depende de ningún β, entonces max. verosimilitud es equivalente a minimizar ~~maximizar~~ la $SR(\hat{\beta}) \Rightarrow \hat{\beta}_{MV} = \hat{\beta}_{MCNL}$

$$\hat{\sigma}_u^2 = \frac{SR(\hat{\beta})}{T}$$

Para muestras grandes, la matriz de cov. del $\hat{\beta}_{MV}$ se puede aprox. por la inversa de la matriz de información.

Este resultado justifica la distrib. normal asintótica de $\hat{\beta}$.

MULTICOLINEALIDAD

Var. explicativas correlacionadas entre sí

↳ exacta : 1 c.l. demás ⇒ $\nexists (X'X)^{-1}$ ⇒ Imposible $\hat{\beta}$ y $\hat{\sigma}_u^2$.

Fácil detección → eliminar la var. causante

↳ aproximada : $\hat{\beta}$ muy sensible

$(X'X)^{-1}$ casi singular ⇒ Var$(\hat{\beta})$ muy grande → IC excesivo

→ Contrastes sesgados

→ Se acepta + fácil H_0

→ R^2 alto y coef = 0

Detección
- Problema de grado, no de clase
- característica de la muestra, no de la poblac.

Cuidado si : $\hat{\beta}$ muy sensible (pequeños cambios datos ⇒ fuerte varicc. $\hat{\beta}$)

$\hat{\beta}_i$ poco signif + desv. típicas altas + R^2global alto

Mt. correlac : • Correlac. alta $\Longrightarrow$ multicolinealidad

$\not\Longleftarrow$

• Regresiones $X_i / X_{j \neq i}$ con R_i^2. $R_i^2 > R^2$global ⇒ multicol.

Mt. $(X'X)$: • $|X'X|$ bajo ⇒ multicolinealidad (¿escala?) → valores propios

λ_i bajos ⇒ $| \ |$ bajo.

Remedios

Obs : MCO sigue siendo lineal + insesgado + min. varianza

⇒ Fines predictivos : MCO ✓, predicción buena

⇒ Fines estructurales : buscar solución (interpret. muy difícil)

↳ Estimación modelo propuesto

↳ Exclusión de variables

↳ Comparar $\hat{\beta}_{MCO}$ y $\hat{\beta}_R$ (Restringido, eliminado 1 var).

MODELOS NO LINEALES

• En las var. exógenas : $\ln X_t \longrightarrow X_t^*$ y MCO.

• En la var. endógena → Problema.

• En los coef : Transformar y estimar, no se puede deshacer cambio

↳ Aproximación lineal mediante desarrollo en serie de Taylor (a partir de $\hat{\beta}$)

$$Y_t \simeq f(X_t, \hat{\beta}) + \left(\frac{\partial f(X_t, \beta)}{\partial \beta}\right)'_{\beta = \hat{\beta}} (\beta - \hat{\beta}) + u_t \quad , t = 1..T$$

$$Y_t^* = \left(\frac{\partial f}{\partial \beta}\right)_{\beta = \hat{\beta}} \cdot \beta + u_t \longrightarrow \text{estimar } \neg 2^{\circ} \ \hat{\beta} \sim \tilde{\beta}$$

$\hat{\beta}_{MCNL}$ = mínimos cuadrados no lineal (misma idea, + complejo)

no insesgado, no único o no existe

$u \to$ Normal ⇒ $\hat{\beta}_{MV} = \hat{\beta}_{MCNL}$