

Tema 7

Incumplimiento de las hipótesis básicas

(actualizadas el 31-08-2022)

Tema 7. Incumplimiento de las hipótesis básicas

7.1 Introducción

7.2 Multicolinealidad

7.3 Contraste de normalidad

7.4 Heterocedasticidad

7.5 Autocorrelación

Bibliografía

- Ezequiel Uriel (2013): Capítulo 6 (epígrafes 6.1 a 6.6)
- Wooldridge (2015): Capítulo 3 (epígrafe 3.4), Capítulo 9 (epígrafe 9.1), Capítulo 8 (epígrafes 8.1 y 8.3) y Capítulo 2 (epígrafe 12.2)
- Stock y Watson (2012): Capítulo 6 (epígrafe 6.7) y Capítulo 9 (epígrafe 9.2)

7.1 Introducción

Las h.e.b son necesarias para que los estimadores sean ELIO, también para hacer contrastes y predicciones

recordamos las propiedades probabilísticas

- En el tema 3 vimos que en el MLB; es decir, **si se cumplen las h.e.b**:
 - los estimadores MCO son **ELIO**
 - $\hat{\beta}_j \longrightarrow N(\beta_j, \sigma_{\hat{\beta}_j}^2)$. A partir de este resultado se obtiene el t -ratio y el estadístico F , con los que hemos hecho **inferencia** y estimación por intervalos
- Por contra, evidentemente, **si las h.e.b no se cumpliesen**, entonces alguno de estos resultados podría no cumplirse

En este tema revisaremos algunas de las h.e.b con el propósito de

- Analizar qué **causas** generan su no cumplimiento
- Qué **consecuencias** se derivan de su no cumplimiento
- Cómo **detectar** si se verifican o no
- Si se detecta un incumplimiento, conocer los **tratamientos** correctores

Recordando las h.e.b

1) Correcta especificación: $y_i = \beta_1 + \beta_2 x_i + u_i$

2) Las u_i son v.a. no observables.

3) $E(u_i) = 0$

4) $Var(u_i) = \sigma^2 \quad i = 1, \dots, N$ **(HOMOCEDASTICIDAD: apartado 7.4)**

5) $Cov(u_i, u_j) = 0$ para $i \neq j$

6) $u_i \longrightarrow N$ **(NORMALIDAD: apartado 7.3)**

7) Los regresores son no estocásticos, o sea, los regresores son fijos

8) La matriz de datos de los regresores debe cumplir que:

- 8.1) $N \geq k$ (Hay que tener al menos tantas observaciones como parámetros β)
- 8.2) Los k regresores deben ser **linealmente independientes**; es decir, no pueden existir relaciones lineales exactas entre los regresores. **(NO COLINEALIDAD PERFECTA: apartado 7.2)**

Resumen de los resultados que obtendremos

- Si se cumplen las h.e.b los estimadores MCO son ELIO
- Para que sean ELIO no es necesaria la hipótesis de normalidad, pero si las perturbaciones se distribuyen normal, entonces los estimadores también y podemos hacer inferencia

Sin embargo, si alguna de las h.e.b no se cumpliera estos resultados pueden dejar de ser válidos

- **[7.2 Colinealidad o multicolinealidad]**. Si hubiese una elevada colinealidad entre los regresores, los estimadores MCO seguirían siendo ELIO, pero su varianza sería elevada: es decir, **serían poco precisos**
- **[7.3 no normalidad]**. Si las perturbaciones no siguiesen una distribución Normal, entonces los estimadores continúan siendo ELIO, pero ya no se distribuirán de forma Normal y, por tanto, tendremos **problemas para hacer contrastes** con los estadísticos habituales (t, F)
- **[7.4 heterocedasticidad]**. Si las perturbaciones no son homocedásticas, los estimadores MCO continúan siendo insesgados, pero **dejan de ser óptimos**

7.2 Multicolinealidad

la colinealidad hace referencia a la existencia de relaciones lineales entre los regresores de forma que su precisión se ve afectada

¿Qué es la colinealidad o multicolinealidad?

- Hay colinealidad cuando los regresores (x) están **correlacionados**
- Uno de los objetivos del MRL es explicar el comportamiento y en función de una serie variables explicativas. Para ello se han de **separar los efectos** de cada uno de los regresores sobre y .
- Si las variables explicativas tienden a moverse conjuntamente (es decir están correlacionadas), el modelo presentará cierto grado de colinealidad y la separación de los efectos individuales de cada x sobre y se verá dificultada. (Ejemplo: la experiencia laboral y la edad como variables explicativas de los salarios)

colinealidad perfecta y colinealidad (a secas)

- Conviene **diferenciar** claramente entre colinealidad perfecta y colinealidad (a secas)

Colinealidad perfecta

- Se trata de un **caso teórico**, ya que en la práctica no suele producirse este tipo de colinealidad
- Para que se produzca tiene que haber un regresor que sea C.L. exacta de otros regresores del modelo. Esto supone un **incumplimiento de la hipótesis 8.2**
- Por ejemplo: **trampa de las ficticias** o el gasto medido en euros y dolares
- En este caso no es posible efectuar estimaciones de los parámetros, **los estimadores no están definidos** por lo que ni siquiera cabe plantearse las propiedades de los estimadores
- Recuerda que la colinealidad perfecta es una curiosidad teórica. En la práctica **no suele darse** y si se diese, no podríamos estimar, lo que nos avisaría de su existencia

Colinealidad no perfecta (o sencillamente colinealidad)

- En la práctica no suelen darse situaciones en las que se presente colinealidad perfecta; Sin embargo si es habitual que las variables explicativas presenten **cierto grado de colinealidad**
- Cuanto más alta sea la correlación entre los regresores, más difícil será **separar sus efectos**, haciendo que aumenten las varianzas de los estimadores MCO, siendo por tanto mayor el riesgo de obtener **estimaciones imprecisas**
- Si la elevada correlación entre los regresores hace que los resultados de estimación sean insatisfactorios, entonces decimos que el modelo sufre de colinealidad.
- La colinealidad es un problema de grado: toda(!!) regresión sufre de este problema, por lo que sólo se dice que existe multicolinealidad cuando se cree que está afectando seriamente a los resultados de regresión.

La colinealidad es parte de un problema más general: la precisión de los estimadores

En el tema 3 vimos que la varianza de los estimadores depende de:

$$\sigma_{\hat{\beta}_j}^2 = \frac{\sigma^2}{N \text{Var}(x_j) (1 - R_j^2)}$$

Cuanto menor sea la varianza, más precisos son los estimadores y, por tanto, más fiables son las estimaciones

- Si aumenta el **tamaño muestral** (N) disminuye la varianza del estimador
- Si aumenta la **variabilidad del regresor** ($\text{Var}(x_j)$) disminuye la varianza del estimador
- Si aumenta la **varianza de las perturbaciones** (σ^2) aumenta la varianza del estimador
- El cuarto factor ($(1 - R_j^2)$) está relacionado con la **colinealidad** (de los regresores)

¿Cómo afecta la colinealidad a la varianza de los estimadores?

- Si hay una elevada colinealidad entre los regresores, los estimadores MCO continúan siendo **ELIO**, pero su **varianza es elevada**; es decir, son **poco precisos**.

$$\sigma_{\hat{\beta}_j}^2 = \frac{\sigma^2}{N \text{Var}(x_j) (1 - R_j^2)}$$

- Si hay una elevada colinealidad entre los regresores, estamos diciendo que x_j está muy relacionado con otro(s) regresores del modelo; y si eso ocurre ... es como si los otros regresores pudiesen explicar el comportamiento de X_j
- El cuarto factor $(1 - R_j^2)$ representa la proporción de la variación total de x_j que no es explicada por el resto de regresores. Si la colinealidad es elevada, entonces R_j^2 será elevado, y por tanto $(1 - R_j^2)$ reducido, lo que llevaría a un aumento de $\sigma_{\hat{\beta}_j}^2$
- Repetimos, si la colinealidad aumenta, entonces R_j^2 **aumenta** y, por tanto, **la varianza de los estimadores aumenta**

Algunos detalles más sobre R_j^2

$$\sigma_{\hat{\beta}_j}^2 = \frac{\sigma^2}{N \text{Var}(x_j) (1 - R_j^2)}$$

- Si $R_j^2 = 1$ estaríamos en el caso de colinealidad perfecta. No podemos obtener los estimadores MCO; es como si la varianza de estos fuese infinita
- Si R_j^2 es elevado, la varianza del estimador también. Esto es un problema para la precisión y fiabilidad del estimador
- Sin embargo, no existe un valor de R_j^2 que nos informe de que la colinealidad es un problema grave
- El problema de la colinealidad ($R_j^2 \rightarrow 1$) es "similar" a tener una muestra pequeña o poca variabilidad de x_j o un fenómeno con mucho ruido (σ^2)
- Es decir, la colinealidad **es uno de los factores** que nos pueden llevar a tener estimadores con insuficiente precisión
- Una fuerte correlación entre 2 regresores, generalmente es irrelevante en la estimación de los efectos de los otros regresores

¿Cómo podemos detectar si nuestro modelo tiene problemas de colinealidad?

- Como la colinealidad es un problema esencialmente muestral, asociado a los datos que tenemos, no se cuenta con un método único para detectar cuando esta constituye un problema grave. Lo que tenemos son una reglas generales (algunas formales y otras informales).

Algunas de ellas son:

- Elevadas correlaciones entre parejas de regresores
- Pequeños cambios en la muestra generan grandes cambios en las estimaciones
- Un R^2 elevado (lo que significa que los regresores explican un elevado porcentaje de y), pero pocos regresores significativos individualmente
- **FAV** (Factor de agrandamiento de la varianza). Gretl lo llama FIV.

Si tenemos un problema de colinealidad, ¿podemos solucionarlo?

Los datos en Economía, generalmente se recogen por recopilación pasiva; así que poco podremos hacer si resulta que tenemos 2 regresores muy colineales en la muestra. Algunas posibles "soluciones"

- **Mejora del diseño muestral:** tratar de obtener una muestra con menos colinealidad. En el ejemplo de la determinación del salario, se puede seleccionar un subconjunto de individuos para los que la correlación entre la edad y la experiencia laboral no esté tan correlacionada.
- **Eliminación de variables:** Si las dos variables están muy correlacionadas, es porque las dos aportan, esencialmente, la misma información. Pero ...
- **Combinar variables:** si son similares conceptualmente, por ejemplo educación del padre y de la madre, o gastos en publicidad.
- **Utilización de información extramuestral:** Por ejemplo, imponer restricciones sobre los parámetros basándose en información extramuestral
- **Transformar las variables:** ratios, tasas de crecimiento, primeras diferencias, desviaciones respecto tendencia. La transformación ha de tener sentido

¿Realmente podemos aliviar el problema de la colinealidad?

- Ya hemos visto que solucionar o aliviar la colinealidad es complicado y no siempre es posible
- Otra forma de intentar aliviar el problema de la colinealidad es intentar que los otros 3 factores que afectan a la precisión de los estimadores contribuyan a reducir la varianza

$$\sigma_{\hat{\beta}_j}^2 = \frac{\sigma^2}{N \text{Var}(x_j) (1 - R_j^2)}$$

1. **Recoger más datos** ($\uparrow N$). Evidentemente esto no siempre es posible
2. **Aumentar la variabilidad del regresor** que queremos estimar su efecto,
 $\uparrow \text{Var}(x_j)$
3. **Reducir σ^2** . ¿Es esto posible?

Para finalizar con la colinealidad 3 reflexiones

- Quizás, a veces, la colinealidad la puede provocar el propio investigador al pedirle demasiado a los datos
- A los investigadores nos gustaría entender todos los fenómenos y estimar de forma precisa los efectos de todos los regresores, pero esto no siempre es posible
- Finalmente, cabe preguntarse si la colinealidad es realmente un problema. Si el objetivo de la investigación es la estimación de los efectos parciales de los regresores, la respuesta es sí; pero si el objetivo fuese, por ejemplo, la predicción, entonces no necesariamente.

Una pregunta de examen

Exemple(Examen 4/7/2006): Una multinacional desitja analitzar els factors que determinen els salaris dels seus treballadors i per a això es disposa d'una mostra per la qual es coneix:

SALARI: salari brut anual del treballador en milers d'euros.

EXPLAB: experiència laboral del treballador en anys.

SEXE: Variable fictícia que pren valor 1 si el treballador és home i 0 en cas contrari.

TAMSUC: grandària de la sucursal mesurat pel nombre de treballadors.

Cuadro 2

Regresores : 1, TAMSUC, EXPLAB, EDAD, SEXO Muestra : 1 - 150 Nº Observaciones : 150				
Regresores	Coeficiente	Desv. Típica	Estadís. t	Prob> t
1	11.483656	0.916949	12.52	0.0000
TAMSUC	0.801402	0.001716	467.03	0.0000
EXPLAB	-1.658855	3.813179	-0.44	0.6642
EDAD	2.937429	3.773577	0.78	0.4376
SEXO	1.364399	0.184895	7.38	0.0000
Media Var. Dependiente: 462.7805 Des. Típ. Var. Depen.: 45.9906 Error Típico Regresión: 1.1257 Suma Cuadrados Resid.: 183.7580 R Cuadrado : 0.9994 R Cuadrado Corregido : 0.9994 Logaritmo de Verosim. : -228.0646 Criterio AIC : 3.1075 Estadístico F(4, 14): 62134.389 Prob > F : 0.0000 Estadís. Durbin-Watson: 1.9727 Est. Autocorrelación : 0.0136				

Cuadro 1

Coeficientes de correlación		
EDAD	EXPLAB	0.999
	SEXO	0.071
	TAMSUC	0.040
	SALARIO	0.346
EXPLAB	SEXO	0.071
	TAMSUC	0.040
	SALARIO	0.346
SEXO	TAMSUC	-0.00
	SALARIO	0.033
TAMSUC	SALARIO	0.950

e) Detecta algun problema en l'estimació del model? Com ho solucionaria?
 Raone la seua resposta.

7.3 Normalidad

La construcción de intervalos de confianza y la realización de contrastes de hipótesis se han desarrollado partiendo de que los estimadores siguen una distribución normal

¿Qué ocurre si la perturbación no sigue una distribución Normal?

- Los estimadores MCO continúan siendo ELIO, pero no se distribuirán de forma Normal; por lo tanto tendremos problemas para hacer contrastes con los estadísticos habituales
- No sabremos la distribución teórica del t -ratio, ni del estadístico F

¿Cómo podemos contrastar si las u se distribuyen Normal?

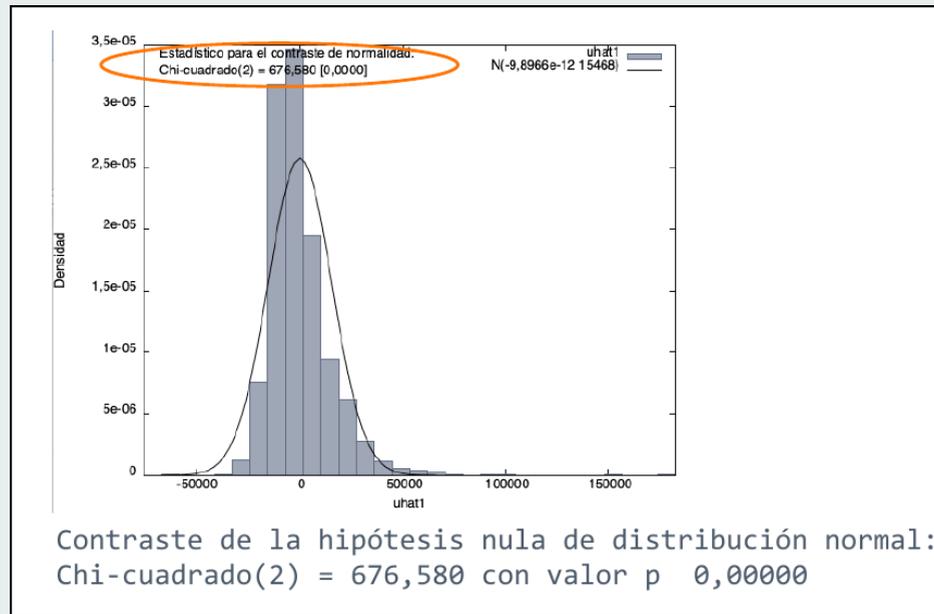
- Piensa que las perturbaciones **no son observables**
- Para contrastar alguna propiedad de u utilizaremos los residuos (\hat{u}), que si son observables una vez se ha estimado el modelo

Contraste de Normalidad

- Para chequear si u es Normal, **contrastaremos si \hat{u} siguen una Normal**
- Para la hipótesis de normalidad existen multitud de contrastes (Kolmogorov-Smirnov, Shapiro-Wilks, Jarque-Bera, ...)
- Utilizaremos el **contraste de Jarque-Bera (JB)**
- El contraste de Jarque-Bera se basa en que en cualquier distribución normal, el coeficiente de asimetría (**S**) es cero y el coeficiente de kurtosis (**K**) es 3.
- La hipótesis nula del contraste de JB es $H_0 : S = 0$ y $K = 3$; es decir la hipótesis nula es Normalidad.
- La alternativa es no-normalidad: $H_1 : S \neq 0$ ó $K \neq 3$
- El estadístico de JB es: $\frac{N}{6} \left(S^2 + \frac{(K-3)^2}{4} \right)$
- Se distribuye asintóticamente como una χ^2 con 2 grados de libertad

Si la muestra es suficientemente grande, el supuesto de normalidad de la perturbación no es necesario y se puede realizar inferencia aproximada

Contraste de Normalidad (ejemplo)



- El valor del estadístico de contraste es 676,58, que es mayor que 5,99 (el valor crítico de una χ^2_2 con dos grados de libertad al 5%). Por tanto rechazamos la hipótesis nula de normalidad de las perturbaciones
- El valor de probabilidad del contraste es 0,0000, es menor que 0.05, por lo tanto, rechazamos la hipótesis nula al 5%
- Como la muestra es de 1695 hogares, el incumplimiento de este supuesto no tiene consecuencias graves. Podemos hacer inferencia (aproximada)

7.4 Heterocedasticidad

Si tenemos heterocedasticidad, los estimadores MCO dejan de ser óptimos; es decir, ...

Si las perturbaciones no son homocedásticas, Los estimadores MCO continúan siendo insesgados, pero ya no serán óptimos

¿Recordáis que significa homocedasticidad? Sí, claro!!!!

4) $Var(u_i) = \sigma^2 \quad i = 1, \dots, N$ **(HOMOCEDASTICIDAD)**

- Si hay homocedasticidad, la variabilidad de las u es la misma para todos los individuos
 - En este caso hay que estimar $k + 1$ parámetros: k coeficientes β más la varianza de las perturbaciones, σ^2
- Si tenemos heterocedasticidad: $Var(u_i) = \sigma_i^2$, la volatilidad será diferente para cada observación.
 - En este caso habría que estimar $k + N$ parámetros. Obviamente esto es imposible dado que tenemos una muestra con N observaciones: tendríamos que suponer/imponer un esquema o patrón para la heterocedasticidad

Heterocedasticidad: causas

- **Errores de especificación:** la especificación errónea de un modelo tanto por omisión de variables como por errores en la forma funcional puede generar problemas de heterocedasticidad
- **Datos de corte transversal:** es muy frecuente que en datos de corte transversal haya heterocedasticidad. Menos frecuente, pero también puede ocurrir en datos de serie temporal.
- Si el modelo incorpora variables como la **renta, riqueza**, tamaño ... es bastante probable que haya heterocedasticidad. Estas variables generalmente implican que las restricciones a las que hacen frente los agentes económico no son uniformes.
- Podemos pensar que la causa última de que exista heterocedasticidad es "simplemente" que los **individuos son heterogeneos** y es imposible recoger en el modelo todos los aspectos o variables que afectan a su comportamiento y decisiones

Heterocedasticidad: ejemplos

- En el ejemplo del examen, habría heterocedasticidad si los estudiantes llegasen al examen con diferentes niveles de nerviosismo o si el examen lo corrigiesen dos profesores con mecanismos de corrección no homogéneos.
- En las decisiones de I+D+i, seguro que las empresas pequeñas tienen menos posibilidades de invertir, así que su nivel esperado de inversión será menor, pero seguramente también su volatilidad: las empresas grandes tienen más margen de acción pueden implementar un importante plan de inversión o no invertir.
- Demanda de bienes de "lujo", cabe esperar que conforme aumente la renta del consumidor, aumente el consumo, pero seguramente también la volatilidad.
- La heterocedasticidad es más probable que aparezca en datos de corte transversal: variables como la renta o el tamaño están muchas veces asociadas a heterocedasticidad.
- La heterocedasticidad también puede aparecer en datos de serie temporal; por ejemplo suele ser muy frecuente en series financieras

Heterocedasticidad: un ejemplo visual

$$\text{Var}(u_i) = \sigma_i^2 \leftarrow$$

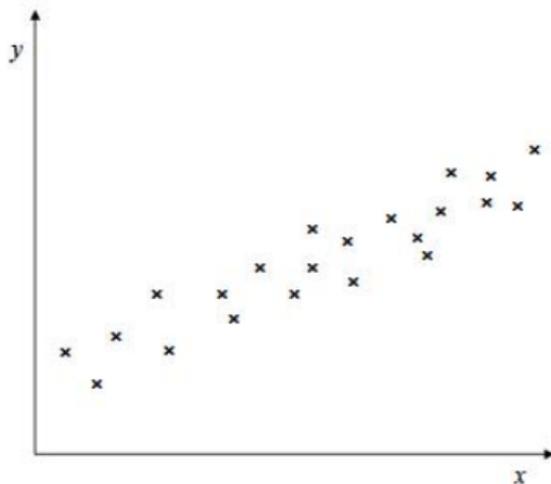


FIGURA 6.1. Diagrama de dispersión correspondiente a un modelo con perturbaciones homoscedásticas.

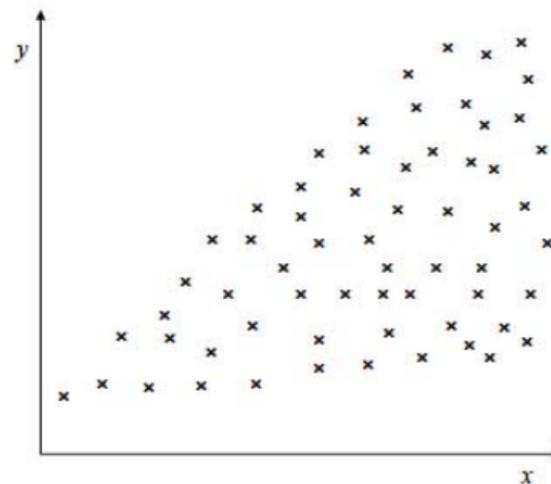


FIGURA 6.2. Diagrama de dispersión correspondiente a un modelo con perturbaciones heteroscedásticas.

Gráfica de Ezequiel Uriel

Heterocedasticidad: consecuencias

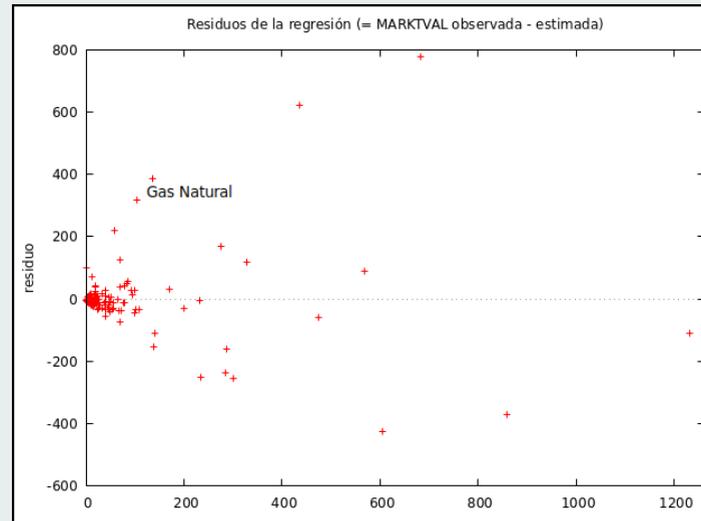
- Los estimadores MCO ya no son lo mejores, no son óptimos: hay un método mejor **MCG**
- La estimación de las varianzas de los estimadores obtenida aplicando la formula usual no es válida cuando existe heterocedasticidad
 - Por lo tanto, los estadísticos t y F basados en dicha estimación de la matriz de covarianzas darán lugar a inferencias erróneas.

Heterocedasticidad: ¿qué hacer?

- La solución teórica está clara: estimar por **MCG** (Mínimo Cuadrados Generalizados)
- Pero ... no siempre es factible
- Si se decide seguir estimando por MCO, aún sabiendo que no es lo óptimo, al menos habrá que calcular las varianzas de los estimadores de forma correcta, habrá que corregir las varianzas para hacerlas robustas a la heterocedasticidad (varianzas HC)

Heterocedasticidad: detección

- Como primera aproximación podemos visualizar los **residuos** MCO frente a \hat{y} o frente al regresor que creemos es causante del problema.



Heterocedasticidad: test de White

- La hipótesis nula es $H_0 : \sigma_i^2 = \sigma^2$ es decir, homocedasticidad
- La hipótesis alternativa es $H_1 : \sigma_i^2 = f(x' s)$ es decir, heterocedasticidad
- El estadístico de White es $W = N * R_{aux}^2$
- El estadístico de White se distribuye asintóticamente como una χ^2 con m grados de libertad. Siendo m el número de regresores de la regresión auxiliar (excluyendo el término independiente)

Supongamos que se ha estimado un modelo y se quiere contrastar si hay heterocedasticidad.

Por ejemplo, este modelo: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$

Para implementar el test de White tenemos que:

- 1) Estimar el modelo por MCO y obtener los residuos
- 2) Se ha de efectuar una regresión auxiliar, en la que el regresando sean los residuos MCO del modelo original al cuadrado, y los regresores sean: los regresores del modelo original, los cuadrados de estos y, opcionalmente, sus productos cruzados (eliminado redundancias). En nuestro ejemplo la regresión auxiliar sería:

$$\widehat{u}^2 = \alpha + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_4 x_2^2 + \gamma_5 x_3^2 + \gamma_6 x_2 x_3 + v_i$$

- 3) Se obtiene el coeficiente de determinación de la ecuación auxiliar R_{aux}^2 para después calcular el valor del estadístico de White $W = N * R_{aux}^2$
- 4) Se efectúa el contraste

Contraste de White: ejemplo

1. Se aplican MCO al modelo:

$$\log(\text{Precio}) = \beta_1 + \beta_2 \text{Superficie} + \beta_3 \text{Habitaciones} + u$$

Y se obtiene el modelo estimado:

$$\log(\widehat{\text{Precio}}_i) = 11,49 + 0,011 \text{Superficie}_i - 0,024 \text{Habitaciones}_i$$

2. Se estima la regresión auxiliar:

$$\hat{u}_i^2 = \alpha + \gamma_2 \text{Superficie}_i^2 + \gamma_3 \text{Habitaciones}_i^2 + \delta_2 \text{Superficie}_i \\ + \delta_3 \text{Habitaciones}_i + \lambda_1 \text{Superficie}_i \text{Habitaciones}_i + e_i$$

$$N = 322 \quad R_{aux}^2 = 0,0567$$

3. El estadístico de White $W = NR_{aux}^2 = 322 \cdot 0,0567 = 18,26$.

4) Los grados de libertad m son 5; por lo tanto hay que comparar el valor del estadístico con las tablas de una χ_5^2 . El valor de tablas, para $\alpha = 0.05$ es 11.07; por lo tanto, se rechaza la hipótesis nula de homocedasticidad

Por lo tanto ...

Contraste de White en Gretl

MCO, usando las observaciones 1-322

Variable dependiente: l_Precio

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	11,4891	0,108406	106,0	1,05e-250	***
Superficie	0,0112037	0,00114948	9,747	8,09e-20	***
Habitaciones	-0,0236781	0,0262914	-0,9006	0,3685	

Contraste de heterocedasticidad de White

MCO, usando las observaciones 1-322

Variable dependiente: uhat^2

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
const	0,305262	0,170462	1,791	0,0743	*
Superficie	0,0109924	0,00258070	4,259	2,71e-05	***
Habitaciones	-0,218235	0,0749136	-2,913	0,0038	***
sq_Superficie	2,31493e-05	1,92512e-05	1,202	0,2301	
X2_X3	-0,00226837	0,000811789	-2,794	0,0055	***
sq_Habitaciones	0,0333616	0,0102386	3,258	0,0012	***

R-cuadrado = 0,056725

Estadístico de contraste: $TR^2 = 18,265607$,
 con valor p = $P(\text{Chi-cuadrado}(5) > 18,265607) = 0,002631$

Y si tenemos heterocedasticidad ¿Qué hacemos?

- Los estimadores MCO ya no son óptimos, hay un método mejor, **MCG**; así que la solución teórica está clara: estimar por **MCG**, pero ... no siempre es factible
- Si tenemos heterocedasticidad pero continuamos estimando por MCO:
 - los coeficientes estimados son "válidos" (al menos los estimadores son insesgados), pero ...
 - la estimación de las desviaciones típicas no están correctamente calculados. La estimación de las varianzas de los estimadores obtenida aplicando la formula usual no es válida; por lo tanto, los estadísticos t y F basados en dicha estimación de la matriz de covarianzas darán lugar a inferencias erróneas.
- Es decir, si se decide seguir estimando por MCO, aún sabiendo que no es lo óptimo, al menos habrá que calcular las varianzas de los estimadores de forma correcta, habrá que corregir las varianzas para hacerlas **robustas bajo heterocedasticidad**
- Al corregir las desviaciones típicas la inferencia vuelve a ser válida

Varianzas robustas en Gretl

The screenshot displays the Gretl software interface, split into two windows. The left window, titled "gretl: especificar modelo", shows the process of selecting variables for a model. The dependent variable is "SALARI". The independent variables are "const", "EXPLAB", and "GENERE". The "GENERE" variable is highlighted in orange. The "Desviaciones típicas robustas" checkbox is checked, and the "HC1" option is selected. The right window, titled "gretl: modelo 1", displays the results of the regression analysis. The model is identified as "Modelo 1: MCO, usando las observaciones 1-200" with the dependent variable "SALARI". The results include a table of coefficients, standard deviations, t-statistics, and p-values for the independent variables, as well as summary statistics for the regression.

gretl: especificar modelo

MCO

const
EXPLAB
GENERE
SALARI

Variable de

SALARI

Selección

Regre

const
EXPLAB
GENERE

Desviaciones típicas robustas

HC1

gretl: modelo 1

Archivo Editar Contrastes Guardar Gráficos Análisis LaTeX

Modelo 1: MCO, usando las observaciones 1-200
Variable dependiente: SALARI
Desviaciones típicas robustas ante heterocedasticidad, variante HC1

	coeficiente	Desv. típica	Estadístico t	valor p
const	21992,1	1724,98	12,75	1,54e-27 ***
EXPLAB	443,037	72,7167	6,093	5,73e-09 ***
GENERE	2647,18	1448,50	1,828	0,0691 *

Media de la vble. dep. 31510,89 D.T. de la vble. dep. 10917,20
Suma de cuad. residuos 1,94e+10 D.T. de la regresión 9912,468
R-cuadrado 0,183879 R-cuadrado corregido 0,175594
F(2, 197) 20,45329 Valor p (de F) 8,49e-09
Log-verosimilitud -2122,586 Criterio de Akaike 4251,172
Criterio de Schwarz 4261,067 Crit. de Hannan-Quinn 4255,176

7.5 Autocorrelación

No lo vemos a ver en detalle, pero ... si tenemos una situación de autocorrelación ...

Autocorrelación en las perturbaciones

- No confundir con la colinealidad (o "correlación" entre los regresores)
- Hipótesis 5ª (NO autocorrelación): $Cov(u_i, u_j) = 0$ para $i \neq j$
- Habrá autocorrelación en las perturbaciones si: $Cov(u_i, u_j) \neq 0$
- La **autocorrelación** es una situación/problema diferente al de la heterocedasticidad, pero sus **consecuencias y posibles soluciones** son, en cierta forma, **similares a los de la heterocedasticidad**.
- Los **contrastos** estadísticos más habituales **para detectar autocorrelación** son: Durbin-Watson, Wallis, h-Durbin, Breusch-Godfrey, Box-Pierce-Ljung, ...

Autocorrelación: un ejemplo visual

- Si tenemos una situación de autocorrelación en las perturbaciones:
 $Cov(u_i, u_j) \neq 0$

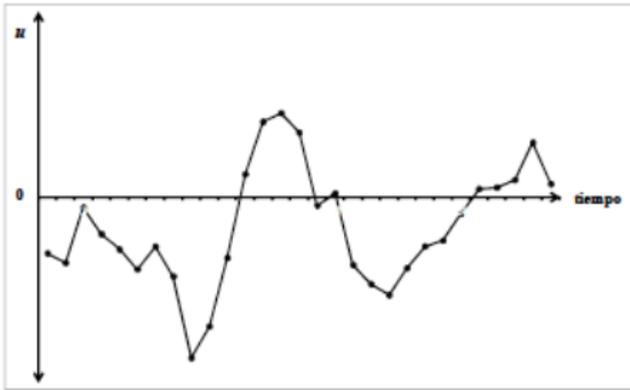


FIGURA 6.4. Gráfico de perturbaciones autocorrelacionadas positivamente.

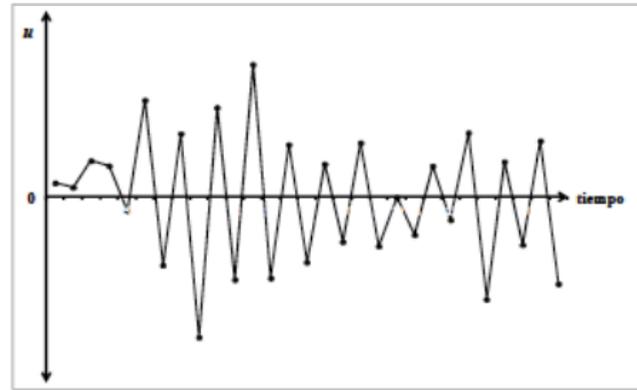


FIGURA 6.5. Gráfico de perturbaciones autocorrelacionadas negativamente.

Gráfica de Ezequiel Uriel

Autocorrelación: causas

- **Errores de especificación:** la especificación errónea de un modelo, tanto por omisión de variables como por errores en la forma funcional, puede generar problemas de autocorrelación
- **Datos temporales:** los fenómenos económicos suelen tener inercias o comportamientos cíclicos. Este comportamiento cíclico puede producir autocorrelación en las perturbaciones; de forma que, en las regresiones con datos de series temporales es muy probable que los valores de la perturbaciones dependan de los valores previos.
- **Transformación de datos:** es posible que al incluir en una regresión variables en diferencias o en tasas aparezcan problemas de autocorrelación.

Autocorrelación: consecuencias

- Si las perturbaciones presentan autocorrelación, al igual que en el caso de la heterocedasticidad, **los estimadores MCO continúan siendo insesgados, pero ya no serán óptimos**
- Los estimadores MCO no son óptimos: hay un método mejor **MCG**
- La estimación de las varianzas de los estimadores obtenida aplicando la fórmula usual no es válida cuando existe autocorrelación
- Por lo tanto, los estadísticos t y F basados en dicha estimación de la matriz de covarianzas darán lugar a inferencias erróneas

Autocorrelación: ¿qué hacer?

- Las soluciones son muy similares a cuando tenemos un problema de heterocedasticidad
- La solución teórica está clara: estimar por **MCG** (Mínimo Cuadrados Generalizados)
- Pero ... no siempre es posible
- Si se decide seguir estimando por MCO, aún sabiendo que no es lo óptimo, al menos habrá que ajustar el estimador de la matriz de varianzas y covarianzas de los estimadores de forma correcta, corrigiéndola para que sea consistente a la presencia de heterocedasticidad y autocorrelación, usando estimadores HAC: Heteroskedasticity and autocorrelation consistent.