UNIVERSIDAD DE GRANADA

Econometría

Temario de la asignatura

Ismael Sallami Moreno

Recursos Ingeniería Informática y Ade

Licencia

Econometría

Ismael Sallami Moreno

https://elblogdeismael.github.io

Índice general

Capítulo 1

Introducción a la econometría

Tema 1 del programa. Qué es la econometría, cómo se construye un modelo y con qué clase de datos se trabaja.

1.1 Econometría y modelos econométricos

La **econometría** mide relaciones económicas: toma una relación que la teoría propone, la escribe como un modelo, la estima con datos y contrasta si los datos la sostienen.

Reúne tres disciplinas, y ninguna basta por su cuenta:

Disciplina	Qué aporta
Teoría económica	qué variables se relacionan y en qué sentido
Estadística y probabilidad	cómo estimar y cómo medir la incertidumbre
Datos económicos	la evidencia sobre la que se estima

La teoría económica dice que el consumo depende de la renta y no dice cuánto. La econometría cuantifica ese *cuánto* y dice además con qué margen de error.

1.1.1 Modelo económico y modelo econométrico

Un **modelo económico** es una relación exacta entre variables:

$$C = f(R)$$

Un **modelo econométrico** añade una perturbación aleatoria y una forma funcional concreta:

$$C_i = \beta_1 + \beta_2 R_i + u_i$$

La diferencia está en u_i, y es toda la diferencia:

	Modelo económico	Modelo econométrico
Relación	exacta	estocástica
Ajuste a los datos	perfecto por construcción	imperfecto, y el residuo se mide
Contrastable	no	sí

1.1.2 La perturbación

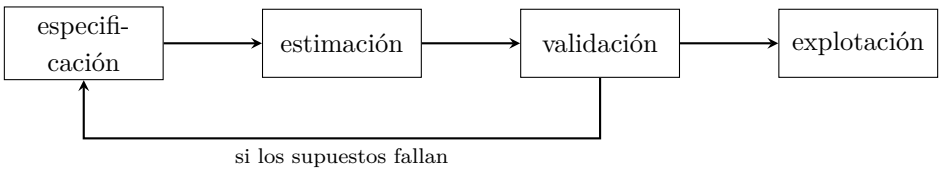
u_i recoge todo lo que afecta a C_i y no está en el modelo. Sus fuentes, y conviene distinguirlas porque no todas se corrigen igual:

Fuente	Qué es
Variables omitidas	factores que influyen y no se han incluido
Errores de medida	la variable observada no es exactamente la teórica
Forma funcional incorrecta	la relación no es lineal y se ha supuesto que sí
Comportamiento aleatorio	las decisiones humanas no son deterministas

La primera es la peligrosa. Si una variable omitida está correlacionada con las incluidas, la perturbación también lo está, y el estimador del tema 2 deja de ser insesgado. Es el **sesgo por variable omitida**, y ningún contraste de los temas 4 a 6 lo detecta: hay que pensarlo al especificar.

1.2 Fases del método econométrico

Fase	Qué se hace
1 · Especificación	elegir variables, forma funcional y estructura de la perturbación
2 · Estimación	calcular los parámetros a partir de la muestra
3 · Validación	contrastar hipótesis y comprobar los supuestos
4 · Explotación	predecir, simular políticas, interpretar



El proceso **no es lineal**: la validación devuelve a la especificación cuando los contrastes rechazan un supuesto, y ese ciclo es donde se pasa la mayor parte del trabajo.

Y una advertencia sobre ese ciclo: cuantas más especificaciones se prueban sobre la misma muestra, más fácil es encontrar una que pase todos los contrastes por azar. Un modelo elegido tras cien pruebas no tiene los niveles de significación que dice tener. La defensa es que la especificación la guíe la teoría, no la búsqueda del mejor ajuste.

1.3 Componentes de un modelo econométrico

Componente	Notación	Qué es
Variable endógena	Y	lo que el modelo explica
Variables exógenas	$X_2, \ldots, X_k$	lo que explica
Parámetros	$\beta_1, \ldots, \beta_k$	lo que se estima
Perturbación	u	lo no observable

El modelo lineal general con k regresores y n observaciones:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i, \qquad i = 1, \ldots, n$$

Y en forma matricial, que es como se trabaja en los temas siguientes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

con $\mathbf{y}$ de orden $n \times 1$, $\mathbf{X}$ de orden $n \times k$ —su primera columna de unos, para el término independiente—, $\boldsymbol{\beta}$ de orden $k \times 1$ y $\mathbf{u}$ de orden $n \times 1$.

1.3.1 Interpretación de los parámetros

β_j es el efecto sobre Y de aumentar X_j en una unidad **manteniendo constantes las demás variables del modelo**. Esa cláusula es lo que distingue el efecto parcial de la correlación simple, y es la razón de estimar con todas las variables a la vez en vez de una a una.

1.3.2 Formas funcionales

El modelo es lineal **en los parámetros**, no necesariamente en las variables. Eso deja sitio a relaciones no lineales:

Forma	Modelo	Interpretación de β_2
Lineal	$Y = \beta_1 + \beta_2 X$	variación de Y por unidad de X
Log-log	$\ln Y = \beta_1 + \beta_2 \ln X$	elasticidad: variación porcentual por variación porcentual
Log-lineal	$\ln Y = \beta_1 + \beta_2 X$	variación porcentual de Y por unidad de X
Lineal-log	$Y = \beta_1 + \beta_2 \ln X$	variación de Y por variación porcentual de X
Cuadrática	$Y = \beta_1 + \beta_2 X + \beta_3 X^2$	el efecto marginal es $\beta_2 + 2\beta_3 X$

La **log-log** es la más usada en economía porque su parámetro es directamente una elasticidad, que es la magnitud con la que la teoría suele razonar.

La **cuadrática** es la primera en la que el efecto de X depende del propio X. Ahí no se puede leer β_2 como el efecto: hay que derivar. Y el punto donde el efecto cambia de signo es $X^* = -\beta_2/(2\beta_3)$.

Lo que **no** se puede tratar así es un modelo no lineal en los parámetros, como $Y = \beta_1 X^{\beta_2} + u$: ahí no hay transformación que lo linealice y hacen falta métodos de estimación distintos.

1.4 Naturaleza de la información

Tipo de datos	Qué es	Problema característico
Serie temporal	una unidad observada a lo largo del tiempo	autocorrelación (tema 6)

Tipo de datos	Qué es	Problema característico
Sección cruzada	muchas unidades en un mismo momento	heteroscedasticidad (tema 5)
Datos de panel	muchas unidades a lo largo del tiempo	los dos, y la heterogeneidad no observable

La correspondencia entre el tipo de datos y el problema no es casual, y ordena la segunda mitad del temario. En una serie temporal, lo que ocurre en un periodo arrastra al siguiente, así que las perturbaciones están correlacionadas entre sí. En una sección cruzada, unidades de tamaños muy distintos —empresas, hogares, países— tienen dispersiones muy distintas, y de ahí la varianza no constante.

1.4.1 Cualidades de los datos

Cualidad	Qué significa
Escala	nominal, ordinal, de intervalo o de razón
Frecuencia	anual, trimestral, mensual, diaria
Fuente	primaria, recogida por el investigador; secundaria, publicada
Precios	corrientes o constantes

Dos cuidados que cambian los resultados y no dan ningún aviso:

- **Trabajar con precios corrientes en una serie larga** mete la inflación dentro del parámetro estimado, que deja de medir lo que se pretendía.
- **Agregar** oculta el comportamiento individual. Una relación que se cumple entre agregados puede no cumplirse para ninguna unidad, y al revés.

1.4.2 Variables cualitativas

Se incorporan con **variables ficticias**, que valen 1 si se da la característica y 0 si no. Con m categorías se incluyen $m-1$ ficticias, y la que se deja fuera es la de referencia: su efecto va dentro del término independiente.

Incluir las m junto con el término independiente produce **multicolinealidad exacta** —las ficticias suman exactamente la columna de unos— y el modelo no se puede estimar. Es la trampa de las ficticias, y es el caso extremo del tema 4.

El planteamiento del método econométrico y la naturaleza de los datos siguen a Gujarati, 2010 y Wooldridge, 2010; el tratamiento de las formas funcionales, también a Novales, 2000 y Stock y Watson, 2012.

Capítulo 2

El modelo lineal I

Tema 2 del programa. Las hipótesis del modelo, la estimación por mínimos cuadrados ordinarios y la medida de la bondad del ajuste.

2.1 Hipótesis del modelo

El modelo, en forma matricial:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Sobre él se hacen estas hipótesis, y cada una tiene su tema en el resto del programa:

	Hipótesis	Si falla
H1	linealidad en los parámetros	error de especificación
H2	$\mathbf{X}$ es no estocástica y de rango completo, $\text{rg}(\mathbf{X}) = k$	multicolinealidad (tema 4)
H3	$\text{E}[\mathbf{u}] = \mathbf{0}$	sesgo en el término independiente
H4	homoscedasticidad: $\text{Var}(u_i) = \sigma^2$ para todo i	heteroscedasticidad (tema 5)
H5	ausencia de autocorrelación: $\text{Cov}(u_i, u_j) = 0$ si $i \neq j$	autocorrelación (tema 6)
H6	$n > k$	el modelo no se puede estimar
H7	normalidad: $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$	la inferencia del tema 3 no vale en muestras pequeñas

H4 y H5 juntas se escriben como

$$\text{E}[\mathbf{u}\mathbf{u}'] = \sigma^2\mathbf{I}_n$$

es decir, matriz de varianzas y covarianzas escalar: la misma varianza en la diagonal y ceros fuera. Los temas 5 y 6 son cada uno una forma de romper esa igualdad: el 5 cambia la diagonal, el 6 llena lo de fuera.

H7 no hace falta para estimar, solo para la inferencia exacta del tema 3. En muestras grandes se puede prescindir de ella, porque los estadísticos convergen a sus distribuciones asintóticas.

2.2 Estimación por mínimos cuadrados ordinarios

El criterio: elegir $\hat{\boldsymbol{\beta}}$ que minimice la suma de los residuos al cuadrado.

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} e_i^2 = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Derivando e igualando a cero se obtienen las **ecuaciones normales**:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

y, si $\mathbf{X}'\mathbf{X}$ es invertible —que es H2—, el estimador:

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

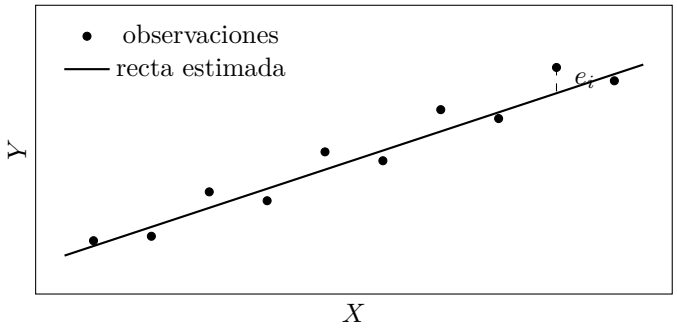
Los residuos no son las perturbaciones. u_i es lo no observable del modelo verdadero; $e_i = y_i - \hat{y}_i$ es lo que queda tras estimar, y es observable. Todos los contrastes de los temas 4 a 6 se hacen sobre los residuos, que es lo único que hay, y de ahí sus limitaciones.

2.2.1 El caso de dos variables

Con $Y_i = \beta_1 + \beta_2 X_i + u_i$ las fórmulas se escriben en escalares:

$$\hat{\beta}_2 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{S_{XY}}{S_{XX}}, \qquad \hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$$

La segunda fórmula dice que **la recta pasa siempre por el punto medio** $(\bar{X}, \bar{Y})$.



Se minimizan las distancias **verticales**, no las perpendiculares. Es una decisión con consecuencias: hace que la regresión de Y sobre X y la de X sobre Y sean rectas distintas, porque cada una minimiza en una dirección.

2.2.2 Propiedades algebraicas

Salen de las ecuaciones normales y se cumplen siempre, sin necesidad de ninguna hipótesis estadística:

Propiedad	Expresión	Condición
Los residuos suman cero	$\sum e_i = 0$	si hay término independiente
Los residuos son ortogonales a los regresores	$\mathbf{X}'\mathbf{e} = \mathbf{0}$	siempre

Propiedad	Expresión	Condición
La recta pasa por las medias	$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$	si hay término independiente
La media de lo ajustado es la de lo observado	$\bar{\hat{y}} = \bar{y}$	si hay término independiente

Las condiciones no son un detalle: **un modelo sin término independiente no cumple tres de las cuatro**, y por eso su coeficiente de determinación puede salir negativo y no se puede comparar con el de un modelo que sí lo tiene.

2.3 Propiedades del estimador

2.3.1 Insesgadez

Sustituyendo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$:

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

y tomando esperanzas, con H2 y H3:

$$\mathrm{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

Solo hacen falta H2 y H3. **La insesgadez no depende de H4 ni de H5**, y de ahí sale la conclusión que gobierna los temas 5 y 6: con heteroscedasticidad o autocorrelación el estimador sigue siendo insesgado, y lo que se rompe es su varianza.

2.3.2 Varianza

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

Aquí sí hacen falta H4 y H5. Y la expresión dice de qué depende la precisión:

- **Menos varianza de la perturbación**, más precisión.
- **Más variabilidad en los regresores**, más precisión: si X apenas varía en la muestra, su efecto no se puede medir.
- **Menos correlación entre regresores**, más precisión. Es el tema 4.

2.3.3 Estimación de σ^2

$$\hat{\sigma}^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k} = \frac{\mathrm{SCR}}{n-k}$$

El divisor es $n-k$ y no n porque se han consumido k grados de libertad al estimar los parámetros. Con ese divisor el estimador es insesgado; con n sale sesgado a la baja, y el sesgo es tanto mayor cuantos más regresores haya.

2.3.4 El teorema de Gauss-Markov

Bajo H1 a H6, el estimador de mínimos cuadrados ordinarios es el de menor varianza dentro de los estimadores lineales e insesgados.

Las tres condiciones importan y se citan juntas por costumbre:

- **Lineal**: en $\mathbf{y}$.
- **Insesgado**: su esperanza es $\boldsymbol{\beta}$.
- **Óptimo**: mínima varianza dentro de esa clase.

No dice que sea el mejor estimador posible. Fuera de la clase de los lineales e insesgados puede haber otros mejores, y aceptando un poco de sesgo se puede reducir mucho la varianza: eso es lo que hace la regresión contraída del tema 4.

Y el teorema **no necesita normalidad**. H7 solo hace falta para la inferencia del tema 3.

2.4 Bondad del ajuste

2.4.1 Descomposición de la varianza

$$\underbrace{\sum (y_i - \bar{y})^2}_{\text{SCT}} = \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{\text{SCE}} + \underbrace{\sum e_i^2}_{\text{SCR}}$$

La descomposición **solo se cumple si hay término independiente**, porque necesita que los residuos sumen cero.

2.4.2 Coeficiente de determinación

$$R^2 = \frac{\text{SCE}}{\text{SCT}} = 1 - \frac{\text{SCR}}{\text{SCT}}, \qquad 0 \le R^2 \le 1$$

Es la proporción de la variación de Y que el modelo explica. Y tiene un defecto grave: R^2 **nunca baja al añadir un regresor**, sea cual sea. Añadir ruido puro como variable explicativa lo sube un poco, así que maximizar R^2 lleva a modelos sobrecargados.

2.4.3 Coeficiente corregido

$$\bar{R}^2 = 1 - \frac{\text{SCR}/(n-k)}{\text{SCT}/(n-1)} = 1 - (1 - R^2)\,\frac{n-1}{n-k}$$

Penaliza por el número de parámetros: solo sube si el regresor nuevo aporta más de lo que cuesta en grados de libertad. Puede salir **negativo**, y eso indica un modelo peor que la simple media.

2.4.4 Criterios de información

Formalizan el mismo compromiso: ajuste frente a número de parámetros.

$$\text{AIC} = \ln\left(\frac{\text{SCR}}{n}\right) + \frac{2k}{n}, \qquad \text{SBC} = \ln\left(\frac{\text{SCR}}{n}\right) + \frac{k \ln n}{n}$$

En los dos, **menor es mejor**. La diferencia está en la penalización: 2 por parámetro en Akaike, $\ln n$ en Schwarz. Como $\ln n > 2$ para $n > 7$, **Schwarz penaliza más y elige modelos más pequeños**, y esa diferencia crece con el tamaño de la muestra.

Tres cuidados al comparar modelos:

- **Solo se comparan modelos con la misma variable endógena.** Un modelo en Y y otro

en $\ln Y$ no son comparables por R^2 ni por AIC: sus SCT son cantidades distintas.
- **Un R^2 alto no valida el modelo.** En series temporales dos variables con tendencia dan R^2 altísimos sin ninguna relación real. Es la regresión espuria, y el tema 6 la trata.
- **Un R^2 bajo no lo invalida.** En sección cruzada con datos individuales, valores de 0,2 son normales y los parámetros pueden estar perfectamente estimados.

El desarrollo del modelo lineal y de las propiedades de mínimos cuadrados sigue a Gujarati, 2010, Novales, 2000 y Greene, 1999; el tratamiento de los criterios de selección, a Wooldridge, 2010 y Maddala, 2001.

Capítulo 3

El modelo lineal II

Tema 3 del programa. Intervalos de confianza, contrastes sobre los parámetros y explotación del modelo.

3.1 De dónde sale la inferencia

Con H7 —normalidad de la perturbación— el estimador es normal:

$$\hat{\boldsymbol{\beta}} \sim N\big(\boldsymbol{\beta},\ \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\big)$$

Como σ^2 no se conoce y se sustituye por $\hat{\sigma}^2$, el estadístico tipificado deja de ser normal y pasa a ser una t de Student con $n-k$ grados de libertad:

$$t_j = \frac{\hat{\beta}_j - \beta_j}{\mathrm{ee}(\hat{\beta}_j)} \sim t_{n-k}, \qquad \mathrm{ee}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2\,[(\mathbf{X}'\mathbf{X})^{-1}]_{jj}}$$

Toda la inferencia del tema sale de ahí, y por eso **todo lo que rompa H4 o H5 invalida el error estándar** y con él los intervalos y los contrastes. Los temas 5 y 6 vuelven sobre esto.

3.2 Intervalos de confianza

$$\hat{\beta}_j \pm t_{\alpha/2,\,n-k}\ \mathrm{ee}(\hat{\beta}_j)$$

Lectura correcta: si se repitiera el muestreo muchas veces, el $100(1-\alpha)\,\%$ de los intervalos construidos así contendría el verdadero β_j. **No** es que β_j esté en este intervalo con probabilidad 0,95: β_j es un número fijo, y lo aleatorio es el intervalo.

Su amplitud crece con $\hat{\sigma}^2$ y con la correlación entre regresores, y decrece con la variabilidad de X_j y con n. Es la misma lista de la varianza del tema 2.

Para σ^2 el intervalo se construye con la χ^2:

$$\left[\frac{(n-k)\hat{\sigma}^2}{\chi^2_{\alpha/2,\,n-k}},\ \frac{(n-k)\hat{\sigma}^2}{\chi^2_{1-\alpha/2,\,n-k}}\right]$$

y es asimétrico, porque la χ^2 lo es.

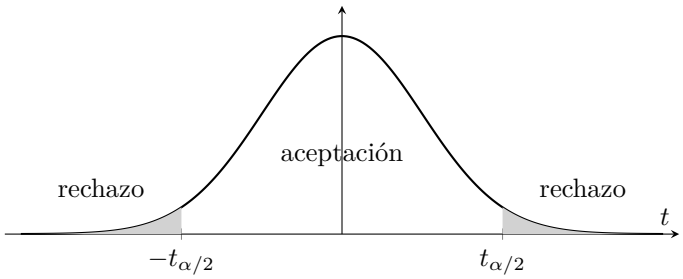
3.3 Contrastes sobre los parámetros

3.3.1 Contraste individual

$$H_0 : \beta_j = 0 \quad \text{frente a} \quad H_1 : \beta_j \neq 0$$

$$t = \frac{\hat{\beta}_j}{\text{ee}(\hat{\beta}_j)}$$

Se rechaza si $|t| > t_{\alpha/2,\, n-k}$. Rechazar significa que la variable es **significativa**: aporta al modelo dado el resto de variables incluidas.



Dos precisiones que se piden en los exámenes:

- **No rechazar no es aceptar.** Que t no supere el valor crítico puede deberse a que el efecto no existe o a que la muestra no da para detectarlo. Con multicolinealidad, el tema 4 produce exactamente ese segundo caso.
- **Significativo no es importante.** Con una muestra grande, un efecto económicamente irrelevante sale significativo. Hay que mirar además el tamaño del coeficiente.

3.3.2 Contraste de significación global

$$H_0 : \beta_2 = \beta_3 = \cdots = \beta_k = 0$$

$$F = \frac{\text{SCE}/(k-1)}{\text{SCR}/(n-k)} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \sim F_{k-1,\, n-k}$$

El término independiente queda fuera de H_0: lo que se contrasta es si el modelo explica algo más que la media.

El contraste conjunto no es la suma de los individuales. Rechazar H_0 con la F y no rechazar ninguno con la t es perfectamente posible, y es el síntoma clásico de la multicolinealidad del tema 4: las variables explican en conjunto y no se puede separar la contribución de cada una.

3.3.3 Contrastes de restricciones lineales

El caso general: q restricciones escritas como $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, con $\mathbf{R}$ de orden $q \times k$.

$$F = \frac{(\text{SCR}_R - \text{SCR}_{NR})/q}{\text{SCR}_{NR}/(n-k)} \sim F_{q,\, n-k}$$

donde SCR_R es la del modelo restringido y SCR_{NR} la del libre. La idea es directa: imponer una

restricción falsa empeora mucho el ajuste, y una verdadera apenas lo empeora.

Restricción	$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$	q
$\beta_2 = 0$	$(0\ 1\ 0\ \cdots)\boldsymbol{\beta} = 0$	1
$\beta_2 = \beta_3$	$(0\ 1\ -1\ 0\ \cdots)\boldsymbol{\beta} = 0$	1
$\beta_2 + \beta_3 = 1$	$(0\ 1\ 1\ 0\ \cdots)\boldsymbol{\beta} = 1$	1
$\beta_2 = \beta_3 = 0$	dos filas	2

La tercera fila es la de **rendimientos constantes a escala** en una función de producción Cobb-Douglas estimada en logaritmos, y es el ejemplo canónico de por qué hacen falta estos contrastes: la teoría propone una relación entre parámetros, no un valor para cada uno.

Con $q = 1$ se cumple $F = t^2$, así que el contraste individual es un caso particular de este.

3.3.4 Cambio estructural

El contraste de Chow comprueba si los parámetros son los mismos en dos submuestras:

$$F = \frac{(\mathrm{SCR}_T - \mathrm{SCR}_1 - \mathrm{SCR}_2)/k}{(\mathrm{SCR}_1 + \mathrm{SCR}_2)/(n_1 + n_2 - 2k)}$$

Necesita $n_1 > k$ y $n_2 > k$, y supone la misma varianza en los dos tramos. Si la varianza cambia, lo que rechaza puede ser la heteroscedasticidad del tema 5 y no el cambio de parámetros.

Una alternativa que evita esa confusión es usar variables ficticias: se define una ficticia para el segundo tramo, se incluyen sus productos con los regresores y se contrasta que sus coeficientes son cero. Además dice **qué** parámetro cambió, y no solo que alguno lo hizo.

3.4 Explotación del modelo

3.4.1 Predicción

Para un vector de valores $\mathbf{x}_0$:

$$\hat{y}_0 = \mathbf{x}_0'\hat{\boldsymbol{\beta}}$$

Y hay dos intervalos distintos, que se confunden con frecuencia:

Qué se predice	Varianza
La **media** de Y dado $\mathbf{x}_0$	$\sigma^2\, \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$
Un **valor individual** de Y	$\sigma^2\, [\,1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\,]$

El sumando 1 de la segunda fila es la varianza de la perturbación del período que se predice, que no desaparece por bien estimado que esté el modelo. Por eso el intervalo individual es siempre más ancho, y no se estrecha aunque n crezca: hay un suelo irreducible.

Las dos varianzas crecen conforme $\mathbf{x}_0$ se aleja de la media muestral. De ahí la regla práctica: **extrapolar lejos del rango de los datos es poco fiable**, y el propio intervalo lo indica al ensancharse.

3.4.2 Medidas de error de predicción

Medida	Fórmula
Error cuadrático medio	$\text{ECM} = \frac{1}{m}\sum(y_t - \hat{y}_t)^2$
Raíz del error cuadrático medio	$\sqrt{\text{ECM}}$, en las unidades de Y
Error absoluto medio	$\frac{1}{m}\sum\lvert y_t - \hat{y}_t\rvert$
Error porcentual absoluto medio	$\frac{100}{m}\sum\lvert(y_t - \hat{y}_t)/y_t\rvert$

La raíz del error cuadrático medio es la más usada porque va en las unidades de la variable y se compara con su media. Y el error porcentual falla cuando algún y_t se acerca a cero, porque el cociente explota.

Y una advertencia de método: **el error de predicción debe medirse fuera de la muestra de estimación**. Medido dentro, siempre sale bueno, y un modelo con muchos regresores sale mejor cuanto más sobreajustado esté.

3.4.3 Simulación

Con el modelo estimado se puede evaluar el efecto de una política: cambiar el valor de una variable exógena y ver qué predice el modelo.

Su límite es la **crítica de Lucas**: los parámetros estimados con datos de un régimen pueden cambiar cuando la política cambia, porque los agentes reaccionan al cambio de reglas. Un modelo estimado bajo unas reglas no predice bien lo que pasará bajo otras distintas, y eso ningún contraste lo detecta.

Los intervalos, los contrastes y la predicción siguen a Gujarati, 2010, Novales, 2000 y Wooldridge, 2010; el tratamiento matricial de las restricciones lineales, a Greene, 1999 y Johnston y DiNardo, 2001.

Capítulo 4

Multicolinealidad

Tema 4 del programa. Qué pasa cuando los regresores están relacionados entre sí: concepto, causas, consecuencias, detección y soluciones.

4.1 Concepto

La **multicolinealidad** es la existencia de relaciones lineales entre las columnas de $\mathbf{X}$. Es un incumplimiento de H2, y tiene dos grados:

Grado	Qué ocurre	Consecuencia
Exacta	una columna es combinación lineal exacta de otras	$\mathbf{X}'\mathbf{X}$ es singular y el modelo no se puede estimar
Aproximada	la relación es fuerte pero no exacta	se puede estimar, y muy mal

La exacta se detecta sola: el programa se detiene o elimina una variable. La **aproximada** es la que importa, porque el modelo se estima, imprime resultados con aspecto normal y esos resultados no valen.

Es un problema de la muestra, no del modelo. Las variables pueden estar relacionadas en estos datos y no estarlo en otros. No es un error de especificación ni una violación de una hipótesis sobre la perturbación: es falta de información para separar dos efectos.

4.2 Causas

Causa	Ejemplo
Variables que se mueven juntas por naturaleza	renta y riqueza, consumo e ingreso
Tendencia común en series temporales	casi todo lo macroeconómico crece con el tiempo
Variables redundantes	incluir la misma magnitud en dos escalas
Retardos de una misma variable	X_t y X_{t-1} están muy correlacionados
Potencias de una variable	X y X^2 en un rango estrecho
Muestra pequeña	con pocos datos casi todo parece correlacionado
La trampa de las ficticias	m ficticias y término independiente: colinealidad exacta

La última es un error de especificación y no un problema de los datos: se corrige dejando fuera una categoría, como dice el tema 1.

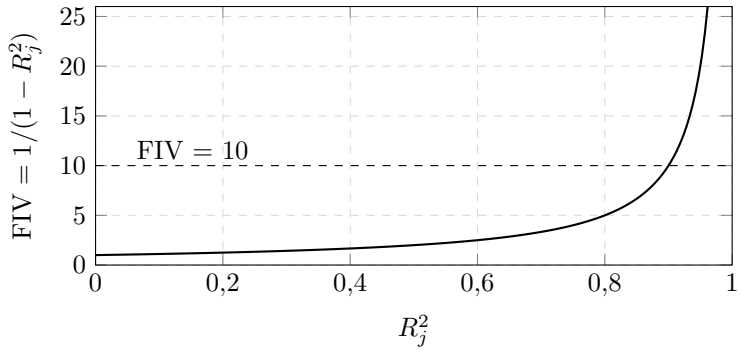
4.3 Consecuencias

Lo primero, y conviene decirlo antes que nada: **el estimador sigue siendo insesgado y óptimo**. Gauss-Markov solo necesita $\text{rg}(\mathbf{X}) = k$, y con multicolinealidad aproximada eso se cumple. El problema no es el sesgo.

El problema es la **varianza**, y se ve en el factor de inflación:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2}{S_{jj}\,(1 - R_j^2)}$$

con R_j^2 el coeficiente de determinación de la regresión de X_j sobre las demás explicativas. Cuando $R_j^2 \to 1$, la varianza se dispara.



De ahí los efectos observables:

Efecto	Por qué
Errores estándar grandes	la varianza está inflada
Estadísticos t pequeños	el denominador es grande
F significativa con todas las t no significativas	el conjunto explica; cada uno por separado no se distingue
Coeficientes con signo contrario al esperado	la estimación es imprecisa y salta de signo
Estimaciones muy sensibles	quitar una observación o una variable las cambia mucho
Intervalos de confianza enormes	son proporcionales al error estándar

La tercera fila es el síntoma característico, y es la que hay que reconocer: si F rechaza y ninguna t lo hace, la primera sospecha es multicolinealidad.

Y algo que la multicolinealidad **no** estropea: la **predicción**. Si los regresores mantienen la misma relación entre sí en el período que se predice, el modelo predice bien aunque no se pueda decir cuánto aporta cada variable. Lo que se pierde es la interpretación de los coeficientes, no el ajuste.

4.4 Detección

Ningún procedimiento la detecta con un sí o un no, porque es cuestión de grado.

4.4.1 Matriz de correlaciones

Correlaciones altas entre pares de regresores, por encima de 0,8, son un indicio. **No es concluyente**: puede haber multicolinealidad grave con todas las correlaciones simples moderadas, si la relación involucra a tres o más variables a la vez.

4.4.2 Factor de inflación de la varianza

$$\mathrm{FIV}_j = \frac{1}{1 - R_j^2}, \qquad \mathrm{Tolerancia}_j = 1 - R_j^2$$

FIV	Lectura habitual
1	ninguna relación con las demás
5	atención
10	problema serio, $R_j^2 = 0{,}9$

Es el procedimiento más informativo porque **identifica qué variable** está implicada, y no solo que hay un problema. El umbral de 10 es convencional, no una frontera con significado estadístico.

4.4.3 Regresiones auxiliares

Regresar cada X_j sobre las demás y contrastar su significación conjunta con la F. Es lo mismo que el FIV, con un contraste formal encima.

4.4.4 Número de condición

$$\kappa = \sqrt{\frac{\lambda_{\text{máx}}}{\lambda_{\text{mín}}}}$$

con λ los valores propios de $\mathbf{X}'\mathbf{X}$. Valores de κ entre 10 y 30 indican multicolinealidad moderada, y por encima de 30, grave.

Las variables deben estar tipificadas antes de calcularlo: si no, un simple cambio de unidades cambia el diagnóstico, que es justo lo que no debe pasar.

4.5 Soluciones

Ordenadas de más a menos recomendables:

4.5.1 No hacer nada

Es una respuesta legítima y a menudo la mejor. Si el objetivo es predecir, si los coeficientes que interesan están bien estimados, o si las variables afectadas son controles cuyo valor no se va a interpretar, la multicolinealidad no impide nada.

4.5.2 Ampliar la muestra

Es la solución de fondo: el problema es falta de información, y más datos la aportan. Con n mayor, S_{jj} crece y la varianza baja.

Su límite es evidente: los datos económicos suelen no estar disponibles, y en series temporales alargar la muestra hacia atrás puede introducir el cambio estructural del tema 3.

4.5.3 Eliminar variables

Quitar una de las variables correlacionadas resuelve el síntoma **y puede introducir sesgo por variable omitida**, que es peor: se cambia un problema de varianza por uno de sesgo, y el sesgo no se ve en los resultados.

Solo es aceptable si la teoría admite que esa variable sobre. Eliminar por criterio estadístico una variable que la teoría exige es sustituir un modelo mal estimado por un modelo mal especificado.

4.5.4 Transformar las variables

- **Primeras diferencias** en series temporales: $\Delta X_t = X_t - X_{t-1}$ elimina la tendencia común, que es la causa más frecuente. A cambio puede introducir autocorrelación, que es el tema 6.
- **Cocientes**: usar magnitudes por habitante o por unidad de producto en vez de los totales.
- **Combinar variables** en un índice, cuando miden lo mismo.
- **Centrar** antes de elevar al cuadrado: X y X^2 dejan de estar tan correlacionados si se usa $(X - \bar{X})^2$.

4.5.5 Regresión contraída

$$\hat{\boldsymbol{\beta}}_{\text{ridge}} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}$$

Sumar λ a la diagonal hace la matriz invertible y estable. El estimador es **sesgado**, y a cambio su varianza es mucho menor; con un λ adecuado su error cuadrático medio es menor que el de mínimos cuadrados.

Es el caso que muestra el alcance real de Gauss-Markov del tema 2: el teorema garantiza mínima varianza **dentro de los insesgados**, y saliendo de esa clase se puede hacer mejor. El precio es elegir λ, que no tiene una regla cerrada.

4.5.6 Componentes principales

Sustituir los regresores por combinaciones lineales suyas ortogonales entre sí, con lo que la multicolinealidad desaparece por construcción. El inconveniente es que las componentes no tienen interpretación económica, así que se pierde exactamente lo que la multicolinealidad estropeaba.

4.6 Un procedimiento de trabajo

1. Calcular la matriz de correlaciones y los FIV.
2. Si algún FIV pasa de 10, identificar qué variables lo producen.
3. Preguntarse **para qué es el modelo**: si es para predecir, seguir.
4. Si es para interpretar, intentar ampliar la muestra o transformar variables.
5. Eliminar una variable solo si la teoría lo permite, y decirlo.
6. Documentar el diagnóstico en el informe, aunque se decida no hacer nada.

El paso 6 no es burocracia: unos coeficientes con signo raro y unas t pequeñas se interpretan de una manera si hay multicolinealidad y de otra si no, y quien lea el informe necesita saberlo.

El tratamiento de la multicolinealidad y sus diagnósticos sigue a Gujarati, 2010 y Novales, 2000; el

número de condición y la regresión contraída, a Greene, 1999 y Maddala, 2001; los ejercicios de aplicación, a García et al., 2017.

Capítulo 5

Heteroscedasticidad

Tema 5 del programa. Cuando la varianza de la perturbación no es constante: concepto, causas, consecuencias, contrastes y estimación.

5.1 Concepto

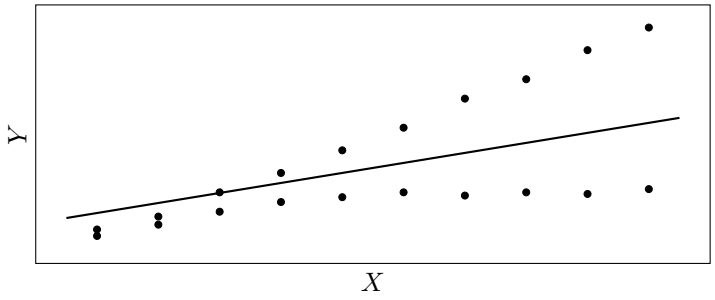
H4 exige que $\mathrm{Var}(u_i) = \sigma^2$ para toda observación. Hay **heteroscedasticidad** cuando esa varianza cambia con i:

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

La matriz de varianzas y covarianzas deja de ser escalar y pasa a ser diagonal con elementos distintos:

$$\mathrm{E}[\mathbf{u}\mathbf{u}'] = \sigma^2\boldsymbol{\Omega}, \qquad \boldsymbol{\Omega} = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n \end{pmatrix}$$

Sigue siendo diagonal, y esa es la diferencia con el tema 6: aquí lo que cambia es la diagonal y los ceros de fuera se mantienen, porque las perturbaciones siguen siendo independientes entre sí.



La nube se abre en abanico: la dispersión alrededor de la recta crece con X. Es la forma en que la heteroscedasticidad se ve, y por eso el primer diagnóstico es siempre un gráfico de residuos.

5.2 Causas

Causa	Ejemplo
Escala muy distinta entre unidades	gasto de hogares con rentas muy diferentes

Causa	Ejemplo
Aprendizaje	los errores se reducen conforme se gana experiencia
Mejora en la recogida de datos	series largas cuya calidad de medida mejora
Valores atípicos	unas pocas observaciones muy alejadas
Datos agregados con distinto número de unidades	medias de grupos de tamaño desigual
Error de especificación	omitir una variable o usar una forma funcional errónea

La última merece atención: **un modelo mal especificado produce residuos con apariencia heteroscedástica**. Si se omite una variable relevante o la relación es en logaritmos y se estima en niveles, los contrastes de este tema rechazan la homoscedasticidad. La corrección entonces no es ponderar, sino arreglar la especificación.

Y es sobre todo un problema de **sección cruzada**, por la razón del tema 1: unidades de tamaños muy distintos tienen dispersiones muy distintas.

5.3 Consecuencias

Propiedad	Con heteroscedasticidad
Insesgadez de $\hat{\boldsymbol{\beta}}$	**se mantiene**
Consistencia	**se mantiene**
Eficiencia	**se pierde**: ya no es de mínima varianza
$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$	**es incorrecta**
Errores estándar, t, F e intervalos	**no válidos**

La primera fila es la que hay que retener: **la estimación no está sesgada; lo que está mal es la medida de su precisión**. Los coeficientes son correctos en media, y todo lo que se dice sobre su significación no vale.

La expresión correcta de la varianza es

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\,\sigma^2\boldsymbol{\Omega}\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

y no hay ninguna razón para que se parezca a la que el programa imprime por defecto. El sesgo puede ir en cualquier dirección: los errores estándar habituales pueden salir demasiado pequeños —y entonces se declara significativo lo que no lo es— o demasiado grandes.

5.4 Procedimientos de detección

5.4.1 Análisis gráfico

Representar e_i^2 o $|e_i|$ frente a $\hat{y}_i$ y frente a cada regresor. Un abanico, un embudo o una curva son indicios; una banda de anchura constante, no.

Es informal y es el primer paso siempre, porque además sugiere **qué forma** tiene la heteroscedasticidad, que es lo que hace falta para corregirla.

5.4.2 Contraste de Goldfeld-Quandt

Supone que la varianza es monótona respecto de una variable Z, normalmente uno de los regresores.

Paso	Qué se hace
1	ordenar las observaciones según Z
2	eliminar las c centrales, en torno a $n/4$
3	estimar el modelo en cada uno de los dos grupos
4	calcular $F = \mathrm{SCR}_2/\mathrm{SCR}_1$ con $\mathrm{SCR}_2 > \mathrm{SCR}_1$

$$F \sim F_{(n-c)/2-k,\ (n-c)/2-k}$$

Eliminar el centro es lo que da potencia al contraste: acentúa la diferencia entre los dos extremos. Su limitación es que **exige saber respecto de qué variable crece la varianza**, y que solo detecta heteroscedasticidad monótona.

5.4.3 Contraste de Breusch-Pagan

Más general: contrasta si la varianza depende linealmente de un conjunto de variables.

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \cdots + \alpha_p Z_{pi}$$

Paso	Qué se hace
1	estimar el modelo original y obtener e_i
2	regresar e_i^2 sobre las Z
3	calcular $LM = nR^2$ de esa regresión auxiliar
4	comparar con χ^2_{p-1}

$$H_0 : \alpha_2 = \cdots = \alpha_p = 0 \quad \text{(homoscedasticidad)}$$

Es un contraste **asintótico**: necesita muestras grandes. Y en su versión original es **sensible a la no normalidad** de la perturbación, lo que puede hacerle rechazar por un motivo distinto del que dice; la versión robusta de Koenker corrige eso.

El contraste de White es de la misma familia y añade a la regresión auxiliar los cuadrados y los productos cruzados de los regresores, con lo que detecta también formas no lineales. Su precio son muchos grados de libertad, y por eso pierde potencia en modelos con muchas variables.

5.4.4 Contraste de Glejser

Regresa $|e_i|$ sobre alguna función de Z:

$$|e_i| = \alpha_1 + \alpha_2 f(Z_i) + v_i$$

con $f(Z) = Z$, $\sqrt{Z}$, $1/Z$ o $1/\sqrt{Z}$. Se contrasta $\alpha_2 = 0$.

Su interés está en que **la forma funcional que resulte significativa indica cómo corregir**: si lo que ajusta es $f(Z) = Z$, la desviación típica es proporcional a Z, y esa es exactamente la ponderación que hay que usar. Los otros contrastes dicen que hay problema y este dice de qué tipo.

Su inconveniente conocido es que el término de error de esa regresión auxiliar no cumple las hipótesis habituales, así que su validez es asintótica.

5.5 Estimación de modelos con heteroscedasticidad

5.5.1 Mínimos cuadrados generalizados

Si $\mathbf{\Omega}$ se conoce:

$$\hat{\boldsymbol{\beta}}_{\text{MCG}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}$$

Este estimador **sí es óptimo** bajo heteroscedasticidad: es el teorema de Gauss-Markov generalizado, o teorema de Aitken.

5.5.2 Mínimos cuadrados ponderados

Es lo mismo, visto como una transformación. Si $\text{Var}(u_i) = \sigma^2 Z_i^2$, se divide toda la ecuación por Z_i:

$$\frac{Y_i}{Z_i} = \beta_1 \frac{1}{Z_i} + \beta_2 \frac{X_i}{Z_i} + \frac{u_i}{Z_i}$$

y la nueva perturbación tiene varianza constante:

$$\text{Var}\left(\frac{u_i}{Z_i}\right) = \frac{\sigma^2 Z_i^2}{Z_i^2} = \sigma^2$$

Sobre el modelo transformado se aplica mínimos cuadrados ordinarios. En términos de peso, cada observación pesa $1/Z_i^2$: **las observaciones más precisas pesan más**, que es la idea de fondo.

Dos cuidados al transformar:

- **El término independiente desaparece como tal**: se convierte en el coeficiente de $1/Z_i$, y la nueva ecuación no tiene término independiente propio.
- **El R^2 del modelo transformado no es comparable** con el del original, porque la variable dependiente ha cambiado.

Si $\mathbf{\Omega}$ no se conoce —el caso normal—, se estima primero, con la forma que sugieran Glejser o el gráfico, y se aplica **mínimos cuadrados generalizados factibles**. Su validez es asintótica, y depende de haber acertado con la forma: **una ponderación equivocada puede empeorar el resultado** respecto de no ponderar.

5.5.3 Errores estándar robustos de White

La alternativa que evita tener que acertar con la forma:

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i e_i^2\,\mathbf{x}_i\mathbf{x}_i'\right)(\mathbf{X}'\mathbf{X})^{-1}$$

Se estima por mínimos cuadrados ordinarios y se corrigen **solo** los errores estándar. Los coeficientes no cambian, y no hace falta especificar $\boldsymbol{\Omega}$.

	Mínimos cuadrados ponderados	Errores robustos
Coeficientes	cambian y son más eficientes	no cambian
Hay que conocer la forma de $\boldsymbol{\Omega}$	sí	no
Validez	asintótica si se estima $\boldsymbol{\Omega}$	asintótica
Riesgo	acertar mal la forma empeora	ninguno equivalente

En la práctica, **los errores robustos son la opción por defecto** cuando el objetivo es la inferencia y la eficiencia no es crítica: son más fáciles de justificar y no dependen de una suposición que puede fallar.

5.6 Un procedimiento de trabajo

1. Mirar el gráfico de residuos frente a $\hat{y}$ y frente a cada regresor.
2. **Comprobar la especificación antes que nada.** Si falta una variable o la forma funcional es errónea, corregir eso primero.
3. Aplicar Breusch-Pagan o White como contraste general.
4. Si rechaza, usar Goldfeld-Quandt o Glejser para averiguar **respecto de qué** variable y con qué forma.
5. Corregir: ponderar si la forma está clara, errores robustos si no.
6. Informar de la corrección aplicada y de por qué.

El paso 2 es el que más se salta y el que más resultados salva: corregir con ponderaciones un problema que era de especificación deja el modelo igual de mal especificado y con la apariencia de haberlo arreglado.

El tratamiento de la heteroscedasticidad y de sus contrastes sigue a Gujarati, 2010 y Wooldridge, 2010; los mínimos cuadrados generalizados y los errores robustos, a Greene, 1999 y Johnston y DiNardo, 2001; la práctica con programas estadísticos, a Fernández-Sánchez et al., 2016.

Capítulo 6

Autocorrelación

Tema 6 del programa. Cuando las perturbaciones están correlacionadas entre sí: concepto, causas, consecuencias, contrastes y estimación.

6.1 Concepto

H5 exige que $\text{Cov}(u_i, u_j) = 0$ para $i \neq j$. Hay **autocorrelación** cuando esa covarianza no es cero, es decir, cuando la perturbación de un período guarda relación con la de otro.

$$\text{E}[\mathbf{u}\mathbf{u}'] = \sigma^2\boldsymbol{\Omega}, \qquad \boldsymbol{\Omega} \neq \mathbf{I}, \text{ con elementos no nulos fuera de la diagonal}$$

Es la otra forma de romper $\text{E}[\mathbf{u}\mathbf{u}'] = \sigma^2\mathbf{I}$. El tema 5 cambiaba la diagonal; este llena lo de fuera.

Aparece casi siempre en **series temporales**, por la razón del tema 1: lo que ocurre en un período arrastra al siguiente.

6.1.1 El esquema autorregresivo de primer orden

El caso habitual:

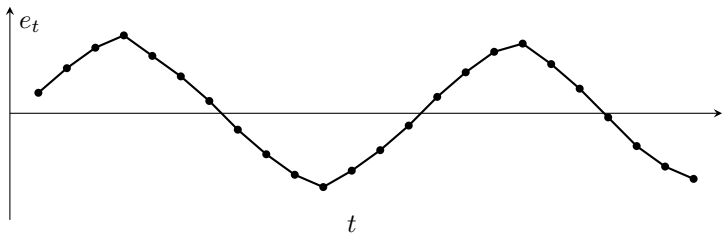
$$u_t = \rho\, u_{t-1} + \varepsilon_t, \qquad |\rho| < 1$$

con ε_t ruido blanco: media cero, varianza constante y sin autocorrelación. De ahí:

$$\text{Var}(u_t) = \frac{\sigma_\varepsilon^2}{1-\rho^2}, \qquad \text{Cov}(u_t, u_{t-s}) = \rho^s \ \text{Var}(u_t)$$

La correlación decae geométricamente con la distancia entre períodos. La condición $|\rho| < 1$ es la de estacionariedad: sin ella la varianza no existe y el proceso no es estable.

Signo de ρ	Qué se ve en los residuos
$\rho > 0$	rachas: varios residuos positivos seguidos y luego varios negativos
$\rho < 0$	alternancia de signo casi observación a observación
$\rho = 0$	sin patrón



6.2 Causas

Causa	Qué ocurre
Inercia económica	las magnitudes económicas tienen ciclos y persistencia
Variable relevante omitida	si la omitida está autocorrelacionada, la perturbación lo hereda
Forma funcional incorrecta	ajustar una recta a una relación curva deja residuos con patrón
Retardos no incluidos	el efecto de X se reparte en varios períodos
Manipulación de los datos	interpolar, suavizar o agregar introduce correlación
Fenómenos de telaraña	la oferta responde con retardo al precio

Las dos del medio son **errores de especificación disfrazados de autocorrelación**, igual que en el tema 5. Es la primera comprobación: si al añadir la variable que falta o al cambiar la forma funcional el patrón desaparece, no había autocorrelación.

6.3 Consecuencias

Son las mismas que en el tema 5, y por la misma razón: lo que falla no es H3 sino la matriz de covarianzas.

Propiedad	Con autocorrelación
Insesgadez	**se mantiene**
Consistencia	**se mantiene**, salvo con endógena retardada
Eficiencia	**se pierde**
Var($\hat{\boldsymbol{\beta}}$) habitual	**incorrecta**
Inferencia	**no válida**

Con $\rho > 0$, que es el caso frecuente, **los errores estándar salen demasiado pequeños**. La dirección del sesgo importa: infla los t, infla el R^2 y hace parecer significativo lo que no lo es. Es más peligroso que el caso contrario, porque el resultado tiene mejor aspecto del que merece.

La salvedad de la segunda fila es importante: si el modelo incluye la variable endógena retardada, Y_{t-1} está correlacionada con u_{t-1} y por tanto con u_t, así que el estimador **deja de ser consistente**. Ahí la autocorrelación no es solo un problema de eficiencia.

6.3.1 Regresión espuria

En series temporales con tendencia aparece un fenómeno específico: dos variables sin ninguna relación pueden dar un R^2 altísimo y coeficientes muy significativos solo porque las dos crecen con el tiempo.

La señal de alarma es un R^2 **mayor que el estadístico de Durbin-Watson**. Esa combinación —ajuste excelente y autocorrelación severa— indica regresión espuria y no un buen modelo.

Se corrige trabajando en diferencias, o comprobando que las variables están cointegradas, es decir, que existe una combinación lineal suya que sí es estacionaria.

6.4 Procedimientos de detección

6.4.1 Gráfico de residuos

Representar e_t frente a t y frente a e_{t-1}. Rachas o alternancia son el indicio, y el segundo gráfico da además el signo de ρ: una nube con pendiente positiva es $\rho > 0$.

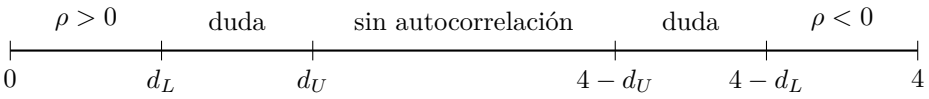
6.4.2 Contraste de Durbin-Watson

El clásico para autocorrelación de primer orden:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \approx 2(1-\hat{\rho})$$

$\hat{\rho}$	d
1	0
0	2
-1	4

La lectura se hace con dos valores críticos, d_L y d_U, y deja **dos zonas sin decisión**:



Sus **condiciones de uso** son estrictas, y se incumplen a menudo:

- el modelo tiene que llevar término independiente;
- los regresores tienen que ser no estocásticos;
- **no puede haber variable endógena retardada** entre los regresores;
- solo detecta autocorrelación de orden 1.

Con endógena retardada, d tiende a 2 aunque haya autocorrelación, así que el contraste dice que no hay problema justo cuando lo hay. Ahí se usa la **h de Durbin**:

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\,\widehat{\mathrm{Var}}(\hat{\beta}_{Y_{t-1}})}} \sim N(0,1)$$

y solo se puede calcular si $n\,\widehat{\mathrm{Var}}(\hat{\beta}_{Y_{t-1}}) < 1$; si no, el radicando es negativo y el estadístico no existe.

6.4.3 Contraste de Breusch-Godfrey

El general, y el que resuelve las limitaciones del anterior.

Paso	Qué se hace
1	estimar el modelo y obtener e_t
2	regresar e_t sobre los regresores originales **y** sobre $e_{t-1}, \ldots, e_{t-p}$
3	calcular $LM = (n-p)R^2$ de esa regresión
4	comparar con χ^2_p

$$H_0 : \rho_1 = \cdots = \rho_p = 0$$

Sus ventajas sobre Durbin-Watson: detecta órdenes superiores al primero, **admite la endógena retardada** y no tiene zona de duda. Su precio es que es asintótico y que hay que elegir p.

6.4.4 Contraste de Ljung-Box

Contrasta conjuntamente que las primeras m autocorrelaciones son nulas:

$$Q = n(n+2)\sum_{s=1}^{m} \frac{\hat{\rho}_s^2}{n-s} \;\sim\; \chi^2_m$$

Se aplica sobre los residuos, y en ese caso los grados de libertad se reducen en el número de parámetros estimados del proceso. Es el contraste habitual para comprobar que un modelo de series temporales ha dejado residuos de ruido blanco.

6.5 Estimación de modelos con perturbaciones autocorrelacionadas

6.5.1 La transformación cuasidiferencial

Si ρ se conociera, restando a la ecuación en t la ecuación en $t-1$ multiplicada por ρ:

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$$

y la nueva perturbación ε_t ya no está autocorrelacionada. Sobre esa ecuación se aplica mínimos cuadrados ordinarios.

Se pierde la primera observación, y con muestras cortas eso es caro. La **transformación de Prais-Winsten** la recupera multiplicándola por $\sqrt{1-\rho^2}$, en vez de descartarla como hace Cochrane-Orcutt.

6.5.2 Estimación de ρ

Como no se conoce, hay que estimarlo:

Método	Cómo
A partir de d	$\hat{\rho} \approx 1 - d/2$
Regresión de residuos	regresar e_t sobre e_{t-1}
Cochrane-Orcutt	iterar: estimar ρ, transformar, reestimar, repetir hasta converger
Hildreth-Lu	barrer ρ en una rejilla y quedarse con el que minimice la suma de residuos al cuadrado
Máxima verosimilitud	estimar $\boldsymbol{\beta}$ y ρ a la vez

Cochrane-Orcutt puede converger a un mínimo local; Hildreth-Lu no, porque explora toda la rejilla, a cambio de más cálculo.

6.5.3 Errores estándar robustos de Newey-West

Como los de White del tema 5, pero robustos además a la autocorrelación:

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\hat{\mathbf{S}}(\mathbf{X}'\mathbf{X})^{-1}$$

con $\hat{\mathbf{S}}$ construida con las autocovarianzas de los residuos hasta un retardo máximo. Los coeficientes no cambian; solo se corrige la inferencia. Es la opción por defecto cuando no se quiere modelar la estructura de la autocorrelación.

6.5.4 Añadir dinámica al modelo

La alternativa de fondo, y a menudo la mejor: si la autocorrelación viene de que el modelo es estático y el fenómeno no lo es, la corrección no es transformar sino **especificar bien**. Incluir Y_{t-1}, o retardos de X, puede hacer desaparecer la autocorrelación porque elimina su causa.

Es el mismo criterio del tema 5: **primero la especificación, después la corrección**.

6.6 Los tres problemas, juntos

	Multicolinealidad	Heteroscedasticidad	Autocorrelación
Hipótesis que rompe	H2	H4	H5
Dónde aparece	cualquier muestra	sección cruzada	series temporales
Insesgadez	se mantiene	se mantiene	se mantiene
Eficiencia	se mantiene	se pierde	se pierde
Varianza estimada	correcta, y grande	incorrecta	incorrecta
Detección	FIV, número de condición	Breusch-Pagan, White, Goldfeld-Quandt, Glejser	Durbin-Watson, h, Breusch-Godfrey, Ljung-Box
Corrección	más datos, transformar, contraer	ponderar, errores robustos	cuasidiferenciar, Newey-West, dinámica

La fila que más se olvida es la quinta. **En multicolinealidad la varianza estimada es correcta**: el problema es que es grande, y los resultados avisan solos con sus t pequeñas. En los otros dos la

varianza estimada es **falsa**, y ahí los resultados no avisan de nada. Por eso los temas 5 y 6 exigen contrastar siempre, y el 4 admite no hacer nada.

El tratamiento de la autocorrelación y de sus contrastes sigue a Gujarati, 2010, Novales, 2000 y Wooldridge, 2010; los métodos de estimación y los errores robustos, a Greene, 1999 y Johnston y DiNardo, 2001; la regresión espuria y la cointegración, a Stock y Watson, 2012 y Pindyck y Rubinfeld, 2001. Los ejercicios de aplicación están en García et al., 2017 y Alonso et al., 2005.

Bibliografía

Alonso, A., Fernández, J., & Gallastegui, I. (2005). *Econometría*. Prentice Hall.

Fernández-Sánchez, P., Salmerón-Gómez, R., & Blanco, V. (2016). *Prácticas de Econometría con Excel, Gretl y RStudio*. Fleming.

García, R. M., Herrerías, J. M., & Palacios, F. (2017). *Econometría. Ejercicios resueltos*. Pirámide.

Greene, W. H. (1999). *Análisis econométrico*. Prentice Hall.

Gujarati, D. N. (2010). *Econometría*. McGraw Hill.

Johnston, J., & DiNardo, J. (2001). *Métodos de econometría*. Vicens-Vives.

Maddala, G. S. (2001). *Introduction to Econometrics* (3.ª ed.). John Wiley & Sons.

Novales, A. (2000). *Econometría* (2.ª ed.). McGraw Hill.

Pindyck, R. S., & Rubinfeld, D. L. (2001). *Econometría, modelos y pronósticos*. McGraw Hill.

Stock, J. H., & Watson, M. M. (2012). *Introducción a la Econometría* (3.ª ed.). Pearson.

Wooldridge, J. M. (2010). *Introducción a la Econometría. Un enfoque moderno* (2.ª ed.). Thomson.