

Incumplimiento de las hipótesis básicas en el término de error

Tomás del Barrio Castro
Miquel Clar López
Jordi Suriñach Caralt

PID_00160619



Universitat Oberta
de Catalunya

www.uoc.edu

Índice

Introducción	5
Objetivos	6
1. Planteamiento general	7
1.1. Matrices de varianzas y covarianzas escalares y no escalares del término de perturbación.....	7
1.1.1. Caso general del modelo de regresión lineal múltiple generalizado (MRLMG): perturbaciones no esféricas (heteroscedasticidad y autocorrelación)	8
1.1.2. Caso particular del MRLMG: perturbaciones esféricas (homoscedasticidad y no autocorrelación)	8
1.1.3. Heteroscedasticidad (pero no autocorrelación): matriz de varianzas y covarianzas diagonal (pero no escalar)	9
1.1.4. Autocorrelación (pero no heteroscedasticidad): matriz de varianzas y covarianzas con elementos no nulos en el exterior de la diagonal principal	10
1.2. Estimación por mínimos cuadrados ordinarios (MCO) de un modelo con perturbaciones no esféricas.....	11
1.2.1. Propiedades de los estimadores $\hat{\mathbf{B}}$ cuando el término de perturbación es no esférico.....	11
1.2.2. Propiedades del estimador $\hat{\sigma}_u^2$ cuando las perturbaciones son no esféricas	13
1.3. Estimación por mínimos cuadrados generalizados (MCG): una doble aproximación.....	14
1.3.1. Primera aproximación a la estimación por MCG: mínimos cuadrados ponderados	14
1.3.2. Segunda aproximación a la estimación por MCG	17
1.4. Propiedades de los estimadores MCG	19
1.4.1. Propiedades de los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$	19
1.4.2. Propiedades del estimador $\hat{\sigma}_{\text{MCG}}^2$	21
1.5. El estimador máximo verosímil (MV) en el MRLMG	21
1.5.1. Propiedades de los estimadores MV	23
1.6. Normalidad del término de perturbación y contraste de normalidad de Jarque-Bera	24
2. Heteroscedasticidad	28
2.1. Definición y causas	28
2.2. Consecuencias de la estimación por MCO.....	33
2.3. Esquemas de dependencia funcional de la varianza	34
2.4. Detección de heteroscedasticidad.....	37

2.4.1. Métodos gráficos	37
2.4.2. Contrastes de heteroscedasticidad.....	39
2.5. Estimación por mínimos cuadrados generalizados (MCG y MCP)	48
3. Autocorrelación	59
3.1. Definición y causas	59
3.1.1. Definición de autocorrelación.....	59
3.1.2. Causas que pueden dar lugar a la presencia de autocorrelación en el término de perturbación	63
3.2. Esquemas de autocorrelación en el término de perturbación	66
3.2.1. Esquema de autocorrelación autorregresivo (AR): el esquema AR(1)	67
3.3. Detección de la autocorrelación	71
3.3.1. Análisis gráfico de los residuos.....	71
3.3.2. Contrastes de autocorrelación.....	73
3.4. Estimación por MCG	84
3.5. Estimación por los métodos Cochrane-Orcutt y Durbin	86
Glosario	91
Bibliografía	93

Introducción

Este módulo didáctico está formado por tres apartados en los que abordamos el incumplimiento de las hipótesis básicas referidas al término de perturbación. A continuación, presentamos de forma más concreta los aspectos asociados a cada apartado:

1) En el primer apartado realizamos un análisis general sobre lo que ocurre con los estimadores de mínimos cuadrados ordinarios cuando el término de perturbación es no esférico (pero sin entrar en detalle con respecto al hecho de si la no esfericidad es una consecuencia de la existencia de heteroscedasticidad o de autocorrelación en el término de perturbación). Así pues, mostramos que los estimadores de mínimos cuadrados ordinarios dejan de ser eficientes y que, para alcanzar la eficiencia, es necesario utilizar otro método de estimación: el **método de mínimos cuadrados generalizados** (MCG). Para finalizar este primer apartado, estudiamos un contraste adecuado de la **hipótesis de normalidad del término de perturbación** y las consecuencias de que ésta se incumpla.

2) A continuación, en el segundo apartado, nos centramos en el **problema de la heteroscedasticidad**. De modo que, una vez definidas y estudiadas las causas que pueden dar lugar a la presencia de ésta, presentamos distintos esquemas funcionales que puede seguir la varianza del término de perturbación. Acto seguido, abordamos la cuestión de cómo podemos detectar si un término de perturbación es heteroscedástico, y presentamos los métodos gráficos y algunos de los contrastes que hay. En último lugar, analizamos la estimación del modelo con perturbaciones heteroscedásticas.

3) En el último apartado de este módulo analizamos el **problema de la autocorrelación**: estudiamos las causas que pueden generar autocorrelación en el término de perturbación, mostramos un tipo de proceso de autocorrelación (el autorregresivo de orden 1), y analizamos los métodos gráficos y variados contrastes para detectar la existencia de autocorrelación. Y ya para acabar, abordamos la estimación del modelo cuando el término de perturbación sigue un proceso autorregresivo de orden 1 mediante diferentes métodos. En concreto, presentamos tres métodos: el de mínimos cuadrados generalizados, y los métodos propuestos por Cochrane-Orcutt y Durbin.

Objetivos

Tras haber trabajado los contenidos de este módulo didáctico, el estudiante tiene que haber alcanzado los siguientes objetivos:


- 1.** Conocer las diferentes formas que puede adoptar la matriz de varianzas y covarianzas del término de perturbación cuando éste es no esférico.
- 2.** Conocer las propiedades que cumplen y las que no cumplen los estimadores de mínimos cuadrados ordinarios cuando el término de perturbación es no esférico, y el porqué.
- 3.** Obtener los estimadores eficientes cuando el término de perturbación es no esférico, ya sea por problemas de heteroscedasticidad, ya de autocorrelación.
- 4.** Saber detectar (gráficamente y mediante contrastes) la existencia de heteroscedasticidad y autocorrelación.
- 5.** Conocer los rasgos característicos básicos de un proceso autorregresivo de orden 1.
- 6.** Averiguar si el término de perturbación cumple la hipótesis de normalidad.

1. Planteamiento general

El modelo de regresión $Y = XB + 1$, llamado **modelo de regresión lineal múltiple estándar**, por lo que respecta al término de perturbación, se supone que cumple las siguientes hipótesis:


- a) Su valor esperado es cero: $E[u_i] = 0 \quad \forall i = 1, \dots, N$.
- b) Presenta homoscedasticidad: todos los términos de perturbación tienen la misma varianza.
- c) No presenta autocorrelación: las covarianzas, es decir, los elementos del exterior de la diagonal principal de la matriz de varianzas y covarianzas, son iguales a cero.
- d) Se distribuye de acuerdo con una ley normal.

En este apartado, en primer lugar hallaremos reflejada la hipótesis de esfericidad del término de perturbación (pero sin entrar en detalle en cuanto a si la causa es la presencia de heteroscedasticidad y/o de autocorrelación), aunque mantendremos el resto.

En los dos apartados posteriores trataremos por separado las diferentes fuentes de no esfericidad: heteroscedasticidad y autocorrelación. 


De hecho, la situación más general que suele darse en la práctica es que las perturbaciones del modelo sean no esféricas, lo cual implica matrices de varianzas y covarianzas no escalares.

Así, al nuevo modelo que estudiamos en este apartado lo denominaremos **modelo de regresión lineal múltiple generalizado** (MRLMG), denominación en la que aparece incorporado el adjetivo *generalizado* para registrar el hecho de que es el caso más general posible con el que podemos encontrarnos.

A continuación, en el subapartado 1.6 trataremos de la hipótesis de normalidad del término de perturbación. 

1.1. Matrices de varianzas y covarianzas escalares y no escalares del término de perturbación

Antes de abordar lo que es el cuerpo central de este apartado, introducimos el subapartado siguiente para familiarizarnos con la notación matricial de los diferentes tipos de matrices de varianzas y covarianzas de los términos de perturbación con las que podemos encontrarnos, suponiendo que se mantiene la hipótesis $E[U] = 0_N$.

 Ved el modelo de regresión lineal múltiple estándar en el módulo "Modelo de regresión lineal múltiple..."

EL MRLM con perturbaciones esféricas

... no es más que un caso particular del MRLMG, como podremos comprobar a lo largo de este apartado.

1.1.1. Caso general del modelo de regresión lineal múltiple generalizado (MRLMG): perturbaciones no esféricas (heteroscedasticidad y autocorrelación)

Por definición, la matriz de varianzas y covarianzas de los términos de perturbación responde a la expresión que vemos a continuación:

$$\begin{aligned} \text{VAR}[\mathbf{U}] &= E\{[\mathbf{U} - E[\mathbf{U}]][\mathbf{U} - E[\mathbf{U}]]'\} = E[\mathbf{U}\mathbf{U}'] = \sigma^2 \Omega_N, \text{ on } \Omega_N \neq \mathbf{I}_N \Rightarrow \\ &\Rightarrow \begin{cases} E[u_i^2] = \sigma_i^2 & \forall i = 1, \dots, N; \\ E[u_i u_j] \neq 0 & \forall i \neq j. \end{cases} \end{aligned}$$

En esta expresión, N es el número de observaciones de las que disponemos y hemos utilizado la hipótesis $E[\mathbf{U}] = 0$.

Desarrollando la matriz $\text{VAR}[\mathbf{U}]$, tenemos la expresión que vemos a continuación:

$$\begin{aligned} \text{VAR}[\mathbf{U}] = E[\mathbf{U}\mathbf{U}'] &= \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \dots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \dots & \sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma_N^2 \end{bmatrix} = \\ &= \sigma^2 \begin{bmatrix} \delta_1 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1N} \\ \gamma_{21} & \delta_2 & \gamma_{23} & \dots & \gamma_{2N} \\ \gamma_{31} & \gamma_{32} & \delta_3 & \dots & \gamma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \gamma_{N3} & \dots & \delta_N \end{bmatrix} = \sigma^2 \Omega_N. \end{aligned}$$

Como podéis apreciar, $\sigma^2 \dots$

... no es más que un factor de escala común a todas las varianzas de los diferentes términos de perturbación, lo cual hace que la matriz Ω_N no sea única.

1.1.2. Caso particular del MRLMG: perturbaciones esféricas (homoscedasticidad y no autocorrelación)

Según este supuesto, la matriz de varianzas y covarianzas de los términos de perturbación es una matriz escalar, es decir, una matriz diagonal con constantes en la diagonal que son la misma, y que denotamos como σ_u^2 :

$$\text{VAR}[\mathbf{U}] = E[\mathbf{U}\mathbf{U}'] = \begin{bmatrix} \sigma^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \sigma^2 \mathbf{I}_N.$$

Entonces tenemos lo siguiente:

$$\Omega_N = \mathbf{I}_N \Rightarrow \mathbf{U}_N \sim \mathbf{N}_N[0, \sigma^2 \mathbf{I}_N],$$

o bien, escrito de otra manera:

$$\begin{aligned} \text{VAR}[\mathbf{U}] &= E\{[\mathbf{U} - E[\mathbf{U}]][\mathbf{U} - E[\mathbf{U}]]'\} = E[\mathbf{U}\mathbf{U}'] = \sigma^2 \mathbf{I}_N \Rightarrow \\ &\Rightarrow \begin{cases} E[u_i^2] = \sigma^2 & \forall i = 1, \dots, N; \\ E[u_i u_j] = 0 & \forall i \neq j. \end{cases} \end{aligned}$$

Como podemos ver, un término de perturbación y no autocorrelacionado, es decir, esférico, tiene asociada una matriz de varianzas y covarianzas escalar, que podemos descomponer en el producto de una constante, σ^2 , por la matriz identidad.

1.1.3. Heteroscedasticidad (pero no autocorrelación): matriz de varianzas y covarianzas diagonal (pero no escalar)

Heteroscedasticidad significa que la varianza de los términos de perturbación no es constante (no es la misma) para todos, es decir, que como mínimo un término de perturbación presenta una varianza diferente de la que presentan el resto de los términos.

La heteroscedasticidad...

... suele darse cuando trabajamos con datos de corte transversal, pero no tenemos que caer en la trampa de considerar que, cuando trabajamos con datos de serie temporal, no puede darse heteroscedasticidad en el término de perturbación.

En consecuencia, Ω_N es, en este caso, una matriz diagonal (todos los elementos del exterior de la diagonal principal son cero) del siguiente tipo:

$$\begin{aligned} \text{VAR}[\mathbf{U}] &= E\{[\mathbf{U} - E[\mathbf{U}]][\mathbf{U} - E[\mathbf{U}]]'\} = E[\mathbf{U}\mathbf{U}'] = \sigma^2 \Omega_N \Rightarrow \\ &\Rightarrow \begin{cases} E[u_i^2] = \sigma_i^2 & \forall i = 1, \dots, N; \\ E[u_i u_j] = 0 & \forall i \neq j. \end{cases} \end{aligned}$$

Es decir:

$$\text{VAR}[\mathbf{U}] = E[\mathbf{U}\mathbf{U}'] = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma_N^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} \delta_1 & 0 & 0 & \dots & 0 \\ 0 & \delta_2 & 0 & \dots & 0 \\ 0 & 0 & \delta_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \delta_N \end{bmatrix} = \sigma^2 \Omega_N.$$

Como podemos

... de la misma manera que ocurría en el subapartado 1.1.1, σ^2 es un factor de escala común a todas las varianzas de los diferentes términos de perturbación y, por lo tanto, la forma de definir la matriz Ω_N no es única.

1.1.4. Autocorrelación (pero no heteroscedasticidad): matriz de varianzas y covarianzas con elementos no nulos en el exterior de la diagonal principal

Autocorrelación significa que hay correlaciones no nulas entre los términos de perturbación correspondientes a diferentes observaciones (en general, diferentes momentos del tiempo), es decir, que como mínimo un elemento (en realidad dos, dado que se trata de una matriz simétrica) del exterior de la diagonal principal de la matriz de varianzas y covarianzas es diferente de cero.

La autocorrelación...

... suele darse cuando trabajamos con datos de serie temporal, aunque también podemos encontrarnos con la presencia de autocorrelación cuando trabajamos con datos de corte transversal.

Según lo que acabamos de decir, si hay autocorrelación, la matriz de varianzas y covarianzas del término de perturbación presenta el aspecto que mostramos a continuación:


$$\begin{aligned} \text{VAR}[U] &= E\{[U - E[U]][U - E[U]]'\} = E[UU'] = \sigma^2 \Omega_N \Rightarrow \\ &\Rightarrow \begin{cases} E[u_i^2] = \sigma^2 & \forall i = 1, \dots, N; \\ E[u_i u_j] \neq 0 & \text{como mínimo para algún } i \neq j. \end{cases} \end{aligned}$$

Así pues, según este supuesto, Ω_N es una matriz no diagonal:

$$\text{VAR}[U] = E[UU'] = \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma^2 & \sigma_{23} & \dots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \sigma^2 & \dots & \sigma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \gamma_{12} & \gamma_{13} & \dots & \gamma_{1N} \\ \gamma_{21} & 1 & \gamma_{23} & \dots & \gamma_{2N} \\ \gamma_{31} & \gamma_{32} & 1 & \dots & \gamma_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \gamma_{N3} & \dots & 1 \end{bmatrix} = \sigma^2 \Omega_N.$$

Como podemos ver en la expresión anterior, teniendo en cuenta que se supone que se cumple la hipótesis básica de homoscedasticidad, es posible obtener el término σ^2 multiplicando fuera de la matriz de varianzas y covarianzas, puesto que es constante para todos los términos de perturbación.

Como ya hemos mencionado antes, en este apartado estudiaremos el MRLM en el contexto de perturbaciones no esféricas en general, es decir, trataremos del supuesto planteado en el subapartado 1.1.1, y dejaremos para los dos apartados siguientes el análisis de los casos 1.1.3 y 1.1.4, respectivamente.

En concreto, se trata de estudiar qué sucede con las propiedades de los estimadores MCO. Tenemos que ver, entonces, si en el supuesto de no esfericidad de los términos de perturbación el método de estimación mencionado sigue proporcionando estimadores que cumplen todas las propiedades deseables, y, en caso de que no sea así, tendremos que analizar qué propiedades se incumplen y encontrar otro método de estimación más adecuado en términos de propiedades de los estimadores. 

1.2. Estimación por mínimos cuadrados ordinarios (MCO) de un modelo con perturbaciones no esféricas


Previo paso a la determinación de las propiedades de los estimadores MCO en este nuevo contexto, puede ser útil recordar la procedencia de este estimador. Así pues, obteníamos el estimador MCO a partir de la minimización de la suma de los residuos al cuadrado $(\sum_{i=1}^N e_i^2)y$, a partir de la resolución de las ecuaciones normales, llegábamos a la expresión del siguiente estimador:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (1.1)$$

siendo su matriz de varianzas y covarianzas:

$$\text{VAR}[\hat{\mathbf{B}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (1.2)$$

Como ya hemos estudiado, en el supuesto de que se cumplan todas las hipótesis básicas del modelo de regresión, los estimadores MCO de los parámetros \mathbf{B} son no sesgados, eficientes (de varianza mínima, es decir, óptimos) y consistentes en términos del error cuadrático medio (ECM); y, en lo que concierne a la estimación de la varianza del término de perturbación, $\hat{\sigma}_u^2$, es no sesgada.

El objetivo de este subapartado es el de comprobar si los estimadores MCO mantienen las propiedades anteriores cuando se incumple la hipótesis de esfericidad de los términos de perturbación, es decir, cuando su matriz de varianzas y covarianzas deja de ser escalar. 

1.2.1. Propiedades de los estimadores $\hat{\mathbf{B}}$ cuando el término de perturbación es no esférico

De acuerdo con lo que hemos visto, podemos representar el contexto de perturbaciones no esféricas de la siguiente forma:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{U};$$


$$\mathbf{U}_N \sim \mathbf{N}_N[\mathbf{0}, \sigma^2\Omega_N], \text{ donde } \Omega_N \neq \mathbf{I}_N.$$


Por otra parte, como ya sabréis, partiendo de la expresión 1.1, podemos escribir el estimador MCO como se sigue:

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U} = \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}, \quad (1.3)$$

lo cual permite escribir:

$$\hat{\mathbf{B}} - \mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}. \quad (1.4)$$


Partiendo de las expresiones anteriores, podemos comprobar si los estimadores $\hat{\mathbf{B}}$ cumplen las propiedades deseables. A continuación vamos a considerar cada una de estas propiedades: 

 Ved la estimación por el método de mínimos cuadrados ordinarios del modelo de regresión lineal múltiple en el subapartado 2.3 del módulo "Modelo de regresión lineal múltiple..."


Nota: recordad que \mathbf{X}' es la matriz transpuesta de \mathbf{X} .


Diferencias entre σ_u^2 y σ^2

Hasta ahora siempre se ha hablado del parámetro σ_u^2 –que es único– en lugar del parámetro σ^2 –que no lo es–, puesto que se supone que las perturbaciones son esféricas.

 Ved el contexto de perturbaciones no esféricas en el subapartado 1.1.1 de este módulo didáctico.

Nota: de ahora en adelante denotaremos la matriz Ω_N simplemente por Ω .

 Ved cómo llegar a esta expresión en la expresión 2.22 del módulo "Modelo de regresión lineal múltiple..."

 Ved las propiedades de los estimadores MCO de los parámetros \mathbf{B} en el subapartado 2.3.2 del módulo "Modelo de regresión lineal múltiple..."

1) En cuanto a la **propiedad de no sesgo**, podemos demostrar que los estimadores, $\hat{\beta}_j$, $j = 1, \dots, k$ continúan siendo estimadores no sesgados. Para verlo, simplemente tomamos esperanzas en 1.3, con lo que llegamos a la siguiente expresión:

$$\begin{aligned} E[\hat{\mathbf{B}}] &= E[\mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}] = E[\mathbf{B}] + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}] = \\ &= \mathbf{B} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{U}] = \mathbf{B}. \end{aligned} \quad (1.5)$$

2) Por lo que respecta a la **eficacia**, tenemos que encontrar previamente la matriz de varianzas y covarianzas de los estimadores en el contexto de perturbaciones no esféricas:

$$\begin{aligned} \text{VAR}[\hat{\mathbf{B}}] &= E[(\hat{\mathbf{B}} - \mathbf{B})(\hat{\mathbf{B}} - \mathbf{B})'] = E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}\mathbf{U}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' E[\mathbf{U}\mathbf{U}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \end{aligned}$$

y, atendido a que $E[\mathbf{U}\mathbf{U}']$ es la matriz de varianzas y covarianzas de los términos de perturbación, $\sigma^2\Omega$, tenemos lo siguiente:

$$\text{VAR}[\hat{\mathbf{B}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \neq \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \quad (1.6)$$

La expresión 1.6 nos permite extraer dos consecuencias:

a) La expresión $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$ ya no es útil para calcular las varianzas y las covarianzas de los estimadores $\hat{\mathbf{B}}$, dado que conduce a errores y que no es la expresión correcta cuando la matriz de varianzas y covarianzas de los términos de perturbación es no escalar.

b) Los estimadores $\hat{\mathbf{B}}$ son ineficientes: no tienen varianza mínima, tal y como demostraremos más adelante. Este hecho representa un verdadero problema, teniendo en cuenta que, cuando un estimador presenta una varianza mayor que la de otro, la distribución de probabilidad de este estimador en torno a su valor poblacional tiene una forma menos puntiaguda, es decir, acumula menos probabilidad en un intervalo dado. De forma equivalente, a la hora de construir un intervalo de confianza para un nivel de significación (α) dado, éste será más ancho que el que resultaría mediante el uso del estimador de varianza menor. Una consecuencia directa de este hecho es que perdemos precisión al hacer predicciones por intervalo, ya que, para niveles de confianza iguales la anchura de éste, será mayor. Por este motivo, acostumbramos a valorar la propiedad de eficiencia como más deseable para un estimador que el no sesgo.

A partir de los resultados 1.5 y 1.6 podemos escribir que, en este contexto más general de perturbaciones no esféricas, los estimadores MCO se distribuyen según una ley normal con el valor esperado y la matriz de varianzas y covarianzas que hallamos a continuación:

$$\hat{\mathbf{B}}_k \sim N_k[\mathbf{B}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]. \quad (1.7)$$

En la expresión 1.6,...

... es interesante que nos percatemos de que, cuando la matriz de varianzas y covarianzas de los términos de perturbación, Ω_N , coincide con la matriz identidad, \mathbf{I}_N (es decir, es escalar), la expresión de la matriz de varianzas y covarianzas de los estimadores coincide con la que ya conocíamos de la asignatura *Introducción a la econometría*, $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, asociada a un MRLM estándar.

3) Acerca de la **consistencia en ECM**, los estimadores $\hat{\mathbf{B}}$ siguen siendo consistentes:

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{EQM}[\hat{\mathbf{B}}] &= \lim_{N \rightarrow \infty} \{[\text{Sesgo}[\hat{\mathbf{B}}]]^2 + \text{VAR}[\hat{\mathbf{B}}]\} = \lim_{N \rightarrow \infty} \text{VAR}[\hat{\mathbf{B}}] = \\ &= \lim_{N \rightarrow \infty} \left[\frac{\sigma^2}{N} \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \left(\frac{\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}}{N} \right) \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \right] = \mathbf{0}. \end{aligned} \quad (1.8)$$

Para obtener el resultado 1.8...

... suponemos que la matriz

$$\left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1}$$

existe cuando $N \rightarrow \infty$ y que es finita.

1.2.2. Propiedades del estimador $\hat{\sigma}_u^2$ cuando las perturbaciones son no esféricas

Por lo que respecta al estimador de la varianza del término de perturbación, en el caso particular de perturbaciones esféricas, sabemos que la expresión no sesgada para estimar el parámetro mencionado es la siguiente:

$$\hat{\sigma}_u^2 = \frac{SCR}{\text{grados de libertad del modelo}} = \frac{\mathbf{e}'\mathbf{e}}{N - k}.$$

En el caso general de heteroscedasticidad y/o autocorrelación, la expresión anterior es sesgada, atendiendo a que $E[\hat{\sigma}_u^2] \neq \sigma_u^2$.

Demostración del sesgo del estimador $\hat{\sigma}_u^2$ con perturbaciones no esféricas

Partimos de la definición y desarrollamos la expresión como veremos ahora:

$$E[\hat{\sigma}_u^2] = E\left[\frac{\mathbf{e}'\mathbf{e}}{N - k} \right] = \frac{1}{N - k} E[\mathbf{e}'\mathbf{e}].$$

Entonces, calculando el valor esperado del sumatorio de los errores al cuadrado en el supuesto de perturbaciones no esféricas, podemos demostrar que:

$$\begin{aligned} E[\hat{\sigma}_u^2] &= \frac{E[\mathbf{e}'\mathbf{e}]}{N - k} = \frac{1}{N - k} \{ \sigma_u^2 [N - k + \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \boldsymbol{\Omega})\mathbf{X}]] \} = \\ &= \sigma_u^2 + \sigma_u^2 \frac{\text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \boldsymbol{\Omega})\mathbf{X}]}{N - k} \neq \sigma_u^2. \end{aligned}$$

Teniendo en cuenta los resultados obtenidos a lo largo de este subapartado, la conclusión que debemos extraer es que, antes de estimar un modelo de regresión por MCO, debemos ser cuidadosos a la hora de comprobar si los términos de perturbación del modelo son esféricos, dado que, como acabamos de ver, si estimamos por MCO un modelo cuando los términos de perturbación son no esféricos (quizá sin ser conscientes), las estimaciones de los parámetros \mathbf{B} obtenidas, si bien serán no sesgadas y consistentes, presentarán unas varianzas estimadas que no serán óptimas y, por lo tanto, al construir intervalos de confianza por los parámetros de población, la anchura de éstos será mayor que si utilizásemos un estimador de varianza mínima.

Tras haber analizado estos resultados, podemos afirmar que la estimación por MCO de un modelo con perturbaciones no esféricas no es una opción completamente satisfactoria. Así pues, tendremos que encontrar un método de estimación alternativo que proporcione estimadores que cumplan todas las propiedades deseables. Este método, que estudiaremos a continuación, se conoce con el

!
Ved el estimador de la varianza del término de perturbación en el contexto de perturbaciones esféricas en el subapartado 2.4.2 del módulo "Modelo de regresión lineal múltiple...".

Por ejemplo,...

... si especificamos un modelo que tiene términos de perturbación no esféricos para explicar el comportamiento de la demanda de un producto que fabrica una empresa determinada, la predicción por intervalo de la demanda esperada para el año siguiente será menos precisa que si utilizamos un estimador óptimo.

nombre de *método de mínimos cuadrados generalizados*, MCG (conocido, también, como *método de mínimos cuadrados ponderados*, MCP). !

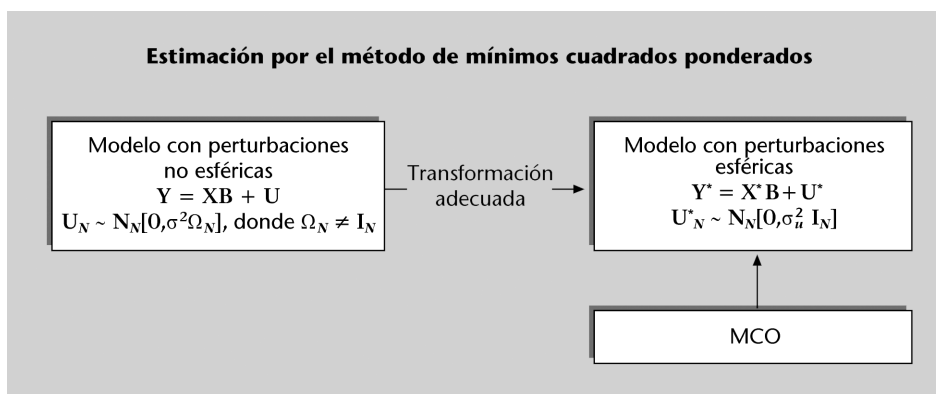
1.3. Estimación por mínimos cuadrados generalizados (MCG): una doble aproximación

A continuación, presentamos dos vías alternativas, pero equivalentes, para llegar a la expresión del estimador MCG. La primera vía consiste en transformar el modelo no esférico en uno esférico, y entonces aplicar MCO en las variables del modelo transformado. En cuanto a la segunda vía, consiste en aplicar la expresión del estimador MCG directamente a los datos originales (es decir, los datos observados). !

1.3.1. Primera aproximación a la estimación por MCG: mínimos cuadrados ponderados

El punto clave de esta aproximación es el siguiente: si estimamos por MCO un modelo que tiene términos de perturbación esféricos, los estimadores $\hat{\mathbf{B}}$ y $\hat{\sigma}_u^2$ como bien sabemos, todas las propiedades deseables. Por lo tanto, si cuando nos hallamos ante un modelo con perturbaciones no esféricas, antes de estimarlo, transformamos las variables de manera que las perturbaciones se convierten en esféricas, podremos aplicar MCO a los nuevos datos y los estimadores cumplirán todas las propiedades deseables. !

El siguiente cuadro muestra la idea anterior mediante un gráfico:



En consecuencia, la aplicación de esta aproximación a la estimación de MCG requiere necesariamente la realización de dos etapas. La primera de ellas consiste en transformar las variables, y la segunda, en aplicar la expresión de MCO a las nuevas variables. A continuación, estudiaremos cada una de estas dos etapas. !

1) Transformación del modelo

Consiste en definir una matriz T de transformación que será la que nos permitirá pasar de un modelo con perturbaciones no esféricas a otro con perturbaciones

esféricas. Esta matriz T tiene que reunir tres requisitos: ser cuadrada*, no singular y todos sus elementos deben ser constantes.

* De dimensión $N \times N$.

Una vez definida la matriz de transformación, se multiplican el vector de observaciones Y y la matriz de observaciones X , y también el vector de los términos de perturbaciones, U , por la matriz T mencionada, y se definen los vectores y la matriz transformados $Y^* = TY$, $X^* = TX$ i $U^* = TU$, de dimensión, respectivamente, N , $N \times k$ y N . De esta manera, obtenemos el modelo transformado siguiente:

$$TY = TXB + TU, \quad (1.9)$$

o, lo que es el mismo:

$$Y^* = X^*B + U^*. \quad (1.10)$$

Con el fin de determinar cuáles tienen que ser los elementos de la matriz T , a continuación analizaremos las propiedades del nuevo término de perturbación, U^* :

- El valor esperado es éste:

$$E[U^*] = E[TU] = TE[U],$$

y, atendido a que U –que es el término de perturbación original, no esférico– tiene esperanza nula por hipótesis, tenemos que:

$$E[U^*] = TE[U] = \mathbf{0}_N. \quad (1.11)$$

En consecuencia, el término de perturbación del modelo transformado también tiene esperanza nula.

- La matriz de varianzas y covarianzas se escribe como vemos a continuación:

$$\begin{aligned} \text{VAR}[U^*] &= E\{[U^* - E[U^*]][U^* - E[U^*]]'\} = E[U^*U^{*'}] = E[TUU'T'] = \\ &= TE[UU']T' = \sigma^2 T\Omega T'. \end{aligned} \quad (1.12)$$

Pues bien, llegados a este punto, tenemos que recordar que la finalidad de transformar el modelo es conseguir un modelo con perturbaciones esféricas. Por lo tanto, tenemos que elegir la matriz T de manera que se cumpla la siguiente relación:

$$T\Omega T' = I_N. \quad (1.13)$$


En consecuencia, la validez de este método dependerá de si hay alguna manera de encontrar una **matriz** T que cumpla 1.13. Para encontrar esta matriz, nos basamos en la propiedad del álgebra matricial que establece que, ante una matriz A simétrica y definida positiva, siempre podemos encontrar una **matriz** P cuadrada y no singular tal que $A = PP'$. En nuestro caso tenemos la matriz Ω , que cumple ambas condiciones (simétrica y definida positiva), puesto que es una

matriz de varianzas y covarianzas, y, por lo tanto, siempre será posible descomponerla de manera que $\Omega = \mathbf{P}\mathbf{P}'$.

De esta manera, si definimos la matriz de transformación \mathbf{T} como \mathbf{P}^{-1} , tenemos la siguiente expresión:

$$\begin{aligned}\text{VAR}[\mathbf{U}^*] &= \sigma^2 \mathbf{T} \Omega \mathbf{T}' = \sigma^2 \mathbf{P}^{-1} \Omega (\mathbf{P}^{-1})' = \sigma^2 \mathbf{P}^{-1} (\mathbf{P}\mathbf{P}') (\mathbf{P}^{-1})' = \sigma^2 (\mathbf{P}^{-1} \mathbf{P}) \mathbf{P}' (\mathbf{P}^{-1})' = \\ &= \sigma^2 (\mathbf{P}^{-1} \mathbf{P}) \mathbf{P}' (\mathbf{P}')^{-1} = \sigma^2 (\mathbf{P}^{-1} \mathbf{P}) (\mathbf{P}' (\mathbf{P}')^{-1}) = \sigma^2 \mathbf{I}_N \mathbf{I}'_N = \sigma^2 \mathbf{I}_N.\end{aligned}$$

A modo de resumen, entonces, al transformar el modelo multiplicando por la matriz $\mathbf{T} = \mathbf{P}^{-1}$, conseguimos un nuevo modelo en el que el término de perturbación es esférico, que es precisamente lo que buscábamos.

Es necesario mencionar que los parámetros \mathbf{B} del modelo transformado 1.10 son los mismos que los del modelo original, aunque las variables (las observaciones) no lo sean, debido a que, con la transformación efectuada, tanto las variables explicativas como la variable endógena quedan sometidas al mismo cambio de escala. 

2) Estimación por MCO del modelo transformado

Teniendo en cuenta que el nuevo vector de términos de perturbación, \mathbf{U}^* , presenta una matriz de varianzas y covarianzas escalar, la estimación por MCO proporcionará estimadores que cumplirán todas las propiedades deseables. Por lo tanto, los estimadores MCG de los parámetros \mathbf{B} a los que llegamos por esta vía están determinados por la siguiente expresión:

$$\hat{\mathbf{B}}_{\text{MCG}} = (\mathbf{X}^* \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{Y}^*, \quad (1.14)$$

siendo \mathbf{X}^* e \mathbf{Y}^* la matriz y el vector transformados definidos con anterioridad.

De la misma manera, como ya sabemos, cuando la matriz de varianzas y covarianzas de los términos de perturbación es escalar, el estimador $\hat{\sigma}_u^2$ viene determinado por:

$$\hat{\sigma}_u^2 = \frac{\mathbf{e}'\mathbf{e}}{N - k}.$$

Pues bien, obtendremos el estimador del parámetro σ^2 , cuando las perturbaciones son no esféricas y mediante la aplicación de la expresión anterior (SCR dividida por los grados de libertad del modelo), pero, a partir del modelo transformado, es decir:

$$\hat{\sigma}_{\text{MCG}}^2 = \frac{\mathbf{e}^* \mathbf{e}^*}{N - k}, \quad (1.15)$$

siendo \mathbf{e}^* el vector de residuos estimados asociados al modelo transformado, esto es:

$$\begin{aligned}\mathbf{e}^* &= \mathbf{Y}^* - \hat{\mathbf{Y}}^* = \mathbf{Y}^* - \mathbf{X}^* \hat{\mathbf{B}}_{\text{MCG}} = \mathbf{Y}^* - \mathbf{X}^* (\mathbf{X}^* \mathbf{X}^*)^{-1} \mathbf{X}^* \mathbf{Y}^* = \\ &= [\mathbf{I}_N - \mathbf{X}^* (\mathbf{X}^* \mathbf{X}^*)^{-1} \mathbf{X}^*] \mathbf{Y}^* = \mathbf{M}^* \mathbf{Y}^*,\end{aligned} \quad (1.16)$$

donde $M^* = [I_N - X^*(X^{*'}X^*)^{-1}X^{*'}]$ presenta las propiedades de idempotencia, ortogonalidad respecto de la matriz X^* y simetría, y su rango coincide con su traza que es $N - k$.

Ved las propiedades de la matriz M en el subapartado 2.4.1 del módulo "Modelo de regresión lineal múltiple...".

1.3.2. Segunda aproximación a la estimación por MCG

Mientras que la aproximación que acabamos de ver consistía en una **aproximación indirecta** (o aproximación en dos etapas), la aproximación que presentamos a continuación supone una **estimación directa**, sin necesidad de que se produzca una transformación previa de los datos.

Para realizar esta estimación directa partimos de la expresión 1.14 y, simplemente sustituyendo X^* por TX , Y^* por TY y $T'T$ por Ω^{-1} , llegamos a la siguiente expresión:

$$\hat{B}_{MCG} = (X'T'TX)^{-1}X'T'TY = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y). \quad (1.17)$$

En cuanto a la estimación de σ^2 , partiendo de 1.15 y teniendo en cuenta que podemos escribir la suma de los cuadrados de los errores (SCR) asociada al modelo transformado como vemos a continuación:

$$\begin{aligned} e^{*'}e^* &= (Y^* - X^*\hat{B}_{MCG})'(Y^* - X^*\hat{B}_{MCG}) = \\ &= (TY - TX\hat{B}_{MCG})'(TY - TX\hat{B}_{MCG}) = (Y - X\hat{B}_{MCG})'T'T(Y - X\hat{B}_{MCG}) = \\ &= (Y - X\hat{B}_{MCG})'\Omega^{-1}(Y - X\hat{B}_{MCG}) = e'\Omega^{-1}e, \end{aligned} \quad (1.18)$$

tenemos que:

$$\hat{\sigma}_{MCG}^2 = \frac{e'\Omega^{-1}e}{N - k}, \quad (1.19)$$


donde, como podemos ver, el estimador de σ^2 se obtiene a partir de los residuos estimados asociados al modelo sin transformar estimado por MCG utilizando la matriz Ω^{-1} .

Forma alternativa de demostrar que $e^{*'}e^* = e'\Omega^{-1}e$

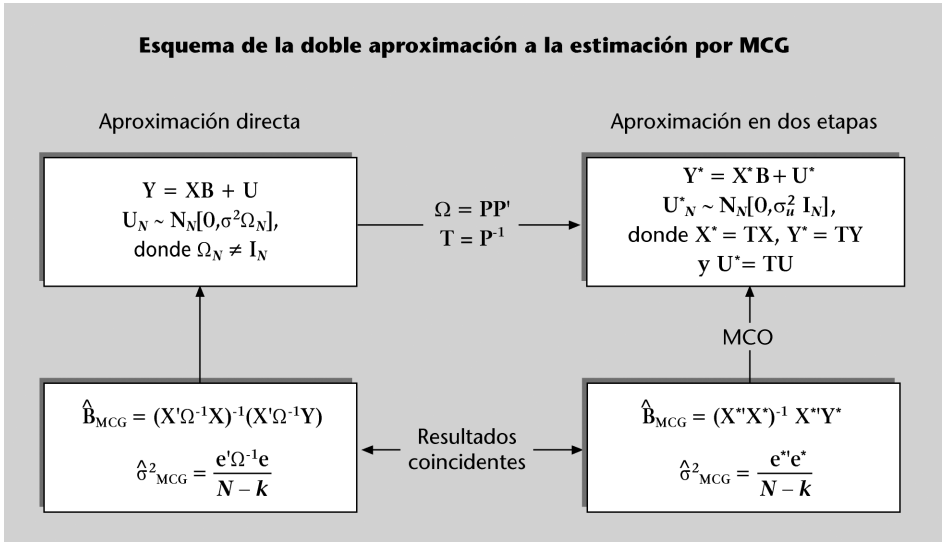
Presentamos, ahora, una forma alternativa a la que acabamos de presentar para llevar a cabo esta demostración:

$$\begin{aligned} e^* &= Y^* - \hat{Y}^* = TY - T\hat{Y} = T(Y - \hat{Y}) = Te \Rightarrow \\ e^{*'}e^* &= (Te)'(Te) = e'T'Te = e'(P^{-1})'P^{-1}e = e'\Omega^{-1}e, \end{aligned}$$

donde tenemos que recordar que $T = P^{-1}$.

Para acabar, sólo nos queda por señalar que calcularemos la suma de los residuos al cuadrado en el contexto de perturbaciones no esféricas a partir de la expresión 1.18, que relaciona ambas aproximaciones. 

A modo de resumen de ambas aproximaciones, a continuación presentamos un cuadro (ved la página siguiente) en el que podemos ver gráficamente de en qué consiste cada una:



Ejemplo de equivalencia de los dos métodos de estimación por MCG

Mediante un ejemplo numérico sencillo, mostraremos la equivalencia de los dos procedimientos alternativos de estimación por MCG presentados a lo largo de este apartado para estimar un modelo con perturbaciones no esféricas. Así, supongamos que deseamos estimar el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \quad \forall i = 1, \dots, N,$$

a partir de la siguiente información muestral:

$$X = \begin{bmatrix} 1 & 3 \\ 1 & 9 \\ 1 & 11 \end{bmatrix}, Y = \begin{bmatrix} 21 \\ 28 \\ 37 \end{bmatrix}.$$

Suponed, asimismo, que la matriz de varianzas y covarianzas de los términos de perturbación es la que presentamos a continuación:

$$\text{VAR}[U] = \sigma^2 \begin{bmatrix} 53 & 59 & 23 \\ 59 & 77 & 29 \\ 23 & 29 & 65 \end{bmatrix},$$

en la que podemos ver que el término de perturbación, además de ser heteroscedástico, presenta autocorrelación, ya que las varianzas son distintas y las covarianzas no son nulas. Como sabemos, en estas condiciones hay que estimar el modelo por MCG, aunque, por lo que conocemos del análisis teórico, este método de estimación admite una doble aproximación:

a) Estimación directa de los parámetros

En este caso, sólo tenemos que aplicar la siguiente expresión a los datos observados:

$$\hat{B}_{MCG} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y).$$

Así, en nuestro ejemplo tenemos:

$$\hat{B}_{MCG} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 3 & 9 & 11 \end{bmatrix} \Omega^{-1} \begin{bmatrix} 1 & 3 \\ 1 & 9 \\ 1 & 11 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 3 & 9 & 11 \end{bmatrix} \Omega^{-1} \begin{bmatrix} 21 \\ 28 \\ 37 \end{bmatrix},$$

siendo la inversa de la matriz de varianzas y covarianzas de los términos de perturbación (tened en cuenta que σ^2 se anula, porque es un escalar que aparece multiplicando y dividiendo):

$$\Omega^{-1} = \begin{bmatrix} 0,1285 & -0,0977 & -0,0018 \\ -0,0977 & 0,09 & -0,0055 \\ -0,0018 & -0,0055 & 0,0185 \end{bmatrix}.$$

Nota

Este ejemplo es meramente académico. En la práctica, no sería conveniente estimar este modelo sólo con tres observaciones, puesto que en exclusiva hay un grado de libertad.



Haciendo las operaciones correspondientes, llegamos al vector de estimadores MCG siguiente:

$$\hat{\mathbf{B}}_{\text{MCG}} = \begin{bmatrix} \hat{\beta}_{1\text{MCG}} \\ \hat{\beta}_{2\text{MCG}} \end{bmatrix} = \begin{bmatrix} 19,64 \\ 1,36 \end{bmatrix}.$$

b) Estimación indirecta (en dos etapas): mínimos cuadrados ponderados

La segunda posibilidad para estimar el modelo del ejemplo consiste en transformarlo previamente, de manera que sea adecuado utilizar la formulación de los estimadores MCO. Para ello, tenemos que multiplicar el modelo por una matriz T:

$$\mathbf{TY} = \mathbf{TXB} + \mathbf{TU}.$$

Obtendremos esta matriz de transformación T a partir de la matriz de varianzas y covarianzas de los términos de perturbación: $\Omega = \mathbf{PP}'$, siendo $\mathbf{T} = \mathbf{P}^{-1}$. Así, la matriz de transformación en el ejemplo es ésta:

$$\mathbf{T} = \frac{1}{180} \begin{bmatrix} -32 & 24 & 20 \\ 56 & -42 & 10 \\ -2 & 24 & -10 \end{bmatrix},$$

y las matrices de observaciones transformadas son las que vemos a continuación:

$$\mathbf{X}^* = \mathbf{TX} = \begin{bmatrix} 0,066 & 1,888 \\ 0,133 & -0,555 \\ 0,066 & 0,555 \end{bmatrix}, \mathbf{Y}^* = \mathbf{TY} = \begin{bmatrix} 4,111 \\ 2,055 \\ 1,444 \end{bmatrix}.$$

Para acabar, sólo nos queda por aplicar la formulación de los estimadores MCO en los nuevos datos. De esta manera, tenemos:

$$(\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{Y}^* = \begin{bmatrix} 19,64 \\ 1,36 \end{bmatrix} = \hat{\mathbf{B}}_{\text{MCG}}.$$

Como podemos ver, llegamos a los mismos resultados que obteníamos mediante el método directo.

Ved. en los apartados 2 y 3 de este módulo didáctico, cómo podemos obtener la matriz T dependiendo de la situación en la que nos encontramos (heteroscedasticidad o autocorrelación).

1.4. Propiedades de los estimadores MCG

En este subapartado estudiaremos las propiedades de los estimadores MCG cuando el término de perturbación es no esférico, y veremos que los $\hat{\mathbf{B}}_{\text{MCG}}$ son los estimadores eficientes de entre la familia de estimadores no sesgados y consistentes, y que el estimador $\hat{\sigma}_{\text{MCG}}^2$ es no sesgado.

1.4.1. Propiedades de los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$


Para estudiar las propiedades de $\hat{\mathbf{B}}_{\text{MCG}}$ partiremos de la expresión 1.17 correspondiente al estimador MCG aplicado en el modelo original*:

$$\hat{\mathbf{B}}_{\text{MCG}} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{Y}), \quad (1.20)$$

y veremos si en efecto es no sesgado, eficiente y consistente en ECM. Para hacer esto, sustituimos en 1.20 la variable endógena por su expresión, $\mathbf{XB} + \mathbf{U}$, y, operando, llegamos a la expresión:

$$\begin{aligned} \hat{\mathbf{B}}_{\text{MCG}} &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{XB}) + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{U}) = \\ &= \mathbf{B} + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{U}). \end{aligned} \quad (1.21)$$

* El modelo original es el modelo sin transformar.

A partir de las expresiones 1.20 y 1.21, podemos deducir las propiedades de $\hat{\mathbf{B}}_{\text{MCG}}$, que son las siguientes: 

1) Con el fin de estudiar si es un estimador sesgado o no sesgado, tomamos esperanzas directamente de 1.21 y obtenemos:

$$E[\hat{\mathbf{B}}_{\text{MCG}}] = E[\mathbf{B}] + E[(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{U})] = \mathbf{B} + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1})E[\mathbf{U}] = \mathbf{B},$$

lo cual muestra que, en efecto, se trata de un estimador no sesgado.

2) En cuanto a la matriz de varianzas y covarianzas de los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$, tenemos lo siguiente:

$$\begin{aligned} \text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] &= E\{[\hat{\mathbf{B}}_{\text{MCG}} - E[\hat{\mathbf{B}}_{\text{MCG}}]][\hat{\mathbf{B}}_{\text{MCG}} - E[\hat{\mathbf{B}}_{\text{MCG}}]]'\} = \\ &= E[(\hat{\mathbf{B}}_{\text{MCG}} - \mathbf{B})(\hat{\mathbf{B}}_{\text{MCG}} - \mathbf{B})'] = E\{[(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{U}][(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{U}]]'\} = \\ &= E\{(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{U}\mathbf{U}'\Omega^{-1}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\} = \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}E[\mathbf{U}\mathbf{U}']\Omega^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} = \\ &= \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\Omega\Omega^{-1}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}. \end{aligned}$$

Es interesante que nos demos cuenta de que...

... si $\Omega = \mathbf{I}_N$, entonces se cumple la siguiente igualdad:
 $\text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] = \text{VAR}[\hat{\mathbf{B}}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$

Todo lo que hemos visto anteriormente permite escribir:

$$\hat{\mathbf{B}}_{\text{MCG}} \sim \mathbf{N}_k[\mathbf{B}, \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}].$$

Volviendo a reanudar el hilo del análisis de la matriz de varianzas y covarianzas de los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$, tenemos que comprobar, una vez conocida su expresión, que son estimadores eficientes (de varianza mínima). En este sentido, el teorema de Aitken (que es el equivalente al de Gauss-Markov, pero para el caso de perturbaciones no esféricas, por lo que podemos decir que es una generalización de éste) muestra que los estimadores de MCG son los de varianza mínima de entre todos los estimadores no sesgados cuando el término de perturbación es no esférico.

3) Para acabar, nos queda por estudiar si los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$ cumplen la propiedad de consistencia en ECM. Y, para verlo, tenemos que aplicar la expresión conocida de ECM, teniendo en cuenta que, como hemos visto antes, son estimadores no sesgados:

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{EQM}(\hat{\mathbf{B}}_{\text{MCG}}) &= \lim_{N \rightarrow \infty} \{[\text{Sesgo}[\hat{\mathbf{B}}_{\text{MCG}}]]^2 + \text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}]\} = \lim_{N \rightarrow \infty} \text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] = \\ &= \lim_{N \rightarrow \infty} \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} = \lim_{N \rightarrow \infty} \left[\frac{\sigma^2}{N} \left(\frac{\mathbf{X}'\Omega^{-1}\mathbf{X}}{N} \right)^{-1} \right] = 0. \end{aligned}$$

Este resultado muestra que los estimadores $\hat{\mathbf{B}}_{\text{MCG}}$ son consistentes en ECM.

Lectura complementaria

Para conocer más detalles sobre la eficiencia de los estimadores MCG, podéis consultar, por ejemplo, la obra:

A. Novales (1993).
Econometría (2.ª ed.).
 Madrid: McGraw-Hill.

Resumiendo, entonces, podemos concluir que los **estimadores MCG** son los estimadores no sesgados, eficientes y consistentes en ECM cuando los términos de perturbación son no esféricos.

1.4.2. Propiedades del estimador $\hat{\sigma}_{MCG}^2$

El estimador $\hat{\sigma}_{MCG}^2$ es un estimador no sesgado. Para comprobarlo, sólo tenemos que tomar esperanzas en la expresión del estimador mencionado:

$$E[\hat{\sigma}_{MCG}^2] = E\left[\frac{\mathbf{e}^* \mathbf{e}^*}{N - k}\right] = \frac{1}{N - k} E[\mathbf{e}^* \mathbf{e}^*]. \quad (1.22)$$

Por este motivo, partiendo de la expresión 1.16 y sustituyendo la \mathbf{Y} por su expresión, obtenemos que:

$$\mathbf{e}^* = \mathbf{M}^* \mathbf{Y}^* = \mathbf{M}^* (\mathbf{X}^* \mathbf{B} + \mathbf{U}^*) = \mathbf{M}^* \mathbf{X}^* \mathbf{B} + \mathbf{M}^* \mathbf{U}^*. \quad (1.23)$$

Teniendo en cuenta que \mathbf{M}^* es ortogonal a \mathbf{X}^* , el primer sumando de 1.23 se anula. Por lo tanto, el sumatorio de los errores al cuadrado queda de la manera que podéis ver a continuación:

$$\begin{aligned} \mathbf{e}^* \mathbf{e}^* &= \mathbf{U}^* \mathbf{M}^* \mathbf{M}^* \mathbf{U}^* = \mathbf{U}^* \mathbf{M}^* \mathbf{U}^* \Rightarrow \\ \Rightarrow E[\mathbf{e}^* \mathbf{e}^*] &= E[\mathbf{U}^* \mathbf{M}^* \mathbf{U}^*] = E[\text{tr}[\mathbf{U}^* \mathbf{M}^* \mathbf{U}^*]] = E[\text{tr}[\mathbf{M}^* \mathbf{U}^* \mathbf{U}^{*'}]] = \text{tr}[E[\mathbf{M}^* \mathbf{U}^* \mathbf{U}^{*'}]] = \\ &= \text{tr}[\mathbf{M}^* E[\mathbf{U}^* \mathbf{U}^{*'}]] = \text{tr}[\mathbf{M}^* \sigma_u^2 \mathbf{I}_N] = \sigma_u^2 \text{tr}[\mathbf{M}^*] = \sigma_u^2 (N - k). \end{aligned} \quad (1.24)$$


Recordad que la traza de cualquier matriz idempotente coincide con su rango. En este caso concreto tenemos:

$$\begin{aligned} \text{tr}[\mathbf{M}^*] &= \text{tr}[\mathbf{I}_N - \mathbf{X}^* (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'}] = \text{tr}[\mathbf{I}_N] - \text{tr}[\mathbf{X}^* \mathbf{X}^* (\mathbf{X}^{*'} \mathbf{X}^*)^{-1}] = \\ &= \text{tr}[\mathbf{I}_N] - \text{tr}[\mathbf{I}_k] = N - k. \end{aligned}$$

Así pues, el estimador $\hat{\sigma}_{MCG}^2$ es no sesgado, atendiendo a que:

$$E[\hat{\sigma}_{MCG}^2] = \frac{1}{N - k} E[\mathbf{e}^* \mathbf{e}^*] = \frac{1}{N - k} \sigma_u^2 (N - k) = \sigma_u^2.$$

1.5. El estimador máximo verosímil (MV) en el MRLMG

Así como hicimos en el contexto del modelo de regresión lineal múltiple estándar, el objetivo de este subapartado es encontrar los estimadores de máxima verosimilitud, pero ahora para aquellos casos en los que los términos de perturbación son no esféricos. 

El supuesto básico para llegar a la expresión del estimador máximo verosímil (MV) es la distribución que siguen los términos de perturbación del modelo. En el caso que analizamos, se supone que se distribuyen según una ley normal:

$$\mathbf{U}_N \sim \mathbf{N}_N[\mathbf{0}, \sigma^2 \Omega_N], \text{ donde } \Omega_N \neq \mathbf{I}_N.$$

En los subapartados 1.3.1 y 1.3.2 de este módulo hallaremos la definición del estimador $\hat{\sigma}_{MCG}^2$.

Ved la matriz \mathbf{M}^* en el subapartado 1.3.1 de este módulo. Ved, también, la traza de \mathbf{M} en el subapartado 2.4.1 del módulo "Modelo de regresión lineal múltiple...".

Una matriz es idempotente...

... cuando, multiplicada por sí misma, da la identidad. En el caso de la matriz \mathbf{M}^* , tenemos que recordar que es simétrica y, en consecuencia, coincide con su matriz invertida. Entonces:

$$\mathbf{M}^* \mathbf{M}^* = \mathbf{M}^* \mathbf{M}^* = \mathbf{I}.$$

Ved la estimación por máxima verosimilitud del MRLMG estándar en el subapartado 2.5 del módulo "Modelo de regresión lineal múltiple...".

Como sabemos, los **estimadores de máxima verosimilitud** son aquellos que resultan de maximizar la función de verosimilitud. La **función de verosimilitud** se construye a partir de la función de distribución conjunta de los términos de perturbación, así que, en primer lugar, tenemos que encontrar esta función.

Ved la función de verosimilitud en el subapartado 2.5 del módulo "Modelo de regresión lineal múltiple...".

Para construir la función de verosimilitud, partimos de la función de densidad de los términos de perturbación y, a partir de ésta, obtenemos la función de densidad conjunta de los N términos de perturbación. En el caso de perturbaciones no esféricas, podemos demostrar que esta función está determinada por la siguiente expresión*:

* En la expresión 1.25, $|\Omega|$ representa el determinante de la matriz Ω y $f(u_i)$ es la función densidad de probabilidad normal.

$$\begin{aligned} P(\mathbf{U}) &= f(u_1, u_2, \dots, u_N) = \prod_{i=1}^N f(u_i) = \\ &= \frac{1}{(2\pi)^{N/2}} \frac{1}{(\sigma^2)^{N/2}} \frac{1}{|\Omega|^{1/2}} \exp\left[-\frac{1}{2\sigma^2} \mathbf{U}' \Omega^{-1} \mathbf{U}\right]. \end{aligned} \quad (1.25)$$

El supuesto de distribución normal multivariante para el vector de los términos de perturbación implica una distribución normal multivariante para el vector de las variables endógenas. De esta manera, podemos demostrar que la función de densidad conjunta de los componentes de la variable endógena coincide con la de los términos de perturbación 1.25. Para pasar de la función de densidad conjunta anterior a la función de verosimilitud, sólo tenemos que sustituir el vector \mathbf{U} por $\mathbf{Y} - \mathbf{XB}$ en 1.25. Entonces nos queda:

$$L(\mathbf{Y}; \mathbf{B}, \sigma^2) = (2\pi)^{-N/2} (\sigma^2)^{-N/2} |\Omega|^{-1/2} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})' \Omega^{-1} (\mathbf{Y} - \mathbf{XB})\right]. \quad (1.26)$$

Como sabemos, el principio de máxima verosimilitud consiste en seleccionar los estimadores de los parámetros desconocidos (\mathbf{B} y σ^2) que maximizan la función de verosimilitud. Por lo tanto, tenemos que igualar la primera derivada de 1.26 a cero. Así, tenemos nos encontramos con que:

$$\frac{\partial L[\mathbf{Y}; \mathbf{B}, \sigma^2]}{\partial \mathbf{B}} = 0, \quad \frac{\partial L[\mathbf{Y}; \mathbf{B}, \sigma^2]}{\partial \sigma^2} = 0.$$

Pero, el cálculo de estas derivadas es complicado, teniendo en cuenta la expresión de la función de verosimilitud. Por este motivo, se acostumbra a trabajar con el logaritmo neperiano de la función de verosimilitud, ya que simplifica de manera importante el cálculo de las derivadas mencionadas:

$$\begin{aligned} \ln L(\mathbf{Y}; \mathbf{B}, \sigma^2) &= \frac{-N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2} \ln |\Omega| - \\ &- \frac{1}{2\sigma^2} (\mathbf{Y} - \mathbf{XB})' \Omega^{-1} (\mathbf{Y} - \mathbf{XB}). \end{aligned} \quad (1.27)$$

Así pues, los estimadores MV de los parámetros del vector \mathbf{B} son aquellos que maximizan la expresión 1.27. Sin embargo, dado que los tres primeros términos no dependen de \mathbf{B} , los estimadores mencionados serán aquellos que maximicen el cuarto término. Con lo que, derivando 1.27 respecto de \mathbf{B} , obtendremos las k ecuaciones normales:

$$\begin{aligned} \frac{\partial \ln L[\mathbf{Y}; \mathbf{B}, \sigma^2]}{\partial \mathbf{B}} &= \frac{\partial (\mathbf{Y}'\Omega^{-1}\mathbf{Y} - 2\mathbf{B}'\mathbf{X}'\Omega^{-1}\mathbf{Y} + \mathbf{B}'\mathbf{X}'\Omega^{-1}\mathbf{X}\mathbf{B})}{\partial \mathbf{B}} = \\ &= -2\mathbf{X}'\Omega^{-1}\mathbf{Y} + 2\mathbf{X}'\Omega^{-1}\mathbf{X}\mathbf{B} = \mathbf{0}. \end{aligned} \quad (1.28)$$


Resolviendo este sistema de k ecuaciones, llegaremos a la expresión de los estimadores $\hat{\mathbf{B}}_{MV}$ que vemos a continuación:

$$\hat{\mathbf{B}}_{MV} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega^{-1}\mathbf{Y}),$$

que, como podemos ver, coinciden con los estimadores de MCG.

Por otra parte, para obtener el estimador máximo verosímil de la varianza del término de perturbación, tenemos que derivar 1.27 con respecto al parámetro de interés y resolver la ecuación que resulta de igualar esta derivada a cero. Así, podemos demostrar que el estimador MV de la varianza del término de perturbación viene determinado por la siguiente expresión:

$$\hat{\sigma}_{MV}^2 = \frac{\mathbf{e}'\Omega^{-1}\mathbf{e}}{N}. \quad (1.29)$$

Como podemos ver en 1.29, el estimador de máxima verosimilitud de la varianza del término de perturbación, a diferencia de lo que sucedía con el de los parámetros \mathbf{B} , no coincide con el de MCG. 

1.5.1. Propiedades de los estimadores MV

A partir de los cálculos que hemos efectuado en el subapartado podemos deducir las siguientes conclusiones: 

a) En primer lugar, por lo que respecta al estimador $\hat{\mathbf{B}}_{MV}$, ya que coincide con el estimador $\hat{\mathbf{B}}_{MCG}$, cumplirá las mismas propiedades que éste. Por lo tanto, se trata de un estimador no sesgado, de varianza mínima y consistente en ECM.

b) En cuanto al estimador $\hat{\sigma}_{MV}^2$, es un estimador sesgado, y, para comprobarlo, tomamos esperanzas en 1.29 y obtenemos:

$$E[\sigma_{MV}^2] = E\left[\frac{\mathbf{e}'\Omega^{-1}\mathbf{e}}{N}\right] = \frac{1}{N}E[\mathbf{e}'\Omega^{-1}\mathbf{e}],$$

y, como sabemos que $E[\mathbf{e}'\Omega^{-1}\mathbf{e}] = \sigma^2(N - k)$, tal y como hemos demostrado en 1.24, llegamos al siguiente resultado:

$$E[\sigma_{MV}^2] = \frac{1}{N} E[\mathbf{e}'\Omega^{-1}\mathbf{e}] = \frac{N - k}{N} \sigma^2 \neq \sigma^2. \quad (1.30)$$

De 1.30 deducimos que el sesgo es:

$$\text{Sesgo}[\hat{\sigma}_{MV}^2] = E[\hat{\sigma}_{MV}^2] - \sigma^2 = \frac{N - k}{N} \sigma^2 - \sigma^2 = -\frac{k}{N} \sigma^2,$$

con lo que se trata de un estimador sesgado por defecto: $\hat{\sigma}_{MV}^2 < \sigma_u^2$.


En la expresión 1.24,...

... se demuestra que $E[\mathbf{e}^*\mathbf{e}^*] = \sigma^2(N - k)$; pero también debemos recordar la expresión 1.18, que relaciona $\mathbf{e}^*\mathbf{e}^*$ con $\mathbf{e}'\Omega^{-1}\mathbf{e}$ y tener en cuenta que $\hat{\mathbf{B}}_{MV} = \hat{\mathbf{B}}_{MCG}$.

Sin embargo, por otra parte, los resultados obtenidos ponen de manifiesto que es un estimador asintóticamente no sesgado, atendiendo a que, cuando $N \rightarrow \infty$, el sesgo tiende a cero, o, lo que es lo mismo:

$$\lim_{N \rightarrow \infty} \frac{N - k}{N} \rightarrow 1 \Rightarrow E[\hat{\sigma}_{MV}^2] \rightarrow \sigma_u^2.$$

1.6. Normalidad del término de perturbación y contraste de normalidad de Jarque-Bera

Validar o refutar la hipótesis de distribución normal del término de perturbación es importante por varios motivos, de entre las cuales podemos destacar las dos que mencionamos a continuación: 

a) Si bien las propiedades de los estimadores MCO y MCG no dependen de la distribución de probabilidad que se suponga para el término de perturbación, la derivación que hemos efectuado para encontrar el estimador MV sí que depende del supuesto de normalidad de las perturbaciones.

b) Todo el proceso de inferencia está basado en el supuesto de distribución normal de las perturbaciones, puesto que los estimadores sólo se distribuyen según una normal si, y sólo si, el término de perturbación se distribuye según una ley normal. En consecuencia, si la hipótesis de normalidad de la perturbación no se cumple, los test de la t de Student y la F de Snedecor que conocemos no son válidos.

En resumidas cuentas, la **normalidad de la perturbación** es una hipótesis básica que se tiene que cumplir, en primer lugar, para la construcción de los estimadores MV tal y como los hemos estudiado y, en segundo lugar, para la realización de inferencia, es decir, los contrastes de hipótesis.

El contraste de normalidad de Jarque-Bera

Las hipótesis nula y alternativa de este contraste son éstas:

$$H_0 : U \sim \text{Normal.}$$

$$H_A : U \not\sim \text{Normal.}$$

El contraste de normalidad de Jarque-Bera está basado en el **coeficiente de asimetría*** y en el **coeficiente de apuntamiento**** de los residuos MCO de la estimación de cuyo modelo queremos contrastar si el término de perturbación se distribuye de acuerdo con una ley normal o no; en otras palabras, se basa en la forma que presenta el histograma de los residuos estimados.

* El término inglés correspondiente a asimetría es *skewness*.

** El término inglés correspondiente a apuntamiento es *kurtosis*.

Los coeficientes de asimetría, b_1 , y de apuntamiento, b_2 , vienen determinados por las siguientes expresiones:

$$b_1 = \frac{\sqrt{N} \sum_{i=1}^N e_i^3}{\left(\sum_{i=1}^N e_i^2 \right)^{3/2}}, \quad b_2 = \frac{N \sum_{i=1}^N e_i^4}{\left(\sum_{i=1}^N e_i^2 \right)^2},$$

donde e_i representa los residuos del ajuste MCO del modelo.


A continuación presentamos la definición del estadístico de prueba:

$$\chi_0^2 = N \left[\frac{b_1^2}{6} + \frac{(b_2 - 3)^2}{24} \right] \sim \chi_2^2.$$

La regla de decisión que utiliza el contraste de Jarque-Bera es la siguiente:

- Si $\chi_0^2 \geq \chi_{2;\alpha}^2 \Rightarrow$ rechazamos $H_0 \Rightarrow U \nmid$ Normal.
- Si $\chi_0^2 < \chi_{2;\alpha}^2 \Rightarrow$ no rechazamos $H_0 \Rightarrow U \sim$ Normal.

De manera intuitiva, la idea que subyace en este test es que, cuanto más asimétrico (la parte derecha de la izquierda) y menos mesocúrtico (más platicúrtico o más leptocúrtico) sea el histograma de los residuos MCO, mayor será el valor del estadístico de prueba, lo cual será un reflejo del hecho de que tiene menos apariencia de distribución normal. De hecho, si la distribución es perfectamente simétrica ($b_1 = 0$) y mesocúrtica ($b_2 = 3$), el estadístico de prueba vale cero, hecho que indica que nos encontramos ante una distribución normal “perfecta” en los residuos MCO. En este caso, el valor de χ_0^2 conducirá, como es lógico, a no rechazar la hipótesis nula de normalidad de los términos de perturbación.

Una forma gráfica de contrastar la normalidad de las perturbaciones consiste en visualizar el histograma de los residuos. Este gráfico transmite básicamente la misma información que el contraste de Jarque-Bera, aunque también permite ver con facilidad si el perfil de la distribución de los residuos se parece o no al de una distribución normal. 

Ejemplo de aplicación del test de Jarque-Bera

A continuación, nos disponemos a presentar una aplicación del test de Jarque-Bera. El ejemplo tiene por objetivo mostrar únicamente cómo hay que llevar a cabo el contraste mencionado, y presentar los resultados dejando a un lado otros conceptos o cuestiones que ya conocemos de los subapartados anteriores.

Los términos *platicúrtico*, *leptocúrtico* y *mesocúrtico*

Los términos *platicúrtico* y *leptocúrtico* indican que la función de distribución considerada tiene una forma más achatada o más puntiaguda que la distribución normal estándar. El término *mesocúrtico* indica que la función de distribución se comporta como una distribución normal estándar.

Así, supongamos que, a partir de 1.000 observaciones correspondientes a una variable endógena y dos variables explicativas, X_2 y X_3 , se ha estimado por MCO el siguiente modelo de regresión:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \forall i = 1, 2, \dots, 1.000.$$

Para comprobar la normalidad del término de perturbación mediante la aplicación del test de Jarque-Bera, tras haber estimado el modelo por MCO (presentamos los resultados de esta estimación en el cuadro de resultados que tenemos continuación), debemos obtener los residuos del ajuste, y, acto seguido, los coeficientes de asimetría y de apuntamiento (presentamos este resultado, junto con otros estadísticos asociados a los residuos, en el cuadro de contraste de normalidad que veremos más adelante) para calcular, finalmente, el estadístico de prueba.

Resultados de la estimación por MCO

Variable dependiente: Y				
Número de observaciones: 1000				
VARIABLE	COEFICIENTE	ERROR ESTAD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	2.9573818	0.0811362	36.449594	0.0000
X_2	5.6300000	0.1053563	53.437706	0.0000
X_3	-2.7647678	0.1030915	-26.818580	0.0000
R-squared	0.785027	Mean of dependent var	4.478648	
Adjusted R-squared	0.784596	S.D. of dependent var	2.065799	
S.E. of regression	0.958772	Sum of squared resid	916.4851	
Log likelihood	-1375.334	F-statistic	1820.397	
Durbin-Watson stat	1.913624	Prob(F-statistic)	0.000000	

Como podemos observar en el siguiente cuadro, los coeficientes de asimetría y de apuntamiento valen, respectivamente, $-0,008076$ (valor muy próximo a cero, lo cual significa que se trata de una distribución muy poco asimétrica) y $2,968900$ (muy próximo a tres; así que, es una distribución casi mesocúrtica). Estos resultados ya nos hacen pensar que no podremos rechazar la hipótesis nula de normalidad.

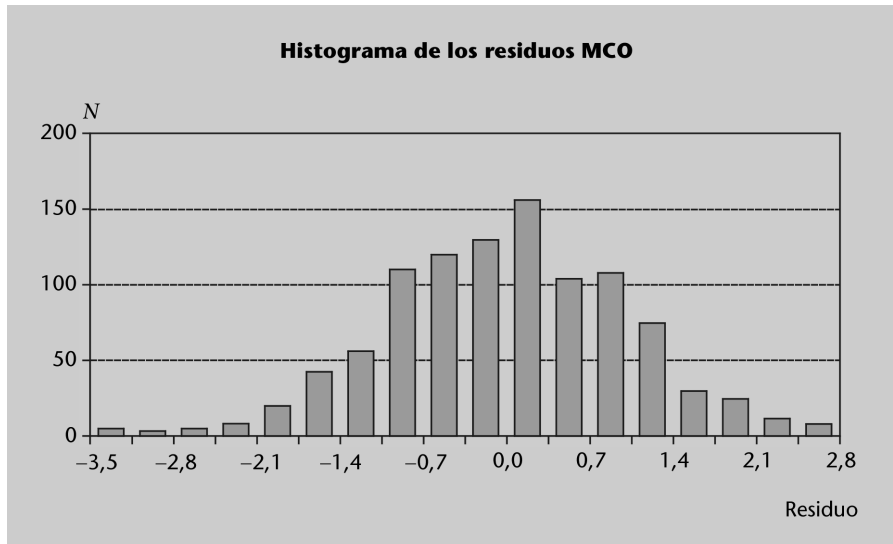
Contraste de normalidad

Rango: 1-1000			
Número de observaciones: 1000			
Skewness	-0.008076	Kurtosis	2.968900
Jarque-Bera normality test stat.	0.051171	Probability	0.974739

Esta sospecha se confirma al calcular el estadístico de prueba del test de Jarque-Bera, que, como pudimos ver en el cuadro anterior, es $0,051171$ y deja a su dere-

cha una probabilidad del 97,4739%, lo cual nos lleva, en efecto, a considerar que el término de perturbación tiene una distribución normal.

Para terminar, mostramos el histograma de los residuos que confirma la normalidad:



2. Heteroscedasticidad

Tras haber estudiado el caso general sobre el incumplimiento de las hipótesis básicas del MRLM por lo que respecta al término de perturbación, veremos el caso particular del incumplimiento del supuesto básico de homoscedasticidad, es decir, el problema de la presencia de perturbaciones heteroscedásticas en el modelo de regresión. Analizaremos las causas y las consecuencias de la existencia de heteroscedasticidad en el modelo de regresión, y la manera de detectar y solucionar el problema en el modelo.

! Ved el caso general sobre el incumplimiento de las hipótesis básicas del MRLM sobre el término de perturbación en el apartado 1 de este módulo didáctico.

2.1. Definición y causas

En este subapartado veremos con claridad en qué consiste la heteroscedasticidad y las posibles causas que generan este problema en torno al modelo de regresión. !

Tal y como vimos, uno de los supuestos que hicimos sobre el comportamiento del término de perturbación es que su varianza era constante e igual para todas las observaciones, es decir, que todos los elementos de la diagonal de la matriz de varianzas y covarianzas del término de perturbación eran constantes e iguales entre sí:

$$\text{VAR}[\mathbf{U}] = \begin{bmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \sigma_u^2 \mathbf{I}_N. \quad (2.1)$$

! Ved las hipótesis sobre el término de perturbación en el MRLM del subapartado 2.2.2 del módulo "Modelo de regresión lineal múltiple...".

Cuando el término de perturbación cumple este supuesto, decimos que es homoscedástico, o bien que se cumple el supuesto de homoscedasticidad.



Situación de homoscedasticidad

En este gráfico, aparece representada una situación de homoscedasticidad en un modelo de regresión lineal simple, en el supuesto de que el término de perturbación se distribuye según una normal. En este gráfico podemos apreciar que la variabilidad del término de perturbación es la misma para todas las observaciones.


Cuando el supuesto de homoscedasticidad no se cumple, es decir, cuando la varianza del término de perturbación no es constante y, por lo tanto, todos los elementos de la diagonal principal de la matriz de varianzas y covarianzas no son iguales entre sí, diremos que el término de perturbación es heteroscedástico o bien que tenemos un **problema de heteroscedasticidad en el modelo**.


En caso de heteroscedasticidad, la matriz de varianzas y covarianzas del término de perturbación dejará de ser igual a 2.1, y responderá a la expresión que vemos a continuación:

$$\text{VAR}[\mathbf{U}] = \begin{bmatrix} \sigma_{u_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{u_2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{u_N}^2 \end{bmatrix} = \sigma^2 \Omega_N; \quad (2.2)$$

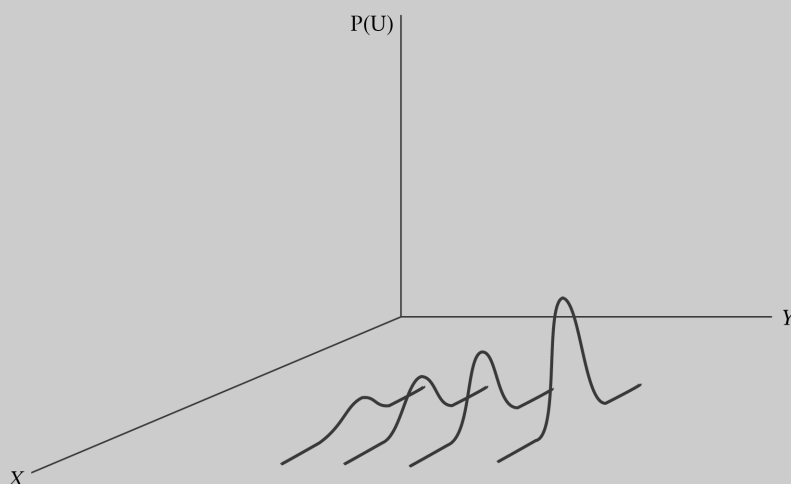
es decir:

$$\text{VAR}[u_i] = \sigma_{u_i}^2 \quad \forall i = 1, \dots, N.$$

En cualquier caso, es importante tener en cuenta que para que haya heteroscedasticidad no todas las varianzas de los términos de perturbación tienen que ser diferentes, sino que sólo es preciso que una de las varianzas de algún término de perturbación sea diferente de la del resto. Esta situación tendrá lugar cuando las varianzas de los términos de perturbación varíen de unos individuos a otros (en caso de trabajar con datos de corte transversal) o cuando varíen a lo largo del tiempo (en caso de trabajar con datos de ámbito temporal). 

 Ved los datos de corte transversal y los datos de ámbito temporal en el subapartado 1.3 del módulo "Modelo de regresión lineal múltiple...".


Situación de heteroscedasticidad en un MRLS



Situación de heteroscedasticidad

En este gráfico aparece registrada la situación en la que la variabilidad de los diferentes términos de perturbación no es la misma, por lo que tendremos una situación de heteroscedasticidad.

Tras haber visto en qué consiste la heteroscedasticidad, analizaremos los motivos que pueden dar lugar al hecho de que aparezca este problema en el modelo de

regresión. Por norma general, las causas se clasifican en tres grupos diferentes: estructurales o teóricas, muestrales y espurias. A continuación, presentamos las **causas de la heteroscedasticidad** con más detalle: 

a) Las **causas teóricas** se producen cuando trabajamos con modelos de corte transversal con unidades muestrales que presentan un comportamiento muy heterogéneo.

Ejemplos de heteroscedasticidad por causas teóricas

Veamos ahora dos modelos que presentan heteroscedasticidad por causas teóricas:

1) Suponemos que tenemos al siguiente modelo:

$$C_i = \beta_1 + \beta_2 Y_i + u_i \quad \forall i = 1, \dots, N,$$

en el que se explica el nivel de consumo, C_i , de diferentes unidades familiares en función de su renta, Y_i . Las familias con niveles de renta bajos tendrán un comportamiento muy homogéneo entre sí, es decir, sus gastos en bienes de consumo representarán una proporción muy elevada de su renta destinada a las necesidades básicas. Por otra parte, las familias con rentas más elevadas tendrán pautas de consumo más heterogéneas entre sí, dado que una vez satisfechas sus necesidades básicas dispondrán de un mayor excedente de renta que pueden destinar a otro tipo de consumo, o bien al ahorro.

En este caso, tendremos un comportamiento heteroscedástico, ya que a medida que la renta aumenta, también se produce un incremento en la variabilidad del consumo, que no se explica por ninguna variable del modelo.


2) Otro ejemplo en el que hallaremos una situación similar a la anterior se encontraría en el siguiente modelo:

$$Div_i = \beta_1 + \beta_2 B_i^0 + u_i \quad \forall i = 1, \dots, N,$$

en el que los dividendos repartidos (Div_i) por las empresas dependen de los beneficios obtenidos (B_i^0). A medida que los beneficios incrementan, la proporción de éstos que cada una de las empresas destinará a retribuir a sus accionistas variará en gran medida. Algunas empresas preferirán utilizar los beneficios para la autofinanciación, otras querrán tener contentos a sus accionistas, etc.

Como hemos visto, los dos ejemplos utilizados tienen en común la existencia de un comportamiento muy heterogéneo en las unidades muestrales en cuanto a su comportamiento respecto de las variables analizadas en el modelo.

En otro módulo de esta asignatura veremos el caso en el que la variable endógena que tenemos que modelizar es cualitativa. En este caso, el modelo de regresión conocido como *modelo de probabilidad lineal* será heteroscedástico.

Ved el modelo de probabilidad lineal en el subapartado 1.1 del módulo "Variables dependientes cualitativas". 

b) Por otra parte, la heteroscedasticidad no tiene por qué surgir en el modelo únicamente por motivos teóricos. En algunas ocasiones, los problemas de heteroscedasticidad aparecen debido a la manera de registrar la información muestral (**causas muestrales**). Estas situaciones se dan cuando no se encuentra disponible la información para la que está especificado el modelo, y hay que trabajar, por ejemplo, con valores medios o agregados para submuestras de la muestra global.

Ejemplo de heteroscedasticidad por causas muestrales

Supongamos que queremos estudiar los gastos en bienes de consumo de las familias catalanas, pero estos datos no están disponibles, por lo que tenemos que trabajar con los gastos en consumo agregado o medio de las comarcas catalanas.

Pues bien, el modelo con datos agrupados presentará heteroscedasticidad si el número de individuos de cada uno de los grupos es diferente, aunque el modelo original no tenga problemas de heteroscedasticidad.

A continuación vamos a ver las dos situaciones antes mencionadas con las que podemos encontrarnos:

- En el caso de que estuviésemos trabajando con **datos en valores medios de los datos iniciales**, la varianza del término de perturbación será inversamente proporcional al número de individuos u observaciones de cada submuestra o grupo.

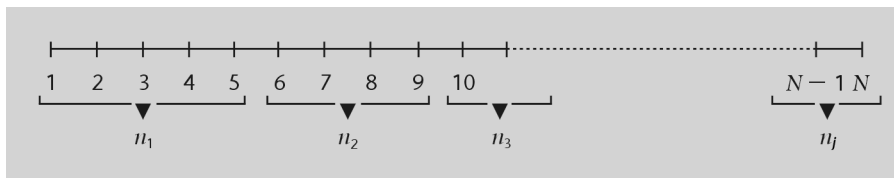
Varianza del término de perturbación cuando se trabaja con datos en valores medios

Para ilustrar esta situación, utilizaremos el siguiente ejemplo: supongamos que tenemos el modelo de regresión lineal simple que mostramos vemos aquí:

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \quad \forall i = 1, \dots, N,$$

$$u_i \sim N[0, \sigma_u^2], \quad E[u_i u_j] = 0 \quad \forall i \neq j,$$

que está correctamente especificado y no presenta ningún problema en cuanto a su término de perturbación. Sin embargo, por problemas con la disponibilidad de los datos, sólo disponemos de los valores agregados en medias de las variables integrantes del modelo para J grupos o submuestras, siendo n_j el número de observaciones o individuos registrados en cada uno de los grupos. Por ejemplo, podríamos tener J grupos como los que podemos ver en el siguiente gráfico, en el que hay 5 observaciones al primer grupo, 4 al segundo, etc.:



Por lo tanto, tendremos que $\sum_{j=1}^J n_j = N$ y el modelo especificado será:

$$\bar{Y}_j = \beta_1^* + \beta_2^* \bar{X}_j + \bar{u}_j \quad \forall j = 1, \dots, J,$$

dónde \bar{Y}_j y \bar{X}_j son los valores medios de los componentes de la variable endógena y de la variable exógena para cada grupo, es decir:

$$\bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_i \quad \forall j = 1, \dots, J,$$

$$\bar{X}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} X_i \quad \forall j = 1, \dots, J.$$

De manera análoga, el término de perturbación del modelo anterior, \bar{u}_j , será la media de los n_j términos de perturbación de cada uno de los grupos:

$$\bar{u}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} u_i \quad \forall j = 1, \dots, J.$$

Así pues, el valor esperado y la varianza del nuevo término de perturbación \bar{u}_j serán:

$$E[\bar{u}_j] = E\left[\frac{1}{n_j} \sum_{i=1}^{n_j} u_i\right] = \frac{1}{n_j} E\left[\sum_{i=1}^{n_j} u_i\right] = \frac{1}{n_j} \sum_{i=1}^{n_j} E[u_i] = 0 \quad \forall j = 1, \dots, J,$$

$$\begin{aligned} \text{VAR}[\bar{u}_j] &= E[\bar{u}_j^2] = E\left[\left(\frac{1}{n_j} \sum_{i=1}^{n_j} u_i\right)^2\right] = \frac{1}{n_j^2} E\left[\left(\sum_{i=1}^{n_j} u_i\right)^2\right] = \\ &= \frac{1}{n_j^2} \sum_{i=1}^{n_j} E[u_i^2] = \frac{1}{n_j^2} n_j \sigma_u^2 = \frac{1}{n_j} \sigma_u^2 \quad \forall j = 1, \dots, J. \end{aligned}$$

Como podemos apreciar, la varianza del nuevo término de perturbación es inversamente proporcional al número de individuos u observaciones de cada una de las submuestras o grupos. En consecuencia, si el número de individuos de cada uno de los grupos es diferente, tendremos un término de perturbación heteroscedástico, debido a que la varianza no será constante para todas las observaciones (grupos en este caso).

Nota: en la expresión de la varianza tenemos en cuenta que
 $E[u_i u_j] = 0 \quad \forall i \neq j.$

En este caso, la matriz de varianzas y covarianzas del término de perturbación será igual a:

$$\text{VAR}[\bar{u}] = \begin{bmatrix} \frac{\sigma_u^2}{n_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma_u^2}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma_u^2}{n_j} \end{bmatrix} = \sigma_u^2 \begin{bmatrix} \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_j} \end{bmatrix} = \sigma_u^2 \Omega_N.$$

Entonces, la matriz de varianzas y covarianzas no será escalar, ya que los elementos de su diagonal principales son diferentes entre sí*.

* La matriz de varianzas y covarianzas sólo sería escalar si todos los subgrupos tuvieran el mismo número de individuos.

- En caso de que trabajemos con **datos agregados**, es decir, cuando se dispone de la suma de los valores adoptados por las variables de interés de los individuos, u observaciones de las submuestras o grupos, la varianza del nuevo término de perturbación será proporcional al número de individuos u observaciones de cada uno de los grupos.

Varianza del término de perturbación cuando trabajamos con datos agregados

Tenemos que plantear el problema de la heteroscedasticidad en términos similares a lo que hemos hecho para el caso de los datos en medias. Así pues, si partimos del modelo visto en el ejemplo de heteroscedasticidad, cuando trabajamos con datos en valores medios de los datos iniciales, pero la información para los diferentes individuos no está disponible y sólo tenemos acceso a los datos agregados, el modelo que podemos utilizar será el siguiente:

$$Y_j^a = \beta_1^* + \beta_2^* X_j^a + u_j^a \quad \forall j = 1, \dots, J,$$

dónde Y_j^a y X_j^a son el valor agregado de la variable endógena y de la variable exógena para cada grupo:

$$Y_j^a = \sum_{i=1}^{n_j} Y_i \quad \forall j = 1, \dots, J,$$

$$X_j^a = \sum_{i=1}^{n_j} X_i \quad \forall j = 1, \dots, J.$$

De manera análoga, el término de perturbación del modelo considerado, u_j , será la suma de los n_j términos de perturbación de cada uno de los grupos:

$$u_j^a = \sum_{i=1}^{n_j} u_i \quad \forall j = 1, \dots, J.$$

El valor esperado y la varianza del nuevo término de perturbación serán:

$$E[u_j^a] = E\left[\sum_{i=1}^{n_j} u_i\right] = E\left[\sum_{i=1}^{n_j} u_i\right] = \sum_{i=1}^{n_j} E[u_i] = 0 \quad \forall j = 1, \dots, J,$$

$$\text{VAR}[u_j^a] = E\left[\left(u_j^a\right)^2\right] = E\left[\left(\sum_{i=1}^{n_j} u_i\right)^2\right] = E\left[\left(\sum_{i=1}^{n_j} u_i\right)^2\right] =$$

$$= \sum_{i=1}^{n_j} E[u_i^2] = n_j \sigma_u^2 \quad \forall j = 1, \dots, J.$$

Nota: en la expresión de la varianza hay que tener en cuenta que $E[u_i u_j] = 0 \quad \forall i \neq j$.

Por lo tanto, la matriz de varianzas y covarianzas del nuevo término de perturbación será:

$$\text{VAR}[u^a] = \begin{bmatrix} n_1 \sigma_u^2 & 0 & \dots & 0 \\ 0 & n_1 \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_j \sigma_u^2 \end{bmatrix} \quad \begin{bmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_j \end{bmatrix}$$

- c) En cuanto a las causas que hemos denominado **causas espurias**, éstas se producen cuando podemos detectar la presencia de heteroscedasticidad en el modelo, pero la que se debe al incumplimiento de otros supuestos básicos del modelo de regresión.

Ejemplo de heteroscedasticidad por causas espurias

La especificación errónea del modelo provoca la aparición de heteroscedasticidad en el modelo de regresión, ya que, si omitimos una variable relevante, el término de perturbación del modelo especificado de forma errónea es heteroscedástico al captar el efecto de la variable omitida. Supongamos que el modelo correctamente especificado es el siguiente:

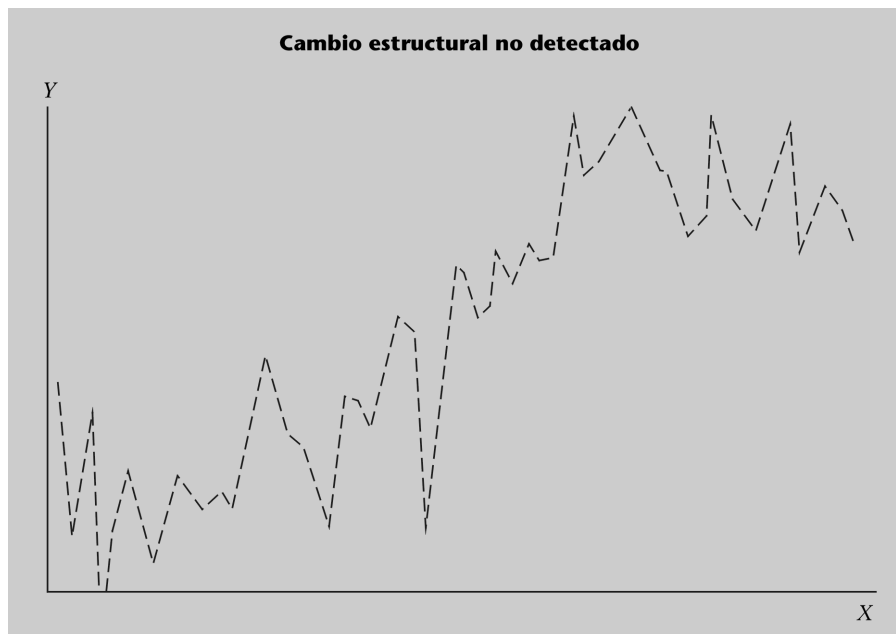
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad \forall i = 1, \dots, N,$$

pero, el investigador ha cometido un error y especifica el siguiente modelo, en el que se ha omitido una de las dos variables explicativas:

$$Y_i = \beta_1 + \beta_2 X_{2i} + v_i \quad \forall i = 1, \dots, N,$$

donde el nuevo término de perturbación será igual a $v_i = u_i + \beta_3 X_{3i}$. Por lo tanto, la varianza de v_i será igual a $\text{VAR}[v_i] = \sigma_u^2 + \beta_3^2 X_{3i}^2$, que no será constante y dependerá de la evolución de la variable omitida.

Asimismo, podemos llegar a detectar la presencia de heteroscedasticidad si hay un cambio estructural no detectado en el modelo analizado, tal y como se muestra en el gráfico que encontramos a continuación:



Tras haber visto en qué consiste la heteroscedasticidad y los posibles motivos que pueden causar su aparición en el modelo de regresión, analizaremos las consecuencias de la estimación por MCO en presencia de heteroscedasticidad.

2.2. Consecuencias de la estimación por MCO

La heteroscedasticidad es un caso particular del que hemos estudiado con respecto a la presencia de perturbaciones no esféricas en el modelo de regresión. Así pues, ya conocemos las consecuencias del uso de los estimadores MCO en presencia de perturbaciones heteroscedásticas. !

Los **estimadores MCO de B**, tal y como vimos, serán lineales, no sesgados, consistentes en ECM, pero ineficientes, es decir, ya no son los estimadores de varianza mínima. Existen otros estimadores que tienen una menor varianza en la familia de los estimadores lineales y no sesgados. !


Ved el caso de presencia de perturbaciones no esféricas en el modelo en el subapartado 1.1.1 de este módulo didáctico. !


Además, no podemos dejar de tener presente que con perturbaciones no esféricas, en este caso heteroscedásticas, para obtener la matriz de varianzas y covarianzas del estimadores MCO de \mathbf{B} , no podemos utilizar la expresión:

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}, \tag{2.3}$$


que dedujimos en otra asignatura, porque estaríamos subestimando la varianza de los estimadores MCO, ya que, tal y como vimos en el apartado anterior, la matriz de varianzas y covarianzas de los estimadores MCO con perturbaciones heteroscedásticas es:

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{2.4}$$


Como hemos estudiado, el método adecuado de estimación en presencia de perturbaciones no esféricas es el de MCG. Mediante este método obtendremos estimadores lineales, no sesgados, eficientes y consistentes en ECM. Para utilizar este método, es necesario conocer la matriz $\mathbf{\Omega}$ con el fin de aplicar cualquiera de las expresiones conocidas de este estimador MCG: 


 Ved la expresión de la matriz de varianzas y covarianzas de los estimadores MCO de \mathbf{B} en el subapartado 2.3.2 del módulo "Modelo de regresión lineal múltiple...".

Método MCG	
Aproximación directa	Aproximación en dos etapa
$\hat{\mathbf{B}}_{\text{MCG}} = (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{Y}$	$\hat{\mathbf{B}}_{\text{MCG}} = (\mathbf{X}^*\mathbf{X}^*)^{-1}\mathbf{X}^*\mathbf{Y}^*$
$\text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] = \sigma^2(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}$	$\text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] = \sigma^2(\mathbf{X}^*\mathbf{X}^*)^{-1}$
	$\mathbf{Y}^* = \mathbf{P}^{-1}\mathbf{Y}$
	$\mathbf{X}^* = \mathbf{P}^{-1}\mathbf{X}$
	$\mathbf{\Omega} = \mathbf{P}\mathbf{P}'$


Esto supone un problema añadido, ya que, excepto en caso de que la heteroscedasticidad tenga causas muestrales, es decir, que esté motivada por la manera de registrar los datos, desconocemos la forma de la matriz $\mathbf{\Omega}$ y tendríamos que estimarla, lo cual supone estimar N parámetros o elementos desconocidos adicionales*. 

* Estos parámetros son los parámetros desconocidos adicionales que aparecen en la expresión 2.2.

Por este motivo tendremos que hacer supuestos o hipótesis sobre la forma funcional de la heteroscedasticidad. En el apartado siguiente analizaremos los supuestos o hipótesis que se utilizan con mayor frecuencia para especificar la forma funcional de la heteroscedasticidad. Más adelante veremos cómo podremos detectar la presencia de la heteroscedasticidad en el modelo y cómo podremos proceder para estimar $\mathbf{\Omega}$. 

 Ved cómo podemos detectar la presencia de la heteroscedasticidad y cómo hay que estimar $\mathbf{\Omega}$ de la forma debida en los subapartados 2.4 y 2.5 del presente módulo didáctico.

2.3. Esquemas de dependencia funcional de la varianza

En este subapartado veremos las hipótesis o **supuestos en torno a la evolución de la varianza del término de perturbación** formuladas con mayor frecuencia. Son las siguientes: 

1) La varianza del término de perturbación es directamente proporcional a una variable Z . Esta variable puede ser una variable exógena, o *explicativa*, o una variable que no esté presente en el modelo:

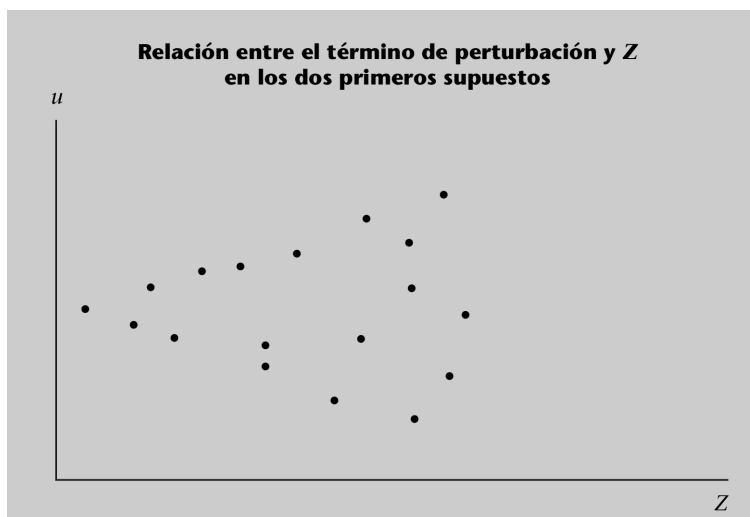
$$\text{VAR}[u_i] = \delta Z_i \quad \delta > 0 \quad \forall i = 1, \dots, N. \quad (2.5)$$

2) La varianza del término de perturbación es directamente proporcional al cuadrado de una variable Z :

$$\text{VAR}[u_i] = \delta Z_i^2 \quad \delta > 0 \quad \forall i = 1, \dots, N. \quad (2.6)$$

El segundo supuesto garantiza la existencia de varianzas positivas para cualquier valor que pueda adoptar la variable explicativa Z , situación que no se garantiza en el primer supuesto.

La relación entre el término de perturbación y la variable Z , en el primer y segundo supuestos, se puede representar mediante la gráfica que vemos a continuación:



En dicha gráfica podemos apreciar que, a medida que se incrementan los valores adoptados por la variable Z , la variabilidad del término de perturbación también aumenta.

3) La varianza del término de perturbación es inversamente proporcional a una variable Z :

$$\text{VAR}[u_i] = \delta \frac{1}{Z_i} \quad \delta > 0 \quad \forall i = 1, \dots, N. \quad (2.7)$$

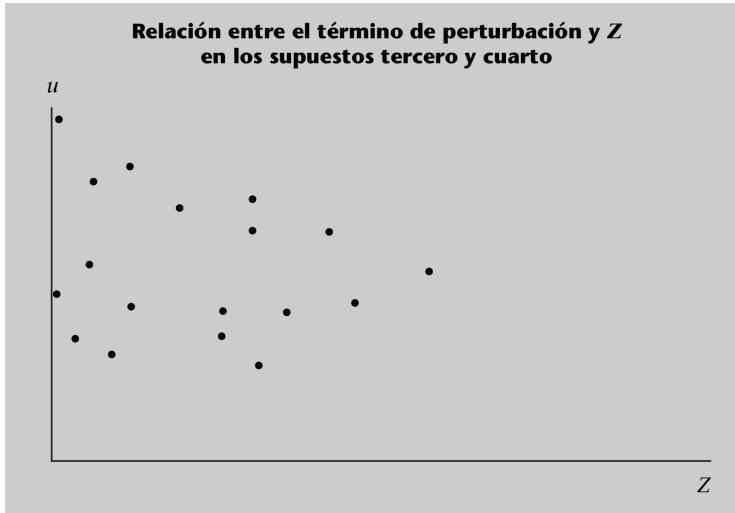
4) La varianza del término de perturbación es inversamente proporcional al cuadrado de la variable Z :

$$\text{VAR}[u_i] = \delta \frac{1}{Z_i^2} \quad \delta > 0 \quad \forall i = 1, \dots, N. \quad (2.8)$$

De la misma manera...

... que en los supuestos 1 y 2, el supuesto 4 garantiza la existencia de varianzas positivas para cualquier valor de Z , algo que no sucede en el supuesto 3.

La relación entre el término de perturbación y la variable Z , en el tercer y cuarto supuestos, se puede representar mediante la gráfica de la página siguiente:



En el gráfico que acabamos de ver podemos apreciar que, a medida que incrementan los valores adoptados por la variable Z , la variabilidad del término de perturbación disminuye.

5) La varianza del término de perturbación es una función lineal de la variable Z . Este supuesto se conoce con el nombre de **heteroscedasticidad aditiva**:

$$\text{VAR}[u_i] = \delta_0 + \delta_1 Z_i \quad \delta_0 > 0 \text{ y } \delta_1 > 0 \quad \forall i = 1, \dots, N. \quad (2.9)$$

6) La varianza del término de perturbación es una función exponencial de la variable Z . Esta hipótesis se conoce con el nombre de **heteroscedasticidad multiplicativa**:

$$\text{VAR}[u_i] = \exp[\delta_0 + \delta_1 Z_i] \quad \forall i = 1, \dots, N. \quad (2.10)$$

Podemos generalizar los supuestos quinto y sexto en un número mayor de variables, de manera que la evolución de la varianza del término de perturbación no dependa únicamente de una sola variable, sino que evolucione en función de un conjunto de variables, como se muestra en la siguiente tabla:

Generalización en un conjunto de variables de los supuestos aditivo y multiplicativo	
Heteroscedasticidad	Evolución de la varianza del término de perturbación
Aditiva	$\text{VAR}[u_i] = \delta_0 + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \dots + \delta_j Z_{ji} \quad \forall i = 1, \dots, N$
Multiplicativa	$\text{VAR}[u_i] = e^{\delta_0 + \delta_1 Z_{1i} + \delta_2 Z_{2i} + \dots + \delta_j Z_{ji}} \quad \forall i = 1, \dots, N$

En un principio hemos comentado que la variable Z o el conjunto de variables en función de las cuales evoluciona la varianza del término de perturbación pueden ser variables exógenas del modelo o variables ajenas a éste. Por norma general, en cualquier caso, supondremos que las variables causantes de la evolución de la varianza pertenecen al conjunto de regresores* del modelo analizado. Es decir, supondremos que la variable o las variables que causan la heteroscedasticidad de los modelos se encuentran entre los regresores.

Los supuestos del 1 al 4,...

... tal y como podemos apreciar, no son más que casos particulares de los casos 5 y 6, después de realizar un cambio de variable y una reparametrización.

* Los regresores son las variables explicativas del modelo.

Y lo haremos de esta manera, porque suponer que la varianza del término de perturbación evoluciona en función de variables ajenas al modelo sería un indicio de que nos encontramos ante un problema de omisión de variables relevantes.

! Ved la omisión de variables relevantes como a causa de heteroscedasticidad en el subapartado 2.1 de este módulo didáctico.

Tras haber visto los posibles esquemas de heteroscedasticidad, estudiaremos cómo podemos detectar la presencia de heteroscedasticidad en un modelo de regresión mediante métodos gráficos y contrastes. !

2.4. Detección de heteroscedasticidad

Para detectar la presencia de heteroscedasticidad en el MRLM, utilizaremos, fundamentalmente, dos tipos de métodos: los **métodos gráficos** y lo que conocemos como **contrastos**.

Entre estos dos tipos de métodos, los más fiables son los contrastes, puesto que, tal y como veremos a continuación, con los primeros sólo podemos obtener indicios de la existencia de heteroscedasticidad, pero no podemos decir con certeza si una variable o un conjunto de variables son los causantes de la existencia de ésta. En primer lugar, analizaremos los métodos gráficos y, a continuación, los contrastes. !

2.4.1. Métodos gráficos

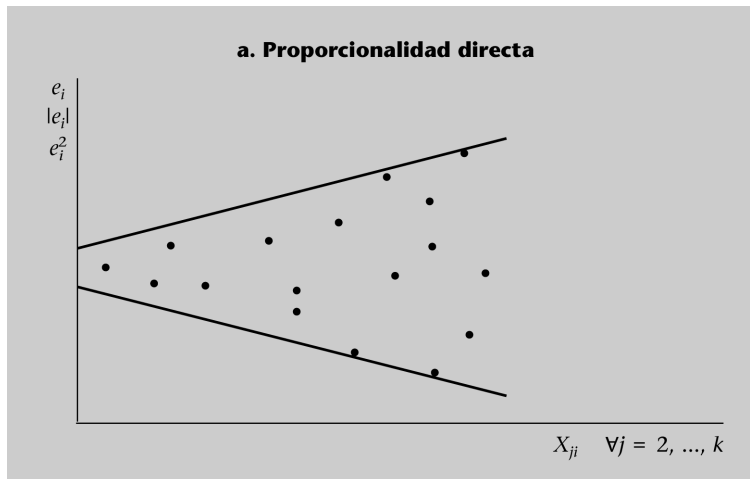
Los métodos gráficos consisten en el análisis gráfico de los residuos obtenidos a partir de la estimación por MCO del modelo analizado.

Los métodos gráficos están basados en el hecho de que los residuos MCO son estimadores consistentes de los términos de perturbación; por lo tanto, si hay problemas de heteroscedasticidad en el término de perturbación del modelo, podemos detectarlos en los residuos.

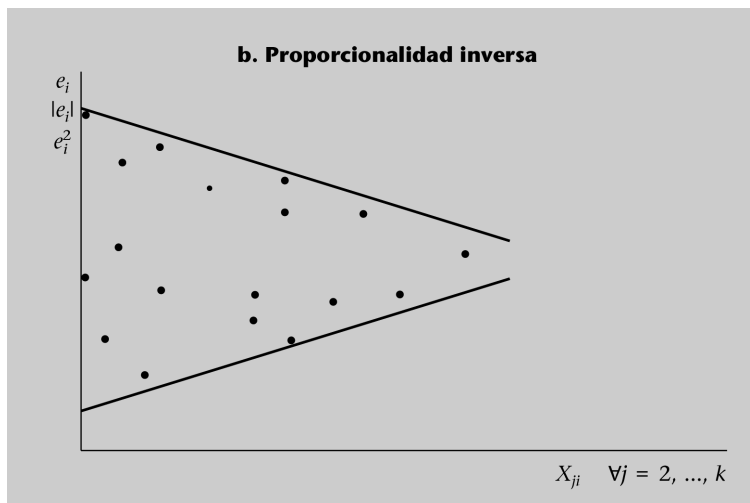
En concreto, estos métodos consisten en representar gráficamente los residuos o diferentes transformaciones de los últimos (el valor absoluto o el cuadrado) en función de cada una de las variables exógenas, o *regresores*, del modelo. Si la variabilidad de los residuos, del valor absoluto o del cuadrado crece o decrece, será un indicio de que la variable exógena seleccionada es la causante de la heteroscedasticidad del modelo.

En los gráficos que presentamos a continuación aparecen registradas algunas de las representaciones gráficas típicas que permiten reconocer una variable como la causante de la heteroscedasticidad:

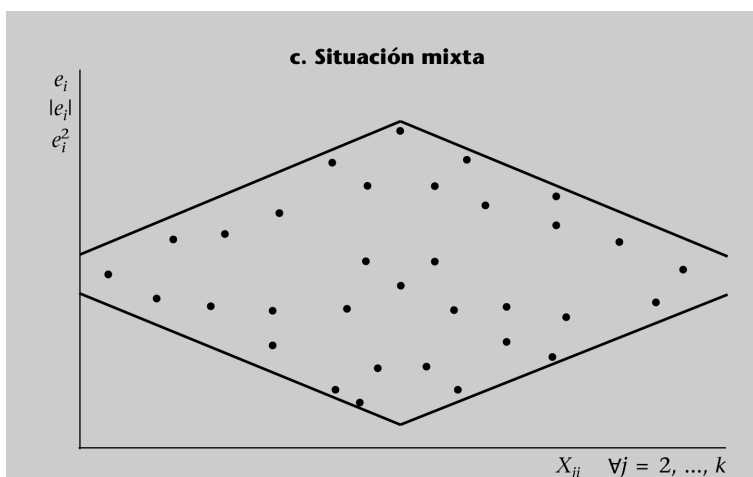
a) En el gráfico de la página siguiente, nos encontraríamos con que la variabilidad del residuo es directamente proporcional a la evolución de la variable exógena:




b) En este otro gráfico aparece la situación contraria, es decir, cuando la variabilidad de los residuos es inversamente proporcional a la evolución de la variable exógena:



c) Para acabar, en el último gráfico que presentamos se muestra una situación que es el compendio de las otras dos: en un principio, la variabilidad de los residuos crece a medida que lo hace la variable, pero, a partir de un punto determinado, la variabilidad de los residuos se vuelve inversamente proporcional a la variable exógena.



2.4.2. Contrastes de heteroscedasticidad

Ahora nos disponemos a estudiar los contrastes principales para detectar la presencia de heteroscedasticidad en el MRLM. Debemos tener en cuenta que no hay un contraste universal para la heteroscedasticidad que nos permita saber con absoluta certeza si el modelo presenta o no presenta problemas de este tipo, puesto que los diferentes rasgos de heteroscedasticidad contrastan la presencia de un determinado tipo de heteroscedasticidad. 

Todos los **contrastos de heteroscedasticidad** que analizaremos tienen las mismas hipótesis nulas y alternativas. En todos ellos, con la hipótesis nula, el modelo será homoscedástico, mientras que, con la hipótesis alternativa, el modelo tendrá problemas de heteroscedasticidad; es decir:

$$\left. \begin{array}{l} \bullet H_0 : \text{VAR}[u_i] = \sigma_u^2 \Rightarrow \text{homoscedasticidad.} \\ \bullet H_A : \text{VAR}[u_i] = \sigma_{u_i}^2 \Rightarrow \text{heteroscedasticidad.} \end{array} \right\} \quad (2.11)$$

Los contrastes que estudiaremos son el de Breusch-Pagan, Golfeld-Quandt, Glesjer y White; veremos sus características principales, cómo utilizarlos y las limitaciones que presentan.

Test de Breusch-Pagan

El test propuesto por Breusch y Pagan es un contraste muy general que permite cubrir un amplio espectro de situaciones. Este test supone que la estructura de heteroscedasticidad presente en el modelo es la siguiente:

$$\text{VAR}[u_i] = \sigma_{u_i}^2 = h(\mathbf{Z}_i' \boldsymbol{\alpha}), \quad (2.12)$$

en el que h es la forma funcional, que supondremos lineal sin pérdida de generalidad, \mathbf{Z}_i , un vector de variables de dimensión $(p + 1)$ en el que se registran la observación i -ésima de las p variables causantes de la heteroscedasticidad, y $\boldsymbol{\alpha}$ es otro vector de orden $(p + 1)$, en el que registraremos los parámetros asociados a cada una de las variables \mathbf{Z}_i :

$$\mathbf{Z}_i = \begin{bmatrix} 1 \\ Z_{1i} \\ Z_{2i} \\ \vdots \\ Z_{pi} \end{bmatrix}; \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}.$$

Por lo tanto, si la forma funcional es lineal, tendremos:

$$\text{VAR}[u_i] = \mathbf{Z}_i' \boldsymbol{\alpha} = (1 \ Z_{1i} \ Z_{2i} \ \dots \ Z_{pi}) \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \dots + \alpha_p Z_{pi}. \quad (2.13)$$

Si escribimos la expresión anterior para todas las observaciones, obtendremos:

$$\begin{aligned}
 \text{VAR}[u_1] &= \sigma_{u_1}^2 = \alpha_0 + \alpha_1 Z_{11} + \alpha_2 Z_{21} + \alpha_3 Z_{31} + \dots + \alpha_p Z_{p1}. \\
 \text{VAR}[u_2] &= \sigma_{u_2}^2 = \alpha_0 + \alpha_1 Z_{12} + \alpha_2 Z_{22} + \alpha_3 Z_{32} + \dots + \alpha_p Z_{p2}. \\
 \text{VAR}[u_3] &= \sigma_{u_3}^2 = \alpha_0 + \alpha_1 Z_{13} + \alpha_2 Z_{23} + \alpha_3 Z_{33} + \dots + \alpha_p Z_{p3}. \\
 &\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \ddots \quad \quad \quad \vdots \\
 \text{VAR}[u_N] &= \sigma_{u_N}^2 = \alpha_0 + \alpha_1 Z_{1N} + \alpha_2 Z_{2N} + \alpha_3 Z_{3N} + \dots + \alpha_p Z_{pN}.
 \end{aligned} \tag{2.14}$$


Y supondremos que tenemos varianzas diferentes para cada uno de los términos de perturbación y que la evolución de éstas depende de un conjunto de variables Z_{ji} .


Así pues, si disponemos de información para estimar 2.14, para saber si hay heteroscedasticidad, sólo tendremos que realizar el contraste de significación global de parámetros, ya que:

$$\text{Si } \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 \Rightarrow \sigma_{u_i}^2 = \alpha_0. \tag{2.15}$$

Entonces, cuando hacemos el contraste, se dará uno de los dos casos siguientes:

- No rechazar la hipótesis nula $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \dots = \alpha_p = 0$ supondría que la varianza del término de perturbación es constante e igual a α_0 , lo cual quiere decir que el modelo original es homoscedástico.
- Por otra parte, si rechazamos la hipótesis nula, la varianza del término de perturbación dependerá de las variables Z_j , no será constante y el modelo será heteroscedástico.

El problema consiste en el hecho de que desconocemos la evolución de la varianza del término de perturbación ($\sigma_{u_i}^2$), con lo que las relaciones 2.14 no son estimables y, por tanto, no podemos llevar a cabo el contraste. Una solución es utilizar una aproximación de la evolución de la varianza del término de perturbación ($\sigma_{u_i}^2$). En este caso, la solución propuesta por Breusch y Pagan consiste en utilizar los residuos normalizados al cuadrado, tal y como veremos. 

Para llevar a cabo el **contraste de Breusch-Pagan**, tenemos que seguir las siguientes etapas: 

1) Estimar el modelo original $Y = XB + U$ por MCO, y obtener los residuos MCO (es decir, e_i).

2) Calcular el residuo normalizado, $\frac{e_i}{\hat{\sigma}_u}$, para cada observación, en la que $\hat{\sigma}_u = \sqrt{\frac{e'e}{N}}$ es la raíz cuadrada del estimador MV de la varianza del término de perturbación.

3) Elevar al cuadrado los residuos normalizados obtenidos en el paso anterior, $\frac{e_i^2}{\hat{\sigma}_u^2}$. Ésta será la aproximación que utilizaremos para captar la evolución de la varianza del término de perturbación ($\sigma_{u_i}^2$).

4) Estimar por MCO la regresión auxiliar que vemos a continuación:

$$\frac{e_i^2}{\hat{\sigma}_u^2} = \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i} + \alpha_3 Z_{3i} + \dots + \alpha_p Z_{pi} + v_i. \quad (2.16)$$

5) Obtener la *SCE* de la regresión anterior. Según la hipótesis nula, la *SCE/2* se distribuye como una χ -cuadrado con p grados de libertad:


$$\frac{SCE}{2} \sim \chi_p^2. \quad (2.17)$$

El parámetro p ...


... es el número de regresores, o *variables explicativas*, de la evolución de la varianza del término de perturbación en 2.16.

La regla de decisión que utilizaremos al hacer el **contraste de Breusch-Pagan** será la siguiente:

- $\frac{SCE}{2} \geq \chi_{p,\alpha}^2 \Rightarrow$ rechazamos la hipótesis nula \Rightarrow no hay homoscedasticidad.
- $\frac{SCE}{2} < \chi_{p,\alpha}^2 \Rightarrow$ no rechazamos la hipótesis nula \Rightarrow hay homoscedasticidad. (2.18)

La interpretación de esta regla de decisión es ésta: si U es homoscedástico, el conjunto de variables registradas en el vector Z_i no tendrán ningún poder explicativo en cuanto a los residuos normalizados; por tanto, la *SCE* será pequeña, el estadístico de prueba también será pequeño y menor que el valor de las tablas de la $\chi_{p,\alpha}^2$. Pero, si nos hallamos en el caso contrario, y la varianza de los términos de perturbación depende del conjunto de variables registradas en Z_i , una parte importante de la variabilidad total de los residuos normalizados al cuadrado será explicada por Z_i ; así pues, la *SCE* y el estadístico de prueba serán grandes y éste último será mayor que el valor de tablas. 

Si finalmente concluimos que hay heteroscedasticidad en el modelo, para poder aplicar MCP tendremos que multiplicar las observaciones de las variables por $\frac{1}{\sqrt{Z_i' \hat{\alpha}}}$, siendo $\hat{\alpha}$ la estimación obtenida en 2.16.

Para finalizar con los aspectos relacionados con el contraste de Breusch-Pagan, únicamente nos queda por comentar las limitaciones de éste, que podemos resumir en los siguientes puntos: 

- Este contraste sólo es estrictamente válido cuando trabajamos con muestras grandes, ya que *SCE/2* solo se distribuye como una χ_p^2 asintóticamente.
- Puede no detectar heteroscedasticidad en el modelo (aunque el término de perturbación sea heteroscedástico) si las variables registradas en Z_i no son sus causantes.

Ejemplo de contraste de Breusch-Pagan

Se ha estimado el siguiente modelo de regresión con dos variables explicativas y un término independiente a partir de una muestra de $N = 100$ observaciones:

$$\hat{Y}_i = 3814,821 - 8,321X_{2i} - 1340,062X_{3i}$$

donde los valores de la t de Student correspondientes a cada uno de los coeficientes estimados son 4.727, 997, 17, 204 y 1.888, 424.

Nuestra intención es comprobar si este modelo tiene problemas de heteroscedasticidad. Para alcanzar este objetivo de contrastar la existencia de heteroscedasticidad, hemos utilizado el contraste de Breusch-Pagan, y hemos obtenido los resultados que mostramos a continuación:

- $\frac{\hat{e}_i^2}{\hat{\sigma}_{u_{MV}}^2} = -0,081 + 0,044X_{2i} - 0,465X_{3i}$.
- $SCE = 162,237$.
- $\frac{SCR}{2} = 81,1185 > \chi_{2,0,05}^2 = 5,99$.

Al plantear la regresión auxiliar anterior, se supone que todos los regresores son los causantes de la heteroscedasticidad en el modelo.

Como bien podemos apreciar, el estadístico de prueba es superior al valor de tablas de una χ -cuadrado; por lo tanto, en función de la evidencia empírica disponible, rechazaríamos la hipótesis nula, es decir, que el modelo original es homoscedástico.

A continuación, presentamos los mismos resultados, pero suponiendo que el conjunto de variables causantes de la heteroscedasticidad se redujese únicamente a cada uno de los regresores por separado:

1) Primer regresor, X_{2i} :

- $\frac{\hat{e}_i^2}{\hat{\sigma}_{u_{MV}}^2} = -1,218 + 0,044X_{2i}$.
- $SCE = 160,746$.
- $\frac{SCE}{2} = 80,373 > \chi_{1,0,05}^2 = 3,84$.

2) Segundo regresor, X_{3i} :


- $\frac{\hat{e}_i^2}{\hat{\sigma}_{u_{MV}}^2} = 1,560 - 0,227X_{3i}$.
- $SCE = 0,356$.
- $\frac{SCE}{2} = 0,178 < \chi_{1,0,05}^2 = 3,84$.

Podemos apreciar que, en el supuesto de que la variable causante de la heteroscedasticidad es X_2 , rechazamos la hipótesis nula de que el término de perturbación es homoscedástico. Por otra parte, cuando suponemos que la variable causante de la heteroscedasticidad es X_3 , la hipótesis nula no es rechazada.

Así pues, podríamos concluir, a partir la información anterior, que la heteroscedasticidad del término de perturbación viene provocada por la variable X_2 .

En apartados sucesivos iremos presentando los resultados obtenidos por este mismo modelo utilizando los diferentes contrastes que iremos explicando, y veremos, entonces, qué conclusiones podemos obtener.

Test de Golfeld y Quandt

A diferencia del test anterior, este test es válido para muestras pequeñas y cuando podemos atribuir la presencia de la heteroscedasticidad a una única variable. Tal y como veremos, se trata de un contraste muy sencillo. 

Este contraste resulta adecuado cuando sospechamos que la varianza del término de perturbación está directa o inversamente relacionada con una variable explicativa.

Para estudiar en qué consiste este test suponemos que la varianza del término de perturbación está relacionada positivamente con uno de los regresores del modelo. Es decir, consideramos los siguientes esquemas de heteroscedasticidad:

- $\text{VAR}[u_i] = \delta X_i \quad \delta > 0.$
 - $\text{VAR}[u_i] = \delta X_i^2 \quad \delta > 0.$
- (2.19)

La situación formulada en 2.19 quedaría registrada en el gráfico que muestra la proporcionalidad directa entre la variabilidad de los residuos y la evolución de la variable exógena.

Ved el gráfico "a. Proporcionalidad directa" en el subapartado 2.4.1 de este módulo didáctico.

Teniendo en cuenta todo lo que acabamos de decir, este contraste es mucho más específico que el anterior y no nos permitirá cubrir un espectro tan amplio como el de Breusch-Pagan. Se trata, por tanto, de un contraste adecuado para contrastar la presencia de heteroscedasticidad en situaciones muy determinadas. !

Para llevar a cabo el contraste de **Golfeld y Quandt** en el supuesto que estamos considerando, tenemos que seguir una serie de pasos que acto seguido podremos ver: !

1) Reordenar las observaciones de todas las variables del modelo en orden creciente, de menor a mayor, según el orden que induce la variable explicativa que suponemos que es la causante de la heteroscedasticidad.

2) Eliminar C observaciones centrales partiendo la muestra en tres submuestras: una primera submuestra que contiene $(N - C) / 2$ observaciones, una segunda submuestra que contiene C observaciones y una tercera que contiene $(N - C) / 2$ observaciones.

En general, el número de observaciones centrales C que queremos eliminar se determina utilizando $C = N/3$ o bien $C = N/4$.

3) Estimar una regresión con la primera y tercera submuestras.

4) Obtener la SCR para cada una de las regresiones: SCR_1 y SCR_3 .

5) Calcular el siguiente estadístico de prueba:

$$G - Q = \frac{SCR_1}{SCR_3} \sim F_{\frac{N-C}{2}-k, \frac{N-C}{2}-k}. \quad (2.20)$$

Como podemos apreciar, el estadístico de prueba se distribuye con la hipótesis nula como una F de Snedecor con $[(N - C) / 2] - k$ en el numerador y en el denominador, es decir, con tantos grados de libertad como los que tiene cada una de las regresiones, a partir de los cuales hemos obtenido las sumas de cuadrados de los errores.

La regla de decisión que tenemos que utilizar al hacer el **contraste de Golfeld y Quandt** es:

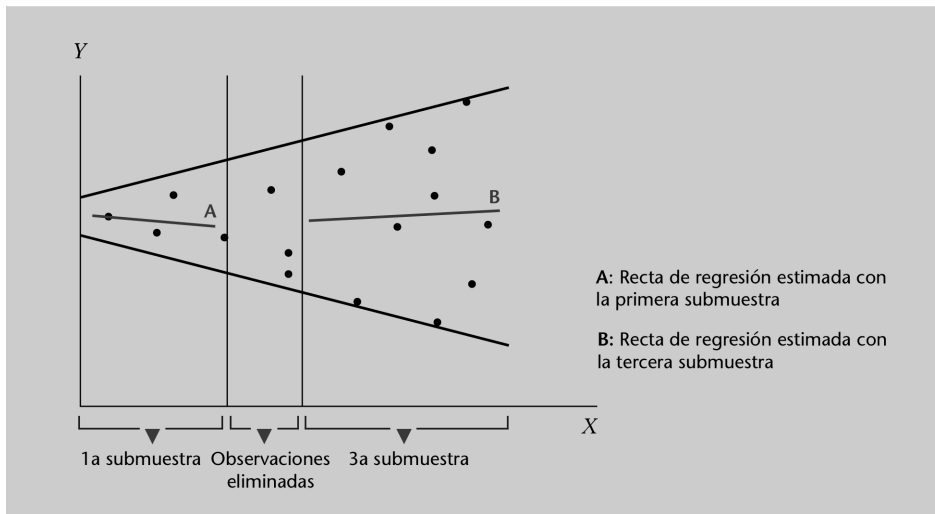
- $G - Q \geq F_{\frac{N-C}{2}-k, \frac{N-C}{2}-k; \alpha} \Rightarrow$ rechazamos $H_0 \Rightarrow$ heteroscedasticidad.
 - $G - Q < F_{\frac{N-C}{2}-k, \frac{N-C}{2}-k; \alpha} \Rightarrow$ no rechazamos $H_0 \Rightarrow$ homoscedasticidad.
- } (2.21)

La interpretación del contraste es sencilla. Si la varianza del término de perturbación es directamente proporcional a la evolución o al cuadrado de la evolución de la variable en función de la cual se han ordenado todas las variables integrantes del modelo, la variabilidad de los residuos en la tercera submuestra tiene que ser mucho mayor que la que se da para la primera; por tanto, la SCR_3 tiene

En la expresión 2.21...

... k es el número de parámetros del modelo de regresión inicial y α es el nivel de significación.

que ser mucho mayor que la SCR_1 . Entonces, el estadístico de prueba será grande y se encontrará por encima del valor de las tablas de la F .



Es importante que tengamos en cuenta que si en el supuesto que estamos considerando definiésemos $\frac{SCR_1}{SCR_3}$, es decir, el estadístico de prueba al revés, la hipótesis nula no se rechazaría nunca, atendiendo a que, en este caso, $SCR_1 < SCR_3$ y, en consecuencia, el estadístico de prueba sería menor que 1 y los valores críticos de la F son siempre superiores a 1. !

Las limitaciones de este contraste son las siguientes: !

- La potencia del test depende, en gran medida, del número C de observaciones centrales que eliminemos. Si C es grande, la primera y tercera submuestras tendrán un número de observaciones reducidas y, por lo tanto, las estimaciones serán poco fiables. Pero, si el número C es reducido, las submuestras primera y tercera serán muy parecidas, y, de esta manera, daremos primacía a la hipótesis de homoscedasticidad.
- Si la heteroscedasticidad es del tipo $VAR[u_i] = \delta \frac{1}{X_{ji}}$ o $VAR[u_i] = \delta \frac{1}{X_{ji}^2}$, es decir, como la que aparece en el gráfico que muestra la proporcionalidad inversa entre la variabilidad de los residuos y la evolución de la variable exógena. Si realizamos el contraste tal y como hemos explicado previamente, nunca rechazaremos la hipótesis nula, ya que $SCR_1 < SCR_3$. Así pues, con una heteroscedasticidad de este tipo nunca rechazaríamos la hipótesis de homoscedasticidad. Para esta última situación hay dos soluciones posibles, y son éstas:
 - Calcular el estadístico de prueba al revés: $\frac{SCR_3}{SCR_1}$.
 - Ordenar las observaciones en orden decreciente, de mayor a menor y utilizar como estadístico de prueba $\frac{SCR_3}{SCR_1}$.
- Este contraste no es válido para cierto tipo de heteroscedasticidad, en caso de que ésta venga provocada por una única variable, si nos encontramos en una

Ved el gráfico "b. Proporcionalidad inversa" en el subapartado 2.4.1 del presente módulo didáctico. !

situación como la que registra el gráfico que muestra una situación mixta de los dos casos anteriores, puesto que, al eliminar las C observaciones centrales, nos quedamos con dos submuestras cuya variabilidad resulta muy parecida.

Vedel gráfico "c. Situación mixta" en el subapartado 2.4.1 de este módulo didáctico.

Ejemplo de contraste de Golfeld y Quandt

Si nos reencontramos con el ejemplo que nos ha servido de modelo en el contraste de Breusch-Pagan, veremos los resultados obtenidos con el contraste de Golfeld y Quandt. El valor de C (número de observaciones centrales que eliminaremos tras haber ordenado las variables en función de la variable causante de la heteroscedasticidad en el modelo, es decir, cada uno de los regresores X_2 y X_3) que hemos utilizado es 30.

Se hace el contraste con cada uno de los regresores por separado, con lo que obtenemos los siguientes resultados:

1) Los resultados que hemos obtenido cuando se supone que la variable causante de la heteroscedasticidad es X_2 son:

- $SCR_1 = 5.958.664,319$; $SCR_3 = 246.791.711,800$.
- $G - Q = \frac{246.791.711,800}{5.958.664,319} = 41,417 > F_{35-3,35-3;0,05} \approx 1,84$.

Al estar el estadístico de prueba por encima del valor de las tablas de la F de Snedecor, la hipótesis nula queda rechazada. Por lo tanto, podemos concluir que la variable X_2 es la causante de la heteroscedasticidad.

2) Cuando se supone que la variable causante de la heteroscedasticidad es X_3 , los resultados son:

- $SCR_1 = 746.699.976,220$; $SCR_3 = 1.261.841.755,372$.
- $G - Q = \frac{1.261.841.755,372}{746.699.976,220} = 1,690 < F_{35-3,35-3;0,05} \approx 1,84$.

Al hallarse el estadístico de prueba por debajo del valor de las tablas de la F de Snedecor, no se rechaza la hipótesis nula. Así pues, podemos concluir diciendo que la variable X_3 no es la variable causante de la heteroscedasticidad.

Test de Glesjer

Este contraste es muy similar al de Breusch-Pagan. Se supone que la forma funcional de la heteroscedasticidad es la siguiente*:

* En la expresión 2.22, el superíndice h expresa una potencia.

$$\text{VAR}[u_i] = \delta_0 + \delta_1 Z_i^h + v_i \quad \forall i = 1, \dots, N, \quad (2.22)$$


donde Z_i es la variable causante de la heteroscedasticidad. Si pudiésemos estimar 2.22 y contrastar la significación individual del parámetro δ_1 , es decir, contrastar la hipótesis nula $H_0 : \delta_1 = 0$, nos encontraríamos en las situaciones siguientes:

- Si rechazamos $H_0 \Rightarrow \delta_1 \neq 0 \Rightarrow$ hay heteroscedasticidad, teniendo en cuenta que:

$$\text{VAR}[u_i] = \hat{\delta}_0 + \hat{\delta}_1 Z_i^h \quad \forall i = 1, \dots, N.$$

- Si no rechazamos $H_0 \Rightarrow \delta_1 = 0 \Rightarrow$ hay homoscedasticidad, teniendo en cuenta que:

$$\text{VAR}[u_i] = \hat{\delta}_0 \quad \forall i = 1, \dots, N,$$

El problema consiste en el hecho de que $\text{VAR}[u_i]$ es desconocida y, por lo tanto, tenemos que utilizar una estimación. La solución propuesta por Glesjer es la siguiente: 

- 1) Estimar por MCO el modelo original.
- 2) Obtener el vector de residuos MCO, e .
- 3) Calcular los residuos en valor absoluto, que será la magnitud propuesta por Glesjer para aproximar la evolución de $\text{VAR}[u_i]$.
- 4) Estimar la siguiente regresión auxiliar:


$$|e_i| = \delta_0 + \delta_1 Z_i^h + w_i \quad \forall i = 1, \dots, N. \quad (2.23)$$

Glesjer propone utilizar los valores de h que vemos a continuación:

$$h = \{1, -1, 1/2, -1/2\}.$$

- 5) Contrastar la hipótesis nula $H_0 : \delta_1 = 0$ en 2.23 mediante el contraste de la t de Student para cada valor de h .

Este contraste es más particular que el de Breusch-Pagan, dado que las variables causantes de la heteroscedasticidad se reducen a una.

Como variable endógena de la regresión 2.23, también podemos utilizar e_i^2 , ya que el objetivo es tener una variable dependiente que aproxime la varianza del término de perturbación con un rango de valor mayor que cero o igual a cero. 

Ejemplo de contraste de Glesjer

Recuperando el ejemplo que hemos utilizado antes, a continuación analizaremos los resultados obtenidos cuando utilizamos el contraste de Glesjer. Aplicamos el test a cada uno de los regresores y obtenemos los siguientes resultados:

- 1) Los resultados, cuando se supone que la variable causante de la heteroscedasticidad es X_2 , son:
 - $|e_i| = -1.279,423 + 82,102X_{2i}$, con desviaciones estándares de 644,641 y 11,082, y valores para la t de Student de -1,985 y 7,408, respectivamente.
 - $|e_i| = 3.205,950 - 6.529,235X_{2i}^{-1}$, con desviaciones estándares de 428,919 y 3.354,432, y valores para la t de Student de 7,474 y -1,946, respectivamente.
 - $|e_i| = -3.400,916 + 933,432X_{2i}^{0,5}$, con desviaciones estándares de 1.019,941 y 143,526, y valores para la t de Student de -3,334 y 6,504, respectivamente.
 - $|e_i| = 4.659,906 - 9.646,098X_{2i}^{-0,5}$, con desviaciones estándares de 654,990 y 2.875,813, y valores para la t de Student de 7,114 y -3,354, respectivamente.

El valor de las tablas de una t de Student con 98 grados de libertad y al 5% de nivel de significación es aproximadamente 1,993. Por lo tanto, para todos los exponentes de X_2 se rechaza la hipótesis nula en el contraste de significación individual, excepto en el caso del exponente -1, en el que estamos muy próximos al valor crítico. A partir de la información obtenida en el contraste de Glesjer, rechazaríamos la hipótesis nula, consistente en el hecho de que el término de perturbación es homoscedástico y podemos suponer que la variable causante de la heteroscedasticidad es X_2 .

Notad que...

... en la expresión 2.23 hay cuatro regresiones:

- $|e_i| = \delta_0 + \delta_1 Z_i + w_i$.
- $|e_i| = \delta_0 + \delta_1 \frac{1}{Z_i} + w_i$.
- $|e_i| = \delta_0 + \delta_1 Z_i^{0,5} + w_i$.
- $|e_i| = \delta_0 + \delta_1 \frac{1}{Z_i^{0,5}} + w_i$.

2) Cuando se plantea el supuesto de que la variable causante de la heteroscedasticidad es X_3 , los resultados son:

- $|e_i| = 6.727,53 - 1.563,272X_{3i}$, con desviaciones estándares de 3.753,254 y 1.511,184, y valores para la t de Student de 1,792 y -1,034, respectivamente.
- $|e_i| = -1.418,353 + 10.462,592X_{3i}^{-1}$, con desviaciones estándares de 3.718,859 y 9.028,222, y valores para la t de Student de -0,381 y 1,159, respectivamente.
- $|e_i| = 10.784,39 - 5.045,368X_{3i}^{0,5}$, con desviaciones estándares de 7.455, 536 y 4.744, 143, y valores para la t de Student de 1, 446 y -1, 063, respectivamente.
- $|e_i| = -5.475,734 + 13.054,394X_{3i}^{-0,5}$, con desviaciones estándares de 7.420,995 y 11.595,832, y valores para la t de Student de -0,738 y 1,126, respectivamente.

En todas las regresiones anteriores no se rechaza la hipótesis nula en el contraste de significación individual del parámetro asociado a las transformaciones de la variable X_3 . Así pues, podemos llegar a la conclusión de que la variable X_3 no es la causante de la presencia de heteroscedasticidad.

Test de White

El último contraste que veremos para detectar la presencia de heteroscedasticidad en el MRLM es el que propuso White. El planteamiento del contraste es el mismo que el de Breusch-Pagan y Glesjer, es decir, intentar explicar la variabilidad del término de perturbación, representada por una expresión diferente en cada test, a partir de alguna variable o conjunto de variables explicativas del modelo. En este caso, el conjunto de variables se extiende a todos los regresores.

Para llevar a cabo este contraste tenemos que seguir los pasos que presentamos a continuación:

- 1) Estimar el modelo original por MCO.
- 2) Obtener los residuos MCO (e_i).
- 3) Estimar una regresión en la que la variable endógena sea los residuos MCO al cuadrado y las variables explicativas, los regresores del modelo original, los regresores al cuadrado y los productos cruzados dos a dos, es decir:

$$e_i^2 = f(1, X_{2i}, \dots, X_{ki}, X_{2i}^2, \dots, X_{ki}^2, X_{2i}X_{3i}, \dots, X_{k-1i}X_{ki}). \quad (2.24)$$

e_i^2 es la aproximación propuesta por White para registrar la evolución de $\text{VAR}[u_i]$.

- 4) Calcular el siguiente estadístico de prueba:

$$N \cdot R^2 \sim \chi_{p-1}^2, \quad (2.25)$$

en el que N y R^2 son el número de observaciones y el coeficiente de determinación de la regresión auxiliar 2.24, respectivamente. Como podemos apreciar, este estadístico se distribuye como una χ -cuadrado con $p - 1$ grados de libertad, donde p es el número de parámetros de la expresión 2.24. Asimismo, podemos calcular los grados de libertad de este estadístico utilizando $[k \times (k + 1) / 2] - 1$, en la que k es el número de parámetros del modelo original para el cual deseamos contrastar la presencia de heteroscedasticidad.

La regla de decisión a utilizar en el **contraste de White** es la siguiente:

- $N \cdot R^2 \geq \chi_{p-1;\alpha}^2 \Rightarrow$ rechazamos $H_0 \Rightarrow$ heteroscedasticidad.
 - $N \cdot R^2 < \chi_{p-1;\alpha}^2 \Rightarrow$ no rechazamos $H_0 \Rightarrow$ homoscedasticidad.
- (2.26)

La lógica del contraste es la siguiente: en este contraste se supone que la heteroscedasticidad viene provocada por todas las variables explicativas. Por lo tanto, si hay heteroscedasticidad, los regresores de 2.24 explicarán una parte importante de la variabilidad de los e_i^2 , el coeficiente de determinación será elevado, y el estadístico de prueba también será grande y se situará por encima del valor de las tablas. En cambio, si no hay heteroscedasticidad, los regresores de 2.24 carecerán de poder explicativo alguno sobre la variable endógena y el coeficiente de determinación será pequeño, así que, el valor del estadístico de prueba se situará por debajo del valor de las tablas.

Ejemplo de contraste de White

Recuperando el ejemplo desarrollado en los contrastes anteriores, y para finalizar, analizaremos los resultados obtenidos mediante el contraste de White, que presentamos a continuación:

- $\hat{\epsilon}_i^2 = 46.037.365,603 - 2.083.390,404X_{2i} - 123.333.588,523X_{3i} + 31.074,88X_{2i}^2 + 1.393.410,04X_{3i}^2 - 1.910,163X_{2i} \cdot X_{3i}$.
- $R^2 = 0,207$; $N = 100$.
- $N \cdot R^2 = 20,7 > \chi_{5;0,05}^2 = 11,1$.

En este contraste, también rechazamos la hipótesis nula, que consiste en el hecho de que el término de perturbación es homoscedástico. La contradicción entre la conclusión que se deriva de este texto y la que obteníamos de los anteriores reside en que, con este contraste, no sabemos cuál de las variables explicativas, o *regresores*, es la causante de la existencia de heteroscedasticidad en el modelo.

2.5. Estimación por mínimos cuadrados generalizados (MCG y MCP)

Tras haber visto algunos de los posibles esquemas de dependencia funcional de la varianza, las propiedades de los estimadores MCO y los métodos para detectar la heteroscedasticidad, podemos afrontar la estimación por MCG en un modelo con perturbaciones heteroscedásticas. Podemos llevar a cabo la estimación por MCG mediante las dos aproximaciones estudiadas, y cada aproximación supone las siguientes dificultades:

a) Para utilizar MCG tenemos que conocer la matriz Ω y, posteriormente, utilizar las expresiones que ya conocemos:

$$\left. \begin{aligned} \bullet \hat{\mathbf{B}}_{\text{MCG}} &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}. \\ \bullet \text{VAR}[\hat{\mathbf{B}}_{\text{MCG}}] &= \sigma^2(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}. \\ \bullet \hat{\sigma}_{\text{MCG}}^2 &= \frac{\mathbf{e}'\Omega^{-1}\mathbf{e}}{N-k}. \end{aligned} \right\} (2.27)$$

Nota: en las expresiones de 2.27, \mathbf{e} es el vector de residuos obtenidos a partir de la estimación MCG.

Si la matriz Ω es desconocida, tenemos que utilizar el supuesto que consideramos más adecuado en cuanto a la forma funcional de la heteroscedasticidad. Tendremos que basar esta decisión en los resultados obtenidos mediante los métodos de detección que hemos visto.

Podéis encontrar los métodos de detección de heteroscedasticidad en el subapartado 2.4 del presente módulo didáctico.

b) Para utilizar MCP (la aproximación en dos etapas) también tenemos que conocer la matriz Ω , con el fin de obtener, a partir de ésta, la matriz $T = P^{-1}$ de transformación. Esta matriz nos va a permitir pasar de un modelo original heteroscedástico a otro modelo homoscedástico.

En primer lugar, estudiaremos cuál será la matriz de transformación que hay que utilizar cuando la matriz Ω es conocida, es decir, cuando la heteroscedasticidad viene determinada por causas muestrales, y, posteriormente, analizaremos aquellos casos en los que es necesario realizar supuestos acerca de la forma funcional de la heteroscedasticidad.

En cuanto a las causas muestrales, nos encontramos con dos casos:

1) Un primer caso es el que se da cuando trabajamos con **datos en medias**. La varianza del término de perturbación será, entonces:

$$\text{VAR}[\bar{u}_j] = \sigma_u^2 \frac{1}{n_j} \quad \forall j = 1, \dots, J;$$

así pues, la matriz Ω tendrá la siguiente expresión:

$$\Omega = \begin{bmatrix} \frac{1}{n_1} & 0 & \dots & 0 \\ 0 & \frac{1}{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{n_j} \end{bmatrix}. \quad (2.28)$$

La matriz P , que cumple $\Omega = PP'$, tiene la expresión:

$$P = \begin{bmatrix} \frac{1}{\sqrt{n_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{n_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{n_j}} \end{bmatrix}. \quad (2.29)$$

Su inversa será la matriz que, cuando multiplicamos por las matrices del modelo, nos permitirá obtener un modelo homoscedástico:

$$\mathbf{P}^{-1} = \mathbf{T} = \begin{bmatrix} \sqrt{n_1} & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{n_j} \end{bmatrix} \quad (2.30)$$

Como podemos apreciar, la matriz \mathbf{P}^{-1} transformará la matriz \mathbf{X} y los vectores \mathbf{Y} y \mathbf{U} , dando lugar a unos nuevos vectores \mathbf{Y}^* y \mathbf{U}^* , y a una nueva matriz \mathbf{X}^* , en la que sus elementos serán el producto de los iniciales por la raíz cuadrada del número de observaciones del grupo correspondiente.

De esta manera, el nuevo término de perturbación (\bar{u}_j^*) será homoscedástico:

$$\text{VAR}[\bar{u}_j^*] = \text{VAR}[\sqrt{n_j}\bar{u}_j] = n_j \text{VAR}[\bar{u}_j] = n_j \frac{1}{n_j} \sigma_u^2 = \sigma_u^2. \quad (2.31)$$

Actuación de la matriz $\mathbf{T} = \mathbf{P}^{-1}$ sobre el modelo

Tal y como podemos apreciar a partir de lo que se hemos visto en el texto, la matriz $\mathbf{T} = \mathbf{P}^{-1}$ transforma la matriz y los vectores del modelo de la siguiente forma:

$$\mathbf{P}^{-1}\bar{\mathbf{Y}} = \begin{bmatrix} \sqrt{n_1} & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{n_j} \end{bmatrix} \begin{bmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_j \end{bmatrix} = \begin{bmatrix} \bar{Y}_1\sqrt{n_1} \\ \bar{Y}_2\sqrt{n_2} \\ \vdots \\ \bar{Y}_j\sqrt{n_j} \end{bmatrix},$$

$$\mathbf{P}^{-1}\bar{\mathbf{X}} = \begin{bmatrix} \sqrt{n_1} & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{n_j} \end{bmatrix} \begin{bmatrix} 1 & \bar{X}_{21} & \dots & \bar{X}_{k1} \\ 1 & \bar{X}_{22} & \dots & \bar{X}_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \bar{X}_{2j} & \dots & \bar{X}_{kN} \end{bmatrix} = \begin{bmatrix} 1\sqrt{n_1} & \bar{X}_{21}\sqrt{n_1} & \dots & \bar{X}_{k1}\sqrt{n_1} \\ 1\sqrt{n_2} & \bar{X}_{22}\sqrt{n_2} & \dots & \bar{X}_{k2}\sqrt{n_2} \\ \vdots & \vdots & \ddots & \vdots \\ 1\sqrt{n_j} & \bar{X}_{2j}\sqrt{n_j} & \dots & \bar{X}_{kN}\sqrt{n_j} \end{bmatrix},$$

$$\mathbf{P}^{-1}\bar{\mathbf{U}} = \begin{bmatrix} \sqrt{n_1} & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{n_j} \end{bmatrix} \begin{bmatrix} \bar{u}_1 \\ \bar{u}_2 \\ \vdots \\ \bar{u}_j \end{bmatrix} = \begin{bmatrix} \bar{u}_1\sqrt{n_1} \\ \bar{u}_2\sqrt{n_2} \\ \vdots \\ \bar{u}_j\sqrt{n_j} \end{bmatrix}.$$

2) De manera similar, un segundo caso se da cuando trabajamos con **datos agregados**. En este caso, sabemos que la varianza del término de perturbación está igual a $\text{VAR}[u_j^a] = n_j\sigma_u^2$. Entonces, la matriz Ω será:

$$\Omega = \begin{bmatrix} n_1 & 0 & \dots & 0 \\ 0 & n_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & n_j \end{bmatrix} \quad (2.32)$$

La matriz \mathbf{P} , que cumple $\Omega = \mathbf{P}\mathbf{P}'$, será:

$$\mathbf{P} = \begin{bmatrix} \sqrt{n_1} & 0 & \dots & 0 \\ 0 & \sqrt{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{n_j} \end{bmatrix} \quad (2.33)$$

Y su inversa será la matriz que, cuando multiplicamos la matriz y los valores del modelo, nos permitirá obtener un modelo homoscedástico:


$$\mathbf{P}^{-1} = \mathbf{T} = \begin{bmatrix} \frac{1}{\sqrt{n_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{n_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{n_j}} \end{bmatrix}. \quad (2.34)$$

De esta manera, el nuevo término de perturbación (u_j^{a*}) será homoscedástico, como podemos apreciar aquí:


$$\text{VAR}[u_j^{a*}] = \text{VAR}\left[\frac{1}{\sqrt{n_j}} u_j^a\right] = \frac{1}{n_j} \text{VAR}[u_j^a] = \frac{1}{n_j} n_j \sigma_u^2 = \sigma_u^2. \quad (2.35)$$

Después de los ejemplos en los que hemos visto cómo tenemos que obtener la matriz $\mathbf{T} = \mathbf{P}^{-1}$, podemos apreciar que, si la forma funcional de la varianza es

$$\text{VAR}[u_i] = \sigma_u^2 f(i),$$

obtenemos los nuevos elementos de las variables transformadas cuando multiplicamos los elementos iniciales de los vectores \mathbf{Y} y \mathbf{U} , y de la matriz \mathbf{X} , por la inversa de la raíz cuadrada de la función, es decir, $\frac{1}{\sqrt{f(i)}}$. 

A continuación, y a modo de ejemplo, obtendremos la matriz $\mathbf{T} = \mathbf{P}^{-1}$ en los dos primeros supuestos de dependencia funcional de la varianza que hemos visto, y, de esta manera, quedará completamente claro lo que acabamos de decir:

Ved los supuestos de dependencia funcional de la varianza más habituales en el subapartado 2.3 de este módulo didáctico. 

1) Cuando la varianza del término de perturbación es directamente proporcional a uno de los regresores, $\text{VAR}[u_i] = \sigma_u^2 X_{ji}$, tendremos las matrices $\mathbf{\Omega}$, \mathbf{P} y $\mathbf{P}^{-1} = \mathbf{T}$ siguientes:

$$\mathbf{\Omega} = \begin{bmatrix} X_{j1} & 0 & \dots & 0 \\ 0 & X_{j2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{jN} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \sqrt{X_{j1}} & 0 & \dots & 0 \\ 0 & \sqrt{X_{j2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{X_{jN}} \end{bmatrix},$$

$$\mathbf{P}^{-1} = \mathbf{T} = \begin{bmatrix} \frac{1}{\sqrt{X_{j1}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{X_{j2}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{X_{jN}}} \end{bmatrix}. \quad (2.36)$$

Por tanto, todos los elementos de las matrices del modelo transformado quedarán multiplicados por la inversa de la raíz cuadrada de X_{ji} , y el modelo que tendremos que estimar será:

$$\frac{Y_i}{\sqrt{X_{ji}}} = \beta_1 \frac{1}{\sqrt{X_{ji}}} + \beta_2 \frac{X_{2i}}{\sqrt{X_{ji}}} + \dots + \beta_j \frac{X_{ji}}{\sqrt{X_{ji}}} + \dots + \beta_k \frac{X_{ki}}{\sqrt{X_{ji}}} + \frac{u_i}{\sqrt{X_{ji}}} \quad \forall i = 1, \dots, N. \quad (2.37)$$

El problema consiste en el hecho de que la interpretación de los parámetros cambia y, como en este caso, no siempre resulta fácil interpretar el significado económico de los nuevos parámetros obtenidos.


2) Cuando la varianza del término de perturbación es directamente proporcional al cuadrado de uno de los regresores $\text{VAR}[u_i] = \sigma_u^2 X_{ji}^2$, tendremos las matrices Ω , \mathbf{P} y $\mathbf{P}^{-1} = \mathbf{T}$ siguientes:

$$\Omega = \begin{bmatrix} X_{j1}^2 & 0 & \dots & 0 \\ 0 & X_{j2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{jN}^2 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} X_{j1} & 0 & \dots & 0 \\ 0 & X_{j2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{jN} \end{bmatrix},$$

$$\mathbf{P}^{-1} = \mathbf{T} = \begin{bmatrix} \frac{1}{X_{j1}} & 0 & \dots & 0 \\ 0 & \frac{1}{X_{j2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{X_{jN}} \end{bmatrix}, \quad (2.38)$$

y el modelo transformado quedará como vemos a continuación:

$$\begin{aligned} \frac{Y_i}{X_{ji}} &= \beta_1 \frac{1}{X_{ji}} + \beta_2 \frac{X_{2i}}{X_{ji}} + \dots + \beta_j \frac{X_{ji}}{X_{ji}} + \dots + \beta_k \frac{X_{ki}}{X_{ji}} + \frac{u_i}{X_{ji}} = \\ &= \beta_1 \frac{1}{X_{ji}} + \beta_2 \frac{X_{2i}}{X_{ji}} + \dots + \beta_j + \dots + \beta_k \frac{X_{ki}}{X_{ji}} + \frac{u_i}{X_{ji}} \quad \forall i = 1, \dots, N. \quad (2.39) \end{aligned}$$

En este caso, a diferencia del anterior, el modelo transformado tiene una interpretación inmediata. Pasamos de analizar la relación entre los valores de las variables, a analizar la relación entre la proporción que representa cada una de las variables sobre el regresor causante de la heteroscedasticidad. 

Otros esquemas de dependencia funcional de la varianza

En cuanto al resto de los esquemas de dependencia funcional de la varianza que hemos propuesto, a continuación podemos ver las matrices $\mathbf{T} = \mathbf{P}^{-1}$ que tenemos que utilizar.

Es fácil comprobar que las matrices $\mathbf{T} = \mathbf{P}^{-1}$ que hay que utilizar en cada uno de los casos son las siguientes:

$$\bullet \text{ VAR}[u_i] = \sigma_u^2 \frac{1}{X_{ji}}; \quad \mathbf{P}^{-1} = \begin{bmatrix} \sqrt{X_{j1}} & 0 & \dots & 0 \\ 0 & \sqrt{X_{j2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{X_{jN}} \end{bmatrix}.$$

- $\text{VAR}[u_i] = \sigma_u^2 \frac{1}{X_{ji}^2}; \mathbf{P}^{-1} = \begin{bmatrix} X_{j1} & 0 & \dots & 0 \\ 0 & X_{j2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{jN} \end{bmatrix}.$

- $\text{VAR}[u_i] = \sigma_u^2(\delta_0 + \delta_1 X_{ji}); \mathbf{P}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{\delta_0 + \delta_1 X_{j1}}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\delta_0 + \delta_1 X_{j2}}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\delta_0 + \delta_1 X_{jN}}} \end{bmatrix}.$

- $\text{VAR}[u_i] = \sigma_u^2 \exp[\delta_0 + \delta_1 X_{ji}];$

$$\mathbf{P}^{-1} = \begin{bmatrix} \frac{1}{\sqrt{\exp[\delta_0 + \delta_1 X_{j1}]}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\exp[\delta_0 + \delta_1 X_{j2}]}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\exp[\delta_0 + \delta_1 X_{jN}]}} \end{bmatrix}.$$

Ejemplo ilustrativo de estimación de un modelo en presencia de heteroscedasticidad

Nuestra intención es explicar el consumo de un conjunto de 14 familias en función de su nivel de renta, es decir:

$$C_i = \beta_1 + \beta_2 R_i + u_i \quad \forall i = 1, \dots, 14.$$

Las datos en torno al consumo y nivel de renta asociados a cada una de las familias aparecen representados en la tabla siguiente:

C_i	R_i
15	10
23	12
13	14
23	16
27	18
15	20
21	22
55	24
47	26
55	28
53	30
33	32
13	34
50	36

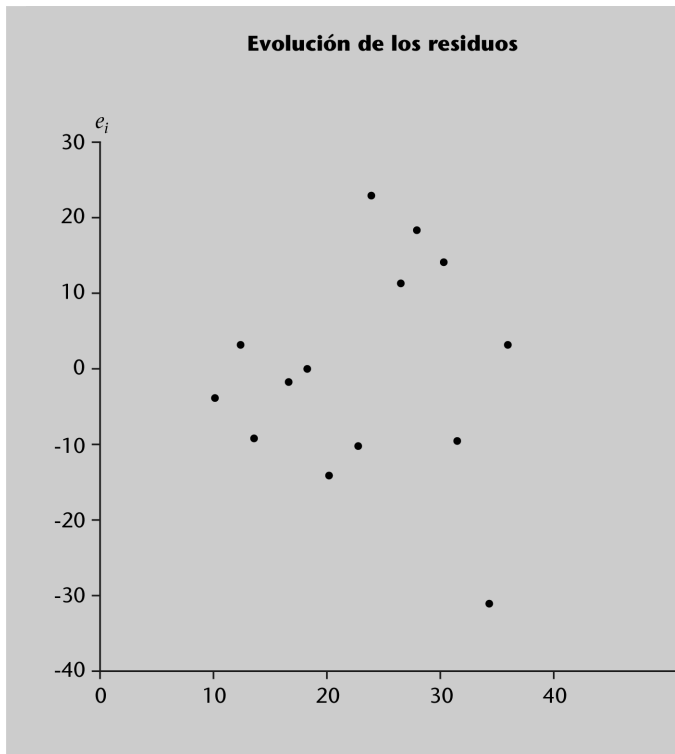
Los resultados obtenidos al estimar el modelo por MCO en el caso particular considerado son los siguientes:

$$\hat{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 14 & 322 \\ 322 & 8.316 \end{bmatrix}^{-1} \begin{bmatrix} 443 \\ 11.194 \end{bmatrix} = \begin{bmatrix} 6,242 \\ 1,104 \end{bmatrix}.$$

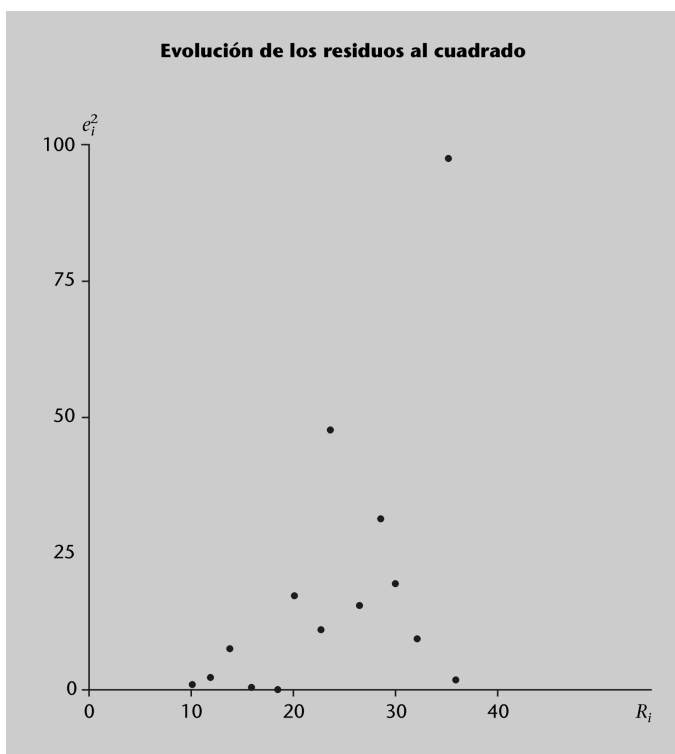
Sin embargo, tenemos la sospecha de que este modelo tiene problemas de heteroscedasticidad, y creemos que la evolución de la varianza del término de perturbación es directamente proporcional a la evolución del cuadrado de la renta familiar; es decir, nos encontraríamos con la situación $\text{VAR}[u_i] = \sigma_u^2 R_i^2, \forall i = 1, \dots, 14$.

Podríamos confirmar esta hipótesis con los siguientes gráficos, en los que se muestra la evolución de los residuos, de los residuos al cuadrado y del valor absoluto de los residuos en función de la variable R , correspondiente a la renta familiar:

- Evolución de los residuos:



- Evolución de los residuos al cuadrado:



b) A continuación, tendremos que obtener la matriz de transformación $T = P^{-1}$, que nos permitirá transformar las matrices del modelo:

$$T = P^{-1} = \begin{bmatrix} \frac{1}{10} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{12} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{14} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{16} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{18} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{20} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{22} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{24} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{26} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{28} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{30} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{32} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{34} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{36} \end{bmatrix}.$$

c) Multiplicando por P^{-1} la matriz X y el vector Y, obtendremos la matriz y el vector transformados:

$$Y^* = P^{-1}Y = \begin{bmatrix} \frac{15}{10} \\ \frac{23}{12} \\ \frac{13}{14} \\ \frac{14}{23} \\ \frac{16}{16} \\ \frac{27}{18} \\ \frac{15}{20} \\ \frac{21}{22} \\ \frac{55}{24} \\ \frac{47}{26} \\ \frac{55}{28} \\ \frac{53}{30} \\ \frac{33}{32} \\ \frac{13}{34} \\ \frac{50}{36} \end{bmatrix}; X^* = P^{-1}X = \begin{bmatrix} \frac{1}{10} \\ \frac{1}{12} \\ \frac{1}{14} \\ \frac{1}{16} \\ \frac{1}{18} \\ \frac{1}{20} \\ \frac{1}{22} \\ \frac{1}{24} \\ \frac{1}{26} \\ \frac{1}{28} \\ \frac{1}{30} \\ \frac{1}{32} \\ \frac{1}{34} \\ \frac{1}{36} \end{bmatrix}.$$

d) Y ya para finalizar, los estimadores MCP serán:

- $\hat{\mathbf{B}}_{\text{MCP}} = (\mathbf{X}^* \mathbf{X})^{-1} \mathbf{X}^* \mathbf{Y}^* = \begin{bmatrix} 0,042 & 0,706 \\ 0,706 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 1,006 \\ 19,620 \end{bmatrix} = \begin{bmatrix} 2,593 \\ 1,271 \end{bmatrix}$.
- $SCR_{\text{MCP}} = \mathbf{e}^{* \prime} \mathbf{e}^* = \mathbf{Y}^{* \prime} \mathbf{Y}^* - \hat{\mathbf{B}}_{\text{MCP}}' \mathbf{X}^* \mathbf{Y}^* = 31,214 - [2,593 \quad 1,271] \begin{bmatrix} 1,006 \\ 19,620 \end{bmatrix} = 3,668$.
- $\widehat{\text{VAR}}[\hat{\mathbf{B}}_{\text{MCP}}] = \hat{\sigma}_u^2 (\mathbf{X}^* \mathbf{X}^*)^{-1} = \frac{3,668}{14 - 2} \begin{bmatrix} 0,042 & 0,706 \\ 0,706 & 14 \end{bmatrix}^{-1} = 0,306 \begin{bmatrix} 156,313 & -7,883 \\ -7,883 & 0,469 \end{bmatrix}$.

Como se puede apreciar, los resultados obtenidos por MCG y MCP son los mismos. Asimismo, es importante que tengamos en cuenta que la interpretación de los parámetros del modelo transformado no es la misma que la del modelo original, ya que el modelo transformado se puede expresar de la siguiente forma:

$$\frac{C_i}{R_i} = \beta_1 \frac{1}{R_i} + \beta_2 + \frac{u_i}{R_i} \quad \forall i = 1, \dots, 14.$$

Entonces, β_1 registra el efecto que tiene una variación unitaria del inverso de la renta familiar en el porcentaje que representa el consumo sobre la renta familiar, y β_2 es el término independiente del modelo.

3. Autocorrelación

Ya hemos estudiado cómo tenemos que abordar el MRLM cuando el término de perturbación es no esférico, porque incumple la hipótesis de homoscedasticidad. En este apartado reproduciremos el análisis, pero, en este caso, para el supuesto en el que el término de perturbación sólo presente autocorrelación. Así pues, en primer lugar recordaremos qué quiere decir *autocorrelación* y estudiaremos algunas de las causas que pueden dar lugar a la presencia de autocorrelación en el término de perturbación del modelo. A continuación, analizaremos uno de los posibles esquemas de autocorrelación, el llamado *esquema autorregresivo de orden 1* (AR(1)) y, tras haber estudiado varios test para contrastar el cumplimiento de la hipótesis de no autocorrelación, nos dispondremos a presentar diferentes métodos para estimar un modelo con perturbaciones autocorrelacionadas. !

! Ved el tratamiento de la heteroscedasticidad en un MRLM en el apartado 2 de este módulo didáctico.

3.1. Definición y causas

En este subapartado veremos qué es la autocorrelación y las causas que pueden dar lugar a la presencia de ésta en el término de perturbación. Además, introduciremos una serie de conceptos nuevos que son básicos para el desarrollo de este apartado, tales como la **autocovarianza** y el **coeficiente de autocorrelación simple**. !

3.1.1. Definición de autocorrelación

La hipótesis de no autocorrelación (que es una de las hipótesis básicas del MRLM) significa que los términos de perturbación están incorrelacionados entre sí (además, si el término de perturbación sigue una distribución normal, la hipótesis de no autocorrelación implica independencia entre los términos de perturbación), lo cual implica que la covarianza entre cualquier par de términos de perturbación es cero*: !

! Podéis ver las hipótesis básicas del MRLM en el subapartado 2.2 del módulo "Modelo de regresión lineal múltiple...".

* Estamos suponiendo que el valor esperado de los términos de perturbación de los diferentes individuos es cero por hipótesis del modelo.

$$\text{COV}[u_i, u_j] = 0 \quad \forall i \neq j. \quad (3.1)$$

Lo que hemos dicho anteriormente se traduce en el hecho de que los elementos de fuera de la diagonal principal de la matriz de varianzas y covarianzas del término de perturbación son todos iguales a cero, lo cual significa que es una matriz diagonal*. Con el supuesto de autocorrelación, habrá, como mínimo, un par de términos de perturbación para los que no se cumplirá que la covarianza entre éstos es cero**. Así, en general, si todos los términos de perturbación están autocorrelacionados entre sí, la matriz de varianzas y covarianzas tiene la siguiente forma:

* Una matriz diagonal sólo tiene elementos distintos de cero en la diagonal principal.
** Es decir, como mínimo, un par de términos de perturbación no serán independientes, con lo que la matriz de varianzas y covarianzas ya no será una matriz diagonal.

$$\text{VAR}[\mathbf{U}] = \sigma^2 \mathbf{\Omega} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \dots & \dots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \dots & \sigma_{NN} \end{bmatrix}, \quad (3.2)$$

en la que σ_{ii} es la varianza del i -ésimo término de perturbación, que, debido que suponemos que se cumple la hipótesis de homoscedasticidad, es constante para todo i : $\text{COV}[u_i, u_i] = \text{VAR}[u_i] = \sigma^2$. Por otra parte, σ_{ij} es la covarianza entre el i -ésimo y el j -ésimo términos de perturbación, que, si hay autocorrelación, será distinta de cero (como mínimo para algún $i \neq j$, $i = 1, \dots, N$, $j = 1, \dots, N$). Así pues, la matriz de varianzas y covarianzas 3.2 queda de la siguiente manera:

$$\text{VAR}[U] = \sigma^2 \Omega = \begin{bmatrix} \sigma^2 & \sigma_{12} & \dots & \dots & \sigma_{1N} \\ \sigma_{21} & \sigma^2 & \dots & \dots & \sigma_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \dots & \dots & \sigma^2 \end{bmatrix}. \quad (3.3)$$

Como veremos más adelante, una de las causas que pueden dar lugar a la presencia de autocorrelación es trabajar con datos de serie temporal; por lo tanto, en este subapartado no hablaremos de la covarianza entre los términos de perturbación correspondientes a diferentes individuos (el i -ésimo y el j -ésimo), como hemos estado haciendo hasta ahora, sino que hablaremos de la covarianza entre un término de perturbación en un instante del tiempo t y el mismo término de perturbación en otro instante del tiempo $t + s$ (o $t - s$). De esta manera, 3.1 queda:

$$\text{COV}[u_t, u_{t-s}] = E[u_t u_{t-s}] \quad \forall s = 0, \pm 1, \pm 2, \dots \quad (3.4)$$

Seguro que recordáis que la estacionariedad de una variable aleatoria se puede definir en sentido estricto (*estacionariedad fuerte*) o en sentido débil. El primer caso implica que la función de probabilidad es invariante con el paso del tiempo, mientras que el segundo implica que los momentos de primer y segundo orden (es decir, el valor esperado, y la varianza y las covarianzas, respectivamente) son finitos e independientes del tiempo. Además, podemos demostrar que, si la variable se comporta según una ley normal y es estacionaria en sentido débil, entonces también lo es en sentido estricto. En cualquier caso, de ahora en adelante supondremos estacionariedad débil.

Llamaremos **autocovarianza** (donde añadimos el prefijo *auto* para denotar que se trata de un mismo término de perturbación) a la covarianza 3.4 y la representaremos por γ_s , siendo s el número de periodos de tiempo transcurridos.

Autocovarianza de orden s

Veamos diferentes ejemplos de autocovarianza de orden s , según su definición:

- Orden 0: si $s = 0$, tenemos:

$$\gamma_0 = \text{COV}[u_t, u_{t-0}] = \text{COV}[u_t, u_t] = \text{VAR}[u_t] = \sigma^2.$$


- Orden 1: si $s = 1$, obtenemos:

$$\gamma_1 = \text{COV}[u_t, u_{t-1}].$$

! Ved las causas de presencia de autocorrelación en el subapartado 3.1.2 de este módulo didáctico.

Notad que en 3.4,...

... de la misma manera que en 3.1, continuamos suponiendo que el valor esperado de los términos de perturbación (ahora, en cualquier instante del tiempo) es cero por hipótesis del modelo.

La autocovarianza cumple las tres propiedades siguientes: 

- a) La autocovarianza es simétrica. Es decir, $\gamma_s = \gamma_{-s} \Rightarrow \text{COV}[u_t, u_{t-s}] = \text{COV}[u_t, u_{t+s}]$.
- b) La autocovarianza entre dos términos de perturbación cualesquiera sólo depende del número de periodos de tiempo (s) que hay entre éstos. En otras palabras, sólo depende del lapso temporal que hay entre los términos de perturbación.
- c) En valor absoluto, las autocovarianzas son, como máximo, iguales a la varianza:

$$|\gamma_s| \leq \gamma_0 \quad \forall s.$$

Esta propiedad garantiza que la matriz de varianzas y covarianzas de los términos de perturbación sea una matriz definida positiva.

El conjunto formado por todas las autocovarianzas ($\gamma_1, \gamma_2, \gamma_3, \dots$) se conoce como **función de autocovarianzas**.

Teniendo en cuenta las propiedades presentadas de la autocovarianza, la matriz de varianzas y covarianzas de los términos de perturbación 3.3 se puede escribirse de la siguiente manera:

$$\text{VAR}[U] = \sigma^2 \Omega = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 \end{bmatrix}. \quad (3.5)$$

Para poder identificar la matriz Ω por separado de σ^2 , tenemos que introducir un concepto nuevo: el **coeficiente de autocorrelación simple**.

El coeficiente de autocorrelación simple entre dos términos de perturbación separados entre sí por s periodos de tiempo, ρ_s , es el cociente entre la autocovarianza de orden s y la varianza:

$$\rho_s = \frac{\gamma_s}{\gamma_0} \quad \forall s = \pm 1, \pm 2, \dots \quad (3.6)$$

El coeficiente de autocorrelación simple cumple las cinco propiedades que vemos a continuación:

- a) Cuando $s = 0$, entonces vale 1:

$$\rho_s = \frac{\gamma_0}{\gamma_0} = 1.$$

Este hecho pone de manifiesto que el coeficiente de autocorrelación entre un término de perturbación y él mismo en el mismo instante del tiempo es 1.

Según la propiedad a,...

... tenemos, por ejemplo:
 $\text{COV}[u_t, u_{t-2}] = \text{COV}[u_t, u_{t+2}]$.

La propiedad b...

... permite escribir, por ejemplo:

- $\gamma_1 = \text{COV}[u_t, u_{t-1}] \Rightarrow$
 $\Rightarrow \text{COV}[u_1, u_2] = \text{COV}[u_2, u_3] =$
 $= \dots = \text{COV}[u_{36}, u_{37}] = \dots$
- $\gamma_2 = \text{COV}[u_t, u_{t-2}] \Rightarrow$
 $\Rightarrow \text{COV}[u_1, u_3] = \text{COV}[u_2, u_4] =$
 $= \dots = \text{COV}[u_{37}, u_{39}] = \dots$

b) El coeficiente de autocorrelación simple no es más que un coeficiente de correlación. Pensad que el coeficiente de correlación simple entre dos variables aleatorias cualesquiera, X e Y , viene determinado por:

$$r_{X,Y} = \frac{\text{COV}[X,Y]}{\sqrt{\text{VAR}[X]} \sqrt{\text{VAR}[Y]}}$$

Reproduciendo la expresión anterior para nuestro caso, obtenemos que:

$$\rho_s = r_{u_t, u_{t-s}} = \frac{\text{COV}[u_t, u_{t-s}]}{\sqrt{\text{VAR}[u_t]} \sqrt{\text{VAR}[u_{t-s}]}} = \frac{\gamma_s}{\sigma\sigma} = \frac{\gamma_s}{\sigma^2} = \frac{\gamma_s}{\gamma_0}$$

Así pues, el coeficiente de autocorrelación simple muestra la relación que hay entre el término de perturbación en dos instantes del tiempo diferentes.

c) Como todo coeficiente de correlación, está acotado entre -1 y 1 :

$$|\rho_s| \leq 1.$$

d) Como las autocovarianzas son simétricas, el coeficiente de autocorrelación cumple la siguiente propiedad:

$$\rho_s = \rho_{-s} \Rightarrow \frac{\gamma_s}{\gamma_0} = \frac{\gamma_{-s}}{\gamma_0}$$

e) Puesto que la autocovarianza sólo depende del número de periodos de tiempo transcurridos, el coeficiente de autocorrelación también dependerá de lo mismo.

El conjunto formado por todos los coeficientes de autocorrelación simple ($\rho_1, \rho_2, \rho_3, \dots$) se conoce como **función de autocorrelación simple, FAS***.

Ejemplo de la propiedad e

La propiedad e permite, por ejemplo, escribir:

$$\begin{aligned} \rho_3 &= \text{correlación}(u_1, u_4) = \\ &= \text{correlación}(u_2, u_5) = \dots = \\ &= \text{correlación}(u_t, u_{t+3}) = \\ &= \text{correlación}(u_t, u_{t-3}). \end{aligned}$$

* En nomenclatura inglesa, ACF.

Teniendo en cuenta las propiedades del coeficiente de autocorrelación simple, podemos escribir la matriz de varianzas y covarianzas de los términos de perturbación 3.5 de la siguiente manera:

$$\begin{aligned} \text{VAR}[\mathbf{U}] &= \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \gamma_{T-3} & \dots & \gamma_0 \end{bmatrix} = \begin{bmatrix} \frac{\gamma_0\gamma_0}{\gamma_0} & \frac{\gamma_1\gamma_0}{\gamma_0} & \frac{\gamma_2\gamma_0}{\gamma_0} & \dots & \frac{\gamma_{T-1}\gamma_0}{\gamma_0} \\ \frac{\gamma_1\gamma_0}{\gamma_0} & \frac{\gamma_0\gamma_0}{\gamma_0} & \frac{\gamma_1\gamma_0}{\gamma_0} & \dots & \frac{\gamma_{T-2}\gamma_0}{\gamma_0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\gamma_{T-1}\gamma_0}{\gamma_0} & \frac{\gamma_{T-2}\gamma_0}{\gamma_0} & \frac{\gamma_{T-3}\gamma_0}{\gamma_0} & \dots & \frac{\gamma_0\gamma_0}{\gamma_0} \end{bmatrix} = \\ &= \gamma_0 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix} = \sigma^2 \Omega. \end{aligned} \tag{3.7}$$

3.1.2. Causas que pueden dar lugar a la presencia de autocorrelación en el término de perturbación

Las causas que pueden generar autocorrelación en el término de perturbación de un modelo de regresión son varias, aunque tenemos que destacar las cuatro siguientes: !

a) Trabajar con datos de serie temporal. Con datos de serie temporal los factores que registra el término de perturbación pueden presentar una determinada tendencia temporal a lo largo del periodo muestral que depende de los anteriores en el tiempo, lo cual significa que los términos de perturbación pueden no ser independientes los unos de los otros.

b) Una especificación errónea del modelo en la parte determinista de éste. En este caso, se dice que hay **autocorrelación espuria**. En este supuesto podemos diferenciar dos situaciones:

- Omisión de variables relevantes. Suponemos que el modelo correctamente especificado es:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \forall t = 1, \dots, T,$$

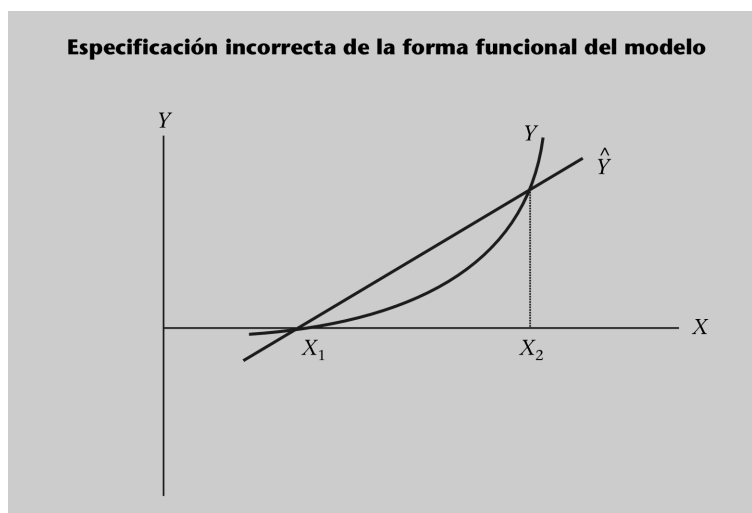
pero, por error, se especifica el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + v_t \quad \forall t = 1, \dots, T.$$

En tal caso, el nuevo término de perturbación, v_t , viene determinado por $v_t = u_t + \beta_3 X_{3t}$. Así pues, siempre que los valores de la variable omitida* dependan de sus valores pasados, esta dependencia se trasladará al término de perturbación, con lo que v_t presentará autocorrelación.

* En este ejemplo, la variable omitida es X_3 .

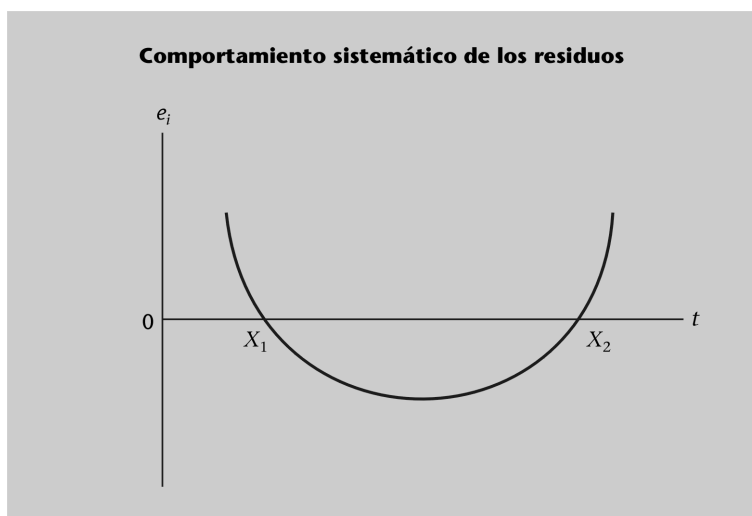
- Especificación incorrecta de la forma funcional del modelo. Imaginémonos que la forma funcional correcta es cuadrática, pero, por error, especificamos una forma lineal:



Si, a partir del gráfico anterior, analizamos los residuos del ajuste, lleguemos a las siguientes conclusiones:

- Para $X < X_1 \rightarrow Y > \hat{Y} \Rightarrow e > 0$.
- Para $X_1 < X < X_2 \rightarrow Y < \hat{Y} \Rightarrow e < 0$.
- Para $X > X_2 \rightarrow Y > \hat{Y} \Rightarrow e > 0$.

Al representar gráficamente estos residuos, observamos que presentan un comportamiento sistemático (no aleatorio), lo que es, como veremos más adelante, señal de autocorrelación:



c) Determinadas transformaciones a las que podamos someter los datos, como por ejemplo, cuando trabajamos con datos temporales y consideramos las diferencias entre observaciones de periodos sucesivos (esta transformación se conoce como *primeras diferencias*).

Ejemplo de trabajo en primeras diferencias

Supongamos el siguiente MRLM, en el que el término de perturbación cumple todas las hipótesis básicas:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t, \quad \forall t = 1, \dots, T,$$

y lo retardamos un periodo:

$$Y_{t-1} = \beta_1 + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + u_{t-1}, \quad \forall t = 2, \dots, T.$$

Si calculamos la diferencia entre los dos modelos anteriores, obtendremos el modelo que vemos aquí:

$$\Delta Y_t = \beta_2 \Delta X_{2t} + \beta_3 \Delta X_{3t} + \Delta u_t, \quad \forall t = 2, \dots, T,$$

en el que Δ representa el operador diferencias, de manera que, por ejemplo, $\Delta Y_t = Y_t - Y_{t-1}$. Si analizamos el término de perturbación del modelo obtenido, veremos que es homoscedástico, aunque presenta autocorrelación. En efecto:

$$\begin{aligned} \text{VAR}[\Delta u_t] &= \text{VAR}[u_t - u_{t-1}] = E[(u_t - u_{t-1})^2] = E[u_t^2 + u_{t-1}^2 - 2u_t u_{t-1}] = \\ &= \text{VAR}[u_t] + \text{VAR}[u_{t-1}] - 2\text{COV}[u_t, u_{t-1}] = \sigma_u^2 + \sigma_u^2 - 0 = 2\sigma_u^2 = \text{constante } \forall t. \end{aligned}$$

Notad, sin embargo, que, aunque Δu_t es homoscedástico, su varianza es el doble que la del término de perturbación del modelo inicial.

En cuanto a la autocorrelación, si calculamos la autocovarianza de primer orden tenemos que:

$$\begin{aligned} \gamma_1 &= \text{COV}[\Delta u_t, \Delta u_{t-1}] = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] = \\ &= E[(u_t u_{t-1} - u_t u_{t-2} - u_{t-1}^2 + u_{t-1} u_{t-2})] = 0 + 0 - \sigma^2 + 0 = -\sigma^2 \neq 0 \Rightarrow \text{autocorrelación.} \end{aligned}$$

De todos modos, si calculamos las autocovarianzas de órdenes superiores, $\gamma_2, \gamma_3, \gamma_4, \dots$, encontraremos que todas son cero. A pesar de esto, este resultado no invalida la existencia de autocorrelación, puesto que, para que haya autocorrelación, es suficiente con que un (dos, teniendo en cuenta la simetría de la matriz) elemento de fuera de la diagonal principal de la matriz de varianzas y covarianzas sea diferente de cero. Así pues, atendiendo a los resultados anteriores, la matriz de varianzas y covarianzas del término de perturbación del modelo transformado viene determinada por:

$$\begin{aligned} \text{VAR}[\Delta U] &= \begin{bmatrix} 2\sigma^2 & -\sigma^2 & 0 & \dots & 0 \\ -\sigma^2 & 2\sigma^2 & -\sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2\sigma^2 \end{bmatrix} = 2\sigma^2 \begin{bmatrix} 1 & \frac{-\sigma_u^2}{2\sigma^2} & 0 & \dots & 0 \\ \frac{-\sigma_u^2}{2\sigma^2} & 1 & \frac{-\sigma^2}{2\sigma^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \\ &= \sigma^2 \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix}. \end{aligned}$$

Recordad que el modelo de perturbación original cumple las hipótesis básicas, por lo que es no autocorrelacionado.

d) Trabajar con determinadas especificaciones dinámicas.

Ejemplo de modelo de retardos distribuidos

Se dice que un modelo es dinámico si no todas las variables hacen referencia al mismo instante de tiempo. Un modelo de este tipo es el llamado *modelo de retardos distribuidos*, que resulta útil cuando la variable endógena depende de los valores de una o más variables han tomado en distintos periodos de tiempo pasados. Pensad, por ejemplo, en el caso de las ventas de una empresa en un periodo t (año, trimestre, etc.). Supongamos, además, con el fin de simplificar, que la única variable explicativa relevante es el gasto en publicidad (obviamos otras variables que sin duda también deberíamos considerar, como el precio de los productos sustitutivos, el precio del mismo producto fabricado por otra empresa del mismo sector, etc.). Con estos supuestos, está claro que las ventas de la empresa en el periodo t no sólo dependerán de la publicidad que la empresa haya hecho en el periodo en curso, sino también de la que haya hecho en los periodos anteriores, si bien, ciertamente, la influencia será cada vez sensiblemente menor a medida que nos alejamos en el tiempo. En cualquier caso, un modelo que registre esta situación vendría determinado por:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \dots + u_t \quad \forall t = 1, \dots, T,$$

dónde Y_t representa las ventas realizadas por la empresa en el periodo t , y $X_t, X_{t-1}, X_{t-2}, \dots$, el gasto en publicidad los periodos $t, t-1, t-2, \dots$. Además, para simplificar, supondremos que el término de perturbación del modelo considerado es esférico y tiene valor esperado cero. Para estimar este modelo, una posibilidad es imponer la llamada *hipótesis de Koyck*, que establece que el efecto de la variable explicativa sobre la variable endógena decrece con el paso del tiempo de acuerdo con una progresión geométrica. En nuestro ejemplo, esto quiere decir que el efecto de la publicidad que la empresa haga en el periodo en curso, t , será mayor sobre las ventas de la empresa en este periodo t que la publicidad que hizo en el periodo anterior, $t-1$, y, al mismo tiempo, ésta será más efectiva que la que hizo hace dos periodos. La forma de introducir esta hipótesis en el modelo es definiendo la siguiente progresión geométrica:

$$\beta_j = \delta^j \beta_0 \quad 0 < \delta < 1 \Rightarrow \begin{cases} \beta_0 = \delta^0 \beta_0 = \beta_0 \\ \beta_1 = \delta^1 \beta_0 < \beta_0 \\ \beta_2 = \delta^2 \beta_0 < \beta_1 \\ \dots \end{cases}.$$

Teniendo en cuenta la última expresión, el modelo queda definido de la manera siguiente:

$$Y_t = \alpha + \beta_0 X_t + \beta_0 \delta X_{t-1} + \beta_0 \delta^2 X_{t-2} + \dots + u_t \quad \forall t = 1, \dots, T,$$

y retardando un periodo la expresión anterior y multiplicando por δ , tenemos:

$$\delta Y_{t-1} = \delta\alpha + \delta\beta_0 X_{t-1} + \beta_0 \delta^2 X_{t-2} + \beta_0 \delta^3 X_{t-3} + \dots + \delta u_{t-1} \quad \forall t = 2, \dots, T.$$

Para finalizar, realizando el cálculo de la diferencia entre las dos últimas expresiones, llegamos al siguiente modelo:

$$\begin{aligned} Y_t - \delta Y_{t-1} &= (1 - \delta)\alpha + \beta_0 X_t + (u_t - \delta u_{t-1}) \Rightarrow \\ \Rightarrow Y_t &= (1 - \delta)\alpha + \beta_0 X_t + \delta Y_{t-1} + (u_t - \delta u_{t-1}) \quad \forall t = 2, \dots, T. \end{aligned}$$

Pues bien, si analizamos el término de perturbación del modelo al que hemos llegado, podemos comprobar que, si bien es homoscedástico:

$$\begin{aligned} \text{VAR}[u_t - \delta u_{t-1}] &= E[(u_t - \delta u_{t-1})^2] = E[u_t^2 + \delta^2 u_{t-1}^2 - 2\delta u_t u_{t-1}] = \\ &= \text{VAR}[u_t] + \text{VAR}[\delta u_{t-1}] - 2\text{COV}[u_t, \delta u_{t-1}] = \sigma^2 + \delta^2 \sigma^2 - 0 = \sigma^2(1 + \delta^2), \end{aligned}$$

presenta autocorrelación:

$$\begin{aligned} \gamma_1 &= E[(u_t - \delta u_{t-1})(u_{t-1} - \delta u_{t-2})] = E(u_t u_{t-1} - \delta u_t u_{t-2} - \delta^2 u_{t-1} u_{t-2}) = \\ &= 0 + 0 - \delta \sigma^2 + 0 = -\delta \sigma^2 \neq 0 \Rightarrow \text{autocorrelación.} \end{aligned}$$

Si calculamos las autocovarianzas de órdenes superiores, $\gamma_2, \gamma_3, \gamma_4, \dots$, encontraremos, igual que sucedía en el supuesto de transformación en primeras diferencias, que todas son cero. Recordad, sin embargo, que este hecho no invalida la existencia de autocorrelación.

Así pues, y de acuerdo con los resultados obtenidos, la matriz de varianzas y covarianzas del término de perturbación del modelo transformado viene determinada por:

$$\begin{aligned} \text{VAR}[u_t - \delta u_{t-1}] &= \begin{bmatrix} \sigma^2(1+\delta^2) & -\delta\sigma^2 & 0 & \dots & 0 \\ -\delta\sigma^2 & \sigma^2(1+\delta^2) & -\delta\sigma^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2(1+\delta^2) \end{bmatrix} = \\ &= \sigma_u^2(1 + \delta^2) \begin{bmatrix} 1 & \frac{-\delta}{1 + \delta^2} & 0 & \dots & 0 \\ \frac{-\delta}{1 + \delta^2} & 1 & \frac{-\delta}{1 + \delta^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \end{aligned}$$

3.2. Esquemas de autocorrelación en el término de perturbación


Como sabemos, ante la presencia de un término de perturbación no esférico (sea heteroscedástico y/o autocorrelacionado, como es este último caso del que estamos tratando) en el modelo de regresión, los estimadores MCO de los parámetros \mathbf{B} , aun manteniendo las propiedades de no sesgo y consistencia, no son eficientes. Así, el método de estimación que ofrece estimadores de varianza mínima (además de no sesgados y consistentes) en estas circunstancias es el de MCG.

El método de estimación de MCG pasa, como sabemos, por conocer la matriz Ω , que, en el supuesto de autocorrelación, está formada por los coeficientes de autocorrelación simple, tal y como hemos tenido ocasión de ver. El problema con el que nos encontramos es que estos coeficientes son desconocidos, por lo que es necesario estimarlos previamente a partir de las T observaciones muestrales de las que disponemos. Sin embargo, esto no es posible, ya que, suponiendo que hay homoscedasticidad*, el número de elementos de la matriz Ω que tenemos que estimar es $T(T - 1) / 2$. Es decir, hay más parámetros por estimar que observaciones.

Ved las propiedades de los estimadores MCO en presencia de un término de perturbación no esférico en el subapartado 1.2 del presente módulo didáctico.

Ved la matriz Ω en el supuesto de autocorrelación en el subapartado 3.1.1 de este módulo didáctico.

* En el supuesto de heteroscedasticidad y autocorrelación, el número de elementos de la matriz Ω que deberíamos estimar es $T(T + 1)/2$.

Tenemos que introducir, en tal caso, algún tipo de parametrización para reducir el número de parámetros que hay que estimar. Por este motivo, se suponen determinados tipos (esquemas) de autocorrelación que puede seguir el término de perturbación. Estos **esquemas de autocorrelación** son de tres tipos: **esquema autorregresivo (AR)**, **esquema media móvil (MI)** y **esquema mixto**. De entre todos ellos, únicamente estudiaremos los primeros, los esquemas autorregresivos. 

Modelo ARMA

Un modelo autorregresivo (AR) en media móvil (MA) al mismo tiempo se conoce como *modelo mixto* o *modelo ARMA*.

3.2.1. Esquema de autocorrelación autorregresivo (AR): el esquema AR(1)

Se dice que un término de perturbación sigue o ha sido generado por un esquema autorregresivo si es una combinación lineal de sus propios retardos y de una variable aleatoria que recibe el nombre de **variable aleatoria ruido blanco**.

Dependiendo del número de retardos del término de perturbación que haya en la combinación lineal, tendremos un esquema autorregresivo de un orden o de otro. Así, si el término de perturbación es una combinación lineal de su primer retardo, y de la variable aleatoria ruido blanco se dice que ha sido generada por un proceso autorregresivo de orden 1, y se simboliza por AR(1):

$$u_t = \phi u_{t-1} + \varepsilon_t$$

en la que ε es el vector de variables aleatorias (formado por los componentes ε_t) ruido blanco, que se caracteriza porque tiene un valor esperado cero ($E[\varepsilon] = \mathbf{0}_T$), y una matriz de varianzas y covarianzas escalar ($\text{VAR}(\varepsilon) = \sigma_\varepsilon^2 \mathbf{I}_T$).

Una vez definido el esquema autorregresivo de orden 1, nos disponemos a estudiar las propiedades que cumple un término de perturbación que sigue este tipo de proceso. En concreto, estudiaremos el valor esperado, la varianza, las autocovarianzas y los coeficientes de autocorrelación simple, con el fin de poder especificar, finalmente, la matriz de varianzas y covarianzas.

Para empezar, enunciaremos las propiedades del término de perturbación en el esquema AR(1):

a) El valor esperado de un término de perturbación generado por un proceso AR(1) es cero:

$$E[u_t] = \phi_1 E[u_{t-1}] + E[\varepsilon_t] \Rightarrow \mu = \phi_1 \mu + 0 \Rightarrow (1 - \phi_1) \mu = 0.$$

Para que se cumpla la igualdad anterior, o bien ϕ_1 tiene que ser 1, o bien μ tiene que ser cero. Entre las dos posibilidades, como veremos al estudiar la varianza, la primera no es admisible; en consecuencia, la única opción es que $\mu = E[u_t] = 0$.

Notad que el ruido blanco,...

... de hecho, no es un concepto nuevo, puesto que es el equivalente al término de perturbación del modelo de regresión cuando éste cumple todas las hipótesis básicas.

b) La varianza del término de perturbación es directamente proporcional a la varianza de la variable aleatoria ruido blanco:

$$\begin{aligned}\text{VAR}[u_t] &= E[u_t^2] = E[(\phi_1 u_{t-1} + \varepsilon_t)^2] = \\ &= \phi_1^2 \text{VAR}[u_{t-1}] + \text{VAR}[\varepsilon_t] + 2\phi_1 \text{COV}[u_{t-1}, \varepsilon_t],\end{aligned}\quad (3.8)$$

y, teniendo en cuenta, por una parte, que suponemos que el término de perturbación cumple la hipótesis de homoscedasticidad, lo cual quiere decir que $\text{VAR}[u_t] = \text{VAR}[u_{t-1}] = \sigma^2 = \gamma_0$, y, por la otra, que el término de perturbación y la variable ruido blanco sólo están correlacionados contemporáneamente, lo cual quiere decir que $\text{COV}[u_{t-1}, \varepsilon_t] = 0$, 3.8 queda como vemos a continuación:

$$\gamma_0 = \phi_1^2 \gamma_0 + \sigma_\varepsilon^2 + 0 \Rightarrow \gamma_0 = \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)}.$$

El resultado obtenido pone de manifiesto que, si ϕ fuera 1, o mayor que 1, entonces la varianza del término de perturbación no estaría bien definida.

Por lo tanto, es necesario que el parámetro ϕ_1 esté acotado entre 1 y -1: $|\phi_1| < 1$.

c) Las autocovarianzas dependen del parámetro ϕ_1 . Para comprobar este hecho, sólo tenemos que calcular las autocovarianzas de diferentes órdenes:

$$\gamma_1 = \text{COV}[u_t, u_{t-1}] = E[u_t u_{t-1}] = E[(\phi_1 u_{t-1} + \varepsilon_t) u_{t-1}] = E[\phi_1 u_{t-1}^2 + \varepsilon_t u_{t-1}] = \phi_1 \gamma_0,$$

$$\gamma_2 = \text{COV}[u_t, u_{t-2}] = E[u_t u_{t-2}] = E[(\phi_1 u_{t-1} + \varepsilon_t) u_{t-2}],$$

y, sustituyendo u_{t-1} por su expresión de acuerdo con un esquema autorregresivo de primer orden ($u_{t-1} = \phi_1 u_{t-2} + \varepsilon_{t-1}$), tenemos que:

$$\gamma_2 = E[(\phi_1^2 u_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t) u_{t-2}] = E[\phi_1^2 u_{t-2}^2 + \phi_1 \varepsilon_{t-1} u_{t-2} + \varepsilon_t u_{t-2}] = \phi_1^2 \gamma_0.$$

Así, en general, podemos comprobar que la autocovarianza para cualquier retardo s viene determinada por:

$$\gamma_s = \text{COV}[u_t, u_{t-s}] = E[u_t u_{t-s}] = \dots = \phi_1^s \gamma_0.$$

d) A partir de los resultados obtenidos por las autocovarianzas, debemos comprobar de inmediato que los coeficientes de autocorrelación simple de un término de perturbación generado por un esquema AR(1) están determinados por:

$$\rho_1 = \frac{\phi_1 \gamma_0}{\gamma_0} = \phi_1; \rho_2 = \frac{\phi_1^2 \gamma_0}{\gamma_0} = \phi_1^2; \dots; \rho_s = \frac{\phi_1^s \gamma_0}{\gamma_0} = \phi_1^s.$$

La varianza del término de perturbación...

... no estaría bien definida en el sentido de que sería cero si ϕ_1 fuera 1, o negativa, si en valor absoluto ϕ_1 fuese mayor que 1.

Notad que...

... en la demostración del cálculo de la autocovarianza se ha hecho uso del hecho de que el término de perturbación y la variable aleatoria ruido blanco únicamente están correlacionados contemporáneamente. Por lo tanto, $E[\varepsilon_t u_{t-1}] = 0$. Por el mismo motivo, en las demostraciones subsiguientes hallamos $E[\varepsilon_{t-1} u_{t-2}] = 0$ y $E[\varepsilon_t u_{t-2}] = 0$.

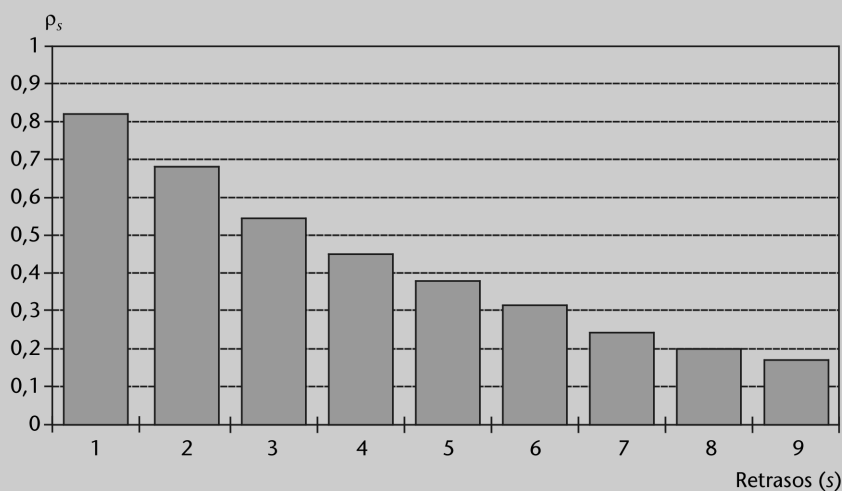
Teniendo en cuenta este resultado, la matriz de varianzas y covarianzas del término de perturbación 3.7 queda como se sigue:

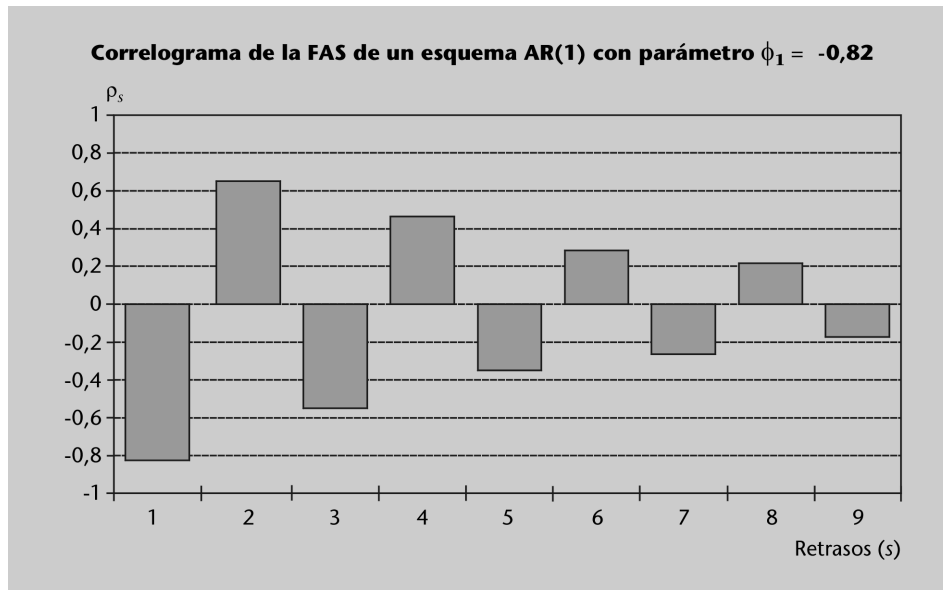
$$\text{VAR}[U]=\gamma_0 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix} = \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{T-1} & \phi_1^{T-2} & \phi_1^{T-3} & \dots & 1 \end{bmatrix},$$

que, como vemos, tan sólo depende de dos parámetros: la varianza de la variable aleatoria ruido blanco, σ_ε^2 , y el parámetro del esquema autorregresivo, ϕ_1 .

Por otra parte, notad que, dado que $|\phi_1| < 1$, los coeficientes de autocorrelación simple decrecen exponencialmente hacia cero: $\rho_1 > \rho_2 > \dots > \rho_s \rightarrow 0$. De esta manera, si los representamos gráficamente (este gráfico se conoce como **correlograma de la FAS**), observaremos un comportamiento decreciente hacia cero; pero, sin embargo, dependiendo del valor del parámetro ϕ_1 , los signos se irán alternando empezando por negativo (si $-1 < \phi_1 < 0$) o serán siempre positivos (si $0 < \phi_1 < 1$). Ved los gráficos que presentamos a continuación, en los que mostramos un ejemplo de cada una de estas situaciones. En concreto, en estos dos gráficos mostramos la magnitud de los primeros nueve coeficientes de autocorrelación simple dispuestos en orden: $\rho_1, \rho_2, \dots, \rho_9$.

Correlograma de la FAS de un esquema AR(1) con parámetro $\phi_1 = 0,82$





Ejemplo de cálculo de la matriz de varianzas y covarianzas del término de perturbación que sigue un esquema AR(1)

Supongamos que un término de perturbación u_t está autocorrelacionado según un esquema AR(1), del cual conocemos que $\phi_1 = 0,6$ y que $\sigma_\varepsilon^2 = 2$. En este ejemplo veremos cómo se calcula el valor esperado, la varianza, las cuatro primeras autocovarianzas, los cuatro primeros coeficientes de autocorrelación simple, escribiremos la matriz de varianzas y covarianzas, y, para finalizar, elaboraremos el correlograma de la FAS correspondiente.

A partir de la información del enunciado, podemos escribir que el esquema AR(1) que sigue el término de perturbación es:

$$u_t = 0,6u_{t-1} + \varepsilon_t, \quad \sigma_\varepsilon^2 = 2.$$

Una vez hemos especificado el modelo, podemos calcular:

- El valor esperado del término de perturbación en este esquema es el siguiente:

$$E[u_t] = 0,6E[u_{t-1}] + E[\varepsilon_t] \Rightarrow \mu = 0,6\mu + 0 \Rightarrow (1 - 0,6)\mu = 0 \Rightarrow \mu = \frac{0}{0,4} = 0.$$

- La varianza es:

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)} = \frac{2}{(1 - 0,6^2)} = \frac{2}{0,64} = 3,125.$$

- Las autocovarianzas que resultan aparecen a continuación:

$$\begin{aligned} \gamma_1 &= \phi_1 \gamma_0 = 0,6 \cdot 3,125 = 1,875; \\ \gamma_2 &= \phi_1^2 \gamma_0 = 0,6^2 \cdot 3,125 = 1,125; \\ \gamma_3 &= \phi_1^3 \gamma_0 = 0,6^3 \cdot 3,125 = 0,675; \\ \gamma_4 &= \phi_1^4 \gamma_0 = 0,6^4 \cdot 3,125 = 0,405. \end{aligned}$$

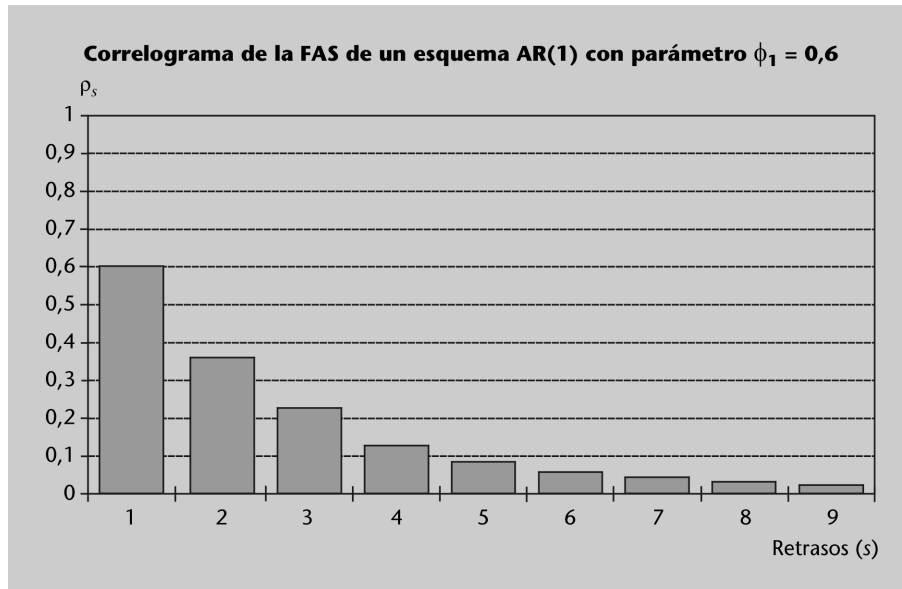
- Los coeficientes de autocorrelación simple que obtenemos en este esquema son:

$$\rho_1 = \phi_1 = 0,6; \rho_2 = \phi_1^2 = 0,36; \rho_3 = \phi_1^3 = 0,216; \rho_4 = \phi_1^4 = 0,1296.$$

A partir de estos resultados, la matriz de varianzas y covarianzas del término de perturbación queda de la siguiente manera:

$$\text{VAR}[U] = \gamma_0 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix} = 3,125 \begin{bmatrix} 1 & 0,6 & 0,36 & 0,216 & \dots \\ 0,6 & 1 & 0,6 & 0,36 & \dots \\ 0,36 & 0,6 & 1 & 0,6 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \dots & 1 \end{bmatrix}.$$

- Para acabar, el correlograma de FAS queda así:



Actividad

3.1. Repetid el ejemplo de resolución del término de perturbación que acabamos de proponer, pero considerando el siguiente esquema autorregresivo de orden 1 para el término de perturbación: $u_t = -0,6u_{t-1} + \varepsilon_t$, $\sigma_\varepsilon^2 = 2$. Observad las diferencias que hallaréis entre ellos.

3.3. Detección de la autocorrelación

En este subpartado estudiaremos dos tipos de instrumentos para detectar la posible presencia de autocorrelación en el término de perturbación. El primero de ellos consiste en llevar a cabo un análisis gráfico de los residuos del ajuste MCO del modelo, y el segundo, en toda una batería de contrastes. ⚠

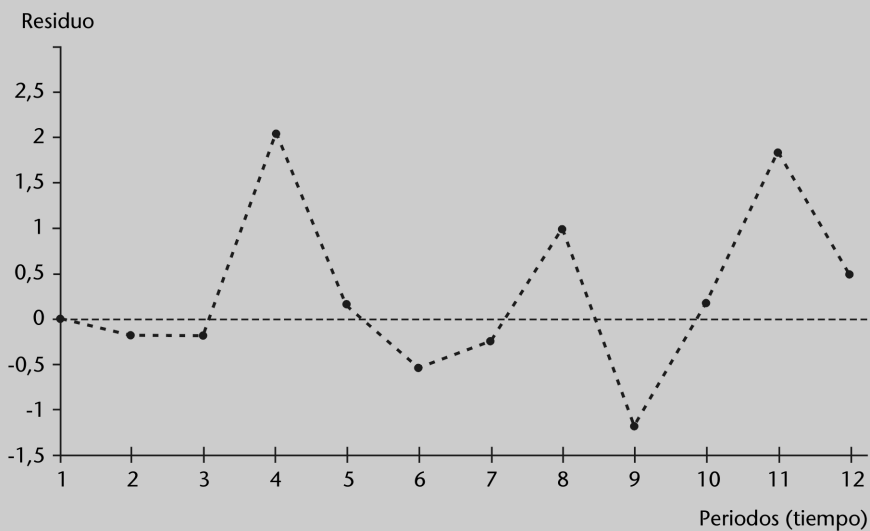
3.3.1. Análisis gráfico de los residuos

Dado el modelo (expresado en términos matriciales) $Y = XB + U$, del cual queremos averiguar si el término de perturbación presenta autocorrelación, tenemos que estimarlo por MCO, obtener el vector de residuos y representar los residuos con respecto al tiempo en una gráfica. En el supuesto de que la nube de puntos resultante presente un comportamiento aleatorio, diremos que no hay autocorrelación, mientras que, si, por el contrario, observamos un patrón de comportamiento sistemático, es un síntoma de la presencia de algún tipo de esquema de autocorrelación en el término de perturbación.

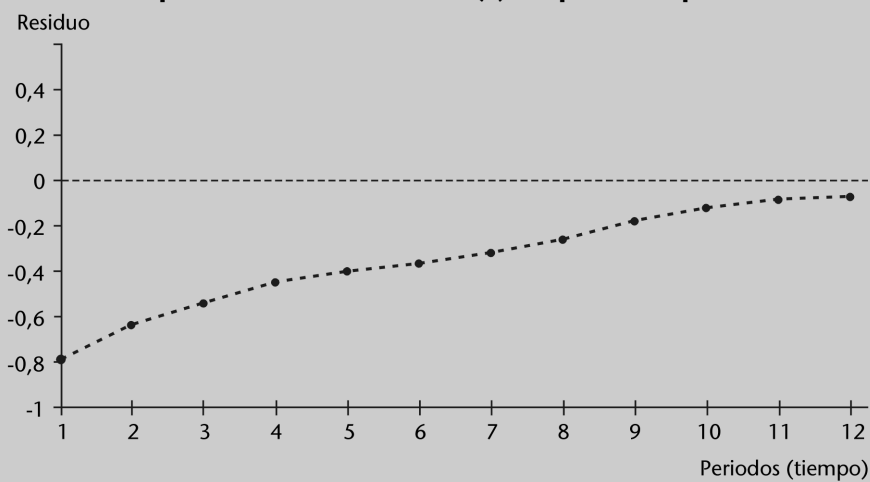
En los gráficos que presentamos a continuación podemos observar distintas situaciones. En la primera, la nube de puntos es completamente aleatoria, mientras que en las cuatro restantes hallamos un patrón de comportamiento sistemático. Además, los gráficos **b** y **c** son los característicos del caso de un esquema de autocorrelación según un proceso autorregresivo de orden 1 con parámetro positivo y negativo, respectivamente. En concreto, en el gráfico **b** observamos un comportamiento atenuado hacia cero, que, en este caso, es siempre negativo

(también podría ser positivo), y en el c, los residuos van alternando su signo formando lo que se llama *dientes de sierra*, y, en este caso, empieza por un valor positivo (aunque el primer residuo puede ser negativo en lugar de positivo).

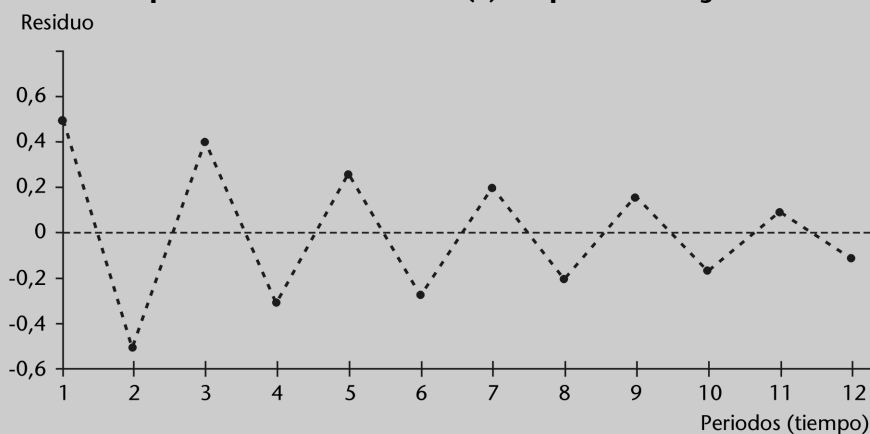
a. Situación de comportamiento aleatorio

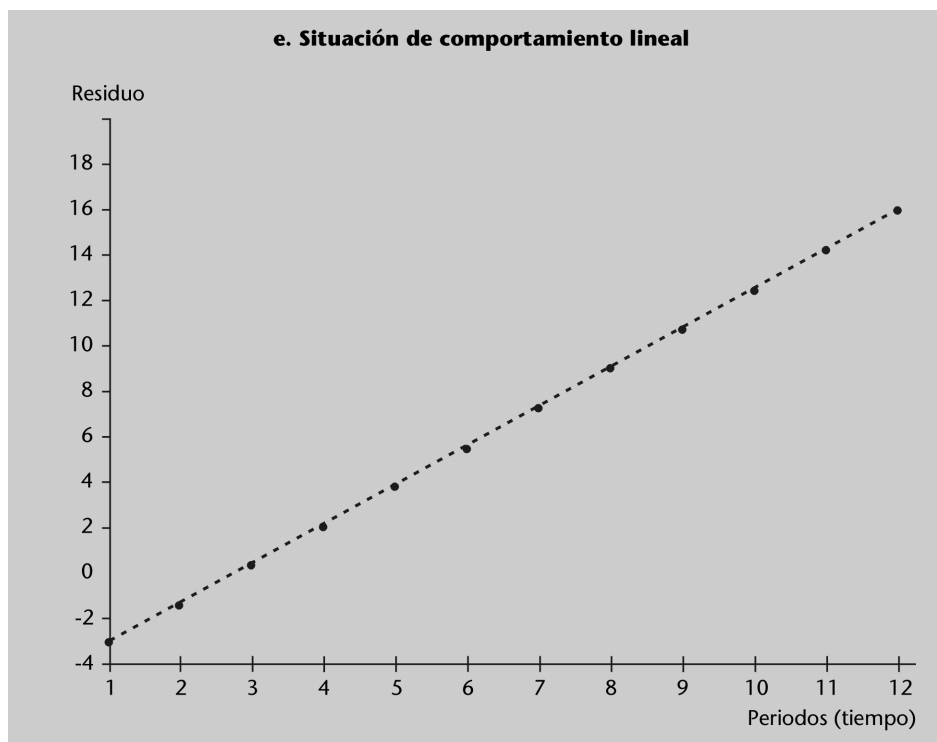
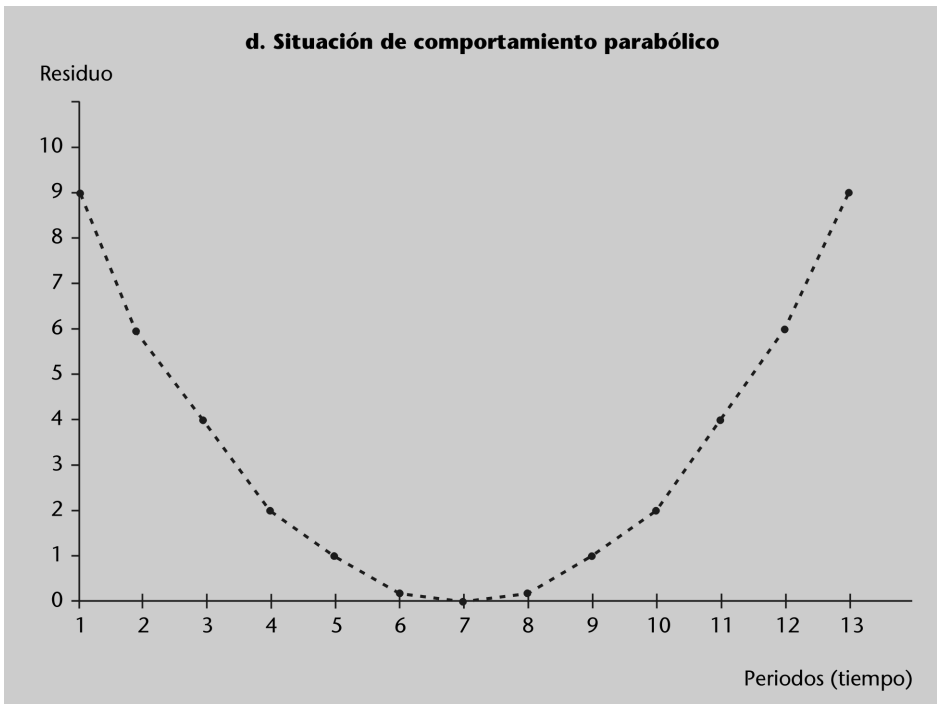


b. Esquema de autocorrelación AR(1) con parámetro positivo



c. Esquema de autocorrelación AR(1) con parámetro negativo





3.3.2. Contrastes de autocorrelación

En este subpartado estudiaremos diferentes test que han sido propuestos para contrastar la presencia de autocorrelación en el término de perturbación. En concreto, estudiaremos cuatro: el test de Durbin-Watson, el de la h de Durbin, el de Breusch-Godfrey, el de la Q de Box-Pierce y de Ljung-Box. !

Test de Durbin-Watson

Se trata de un test clásico que la mayor parte de paquetes estadísticoeconómicos incorporan por defecto al facilitar los resultados de la estimación de un modelo.

Dado un modelo de regresión, el test de Durbin-Watson permite contrastar únicamente si el término de perturbación del modelo está autocorrelacionado según un esquema AR(1).

La hipótesis nula del test es:

H_0 : no autocorrelación según AR (1) (o, de manera alternativa, $\gamma_1 = 0$, $\phi_1 = 0$ o $\rho_1 = 0$).

Es un test en una única cola, lo cual quiere decir que la hipótesis alternativa es:

$$H_A : u_t \sim \text{AR}(1) \text{ con } \phi_1 > 0,$$

o bien:

$$H_A : u_t \sim \text{AR}(1) \text{ con } \phi_1 < 0.$$

El estadístico de prueba, que simbolizaremos por DW , se define como vemos a continuación:

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}, \quad (3.9)$$

En la expresión 3.9,...

... el sumatorio del numerador empieza en $t = 2$ porque se pierde el residuo correspondiente al primer periodo, al estar retardando un periodo, e_{t-1} .

donde e_t simboliza el residuo de la estimación MCO del modelo de regresión del cual queremos contrastar si el término de perturbación sigue un comportamiento autorregresivo del tipo AR(1).

El estadístico de prueba, tal y como aparece registrado en 3.9, no sabemos entre qué valores está acotado. Para averiguarlo, desarrollando el cuadrado del numerador y operando, obtenemos:

$$\begin{aligned} DW &= \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = \frac{\sum_{t=2}^T e_t^2 + \sum_{t=2}^T e_{t-1}^2 - 2 \sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2} \cong \frac{2 \left(\sum_{t=1}^T e_t^2 - \sum_{t=2}^T e_t e_{t-1} \right)}{\sum_{t=1}^T e_t^2} = \\ &= 2 \left[1 - \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2} \right]. \end{aligned}$$

En el desarrollo de DW...

... hemos partido del hecho de que, si T es suficientemente grande, entonces:

$$\sum_{t=2}^T e_t^2 \cong \sum_{t=2}^T e_{t-1}^2 \cong \sum_{t=1}^T e_t^2.$$

Por otra parte, como sabéis, el coeficiente de autocorrelación simple de primer orden, ρ_1 , se define de la siguiente manera:

$$\rho_1 = \frac{\text{COV}[u_t, u_{t-1}]}{\sqrt{\text{VAR}[u_t]} \sqrt{\text{VAR}[u_{t-1}]}} = \frac{\gamma_1}{\gamma_0}.$$

Por lo tanto, una estimación de este coeficiente vendrá determinada por:

$$\hat{\rho}_1 = \frac{\text{COV}[e_t, e_{t-1}]}{\sqrt{\text{VAR}[e_t]} \sqrt{\text{VAR}[e_{t-1}]}} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2}. \quad (3.10)$$

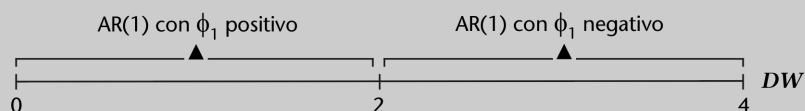
Así pues, si sustituimos 3.10 en la expresión del estadístico DW que hemos encontrado antes, tenemos:

$$DW \cong 2 \left[1 - \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2} \right] = 2(1 - \hat{\rho}_1). \quad (3.11)$$

La expresión 3.11 permite acotar el rango de valores que puede tomar el estadístico DW , ya que sólo depende de $\hat{\rho}_1$ y, como sabéis, $-1 \leq \hat{\rho}_1 \leq 1$. Por lo tanto, podemos distinguir las siguientes situaciones límite:

- Si $\hat{\rho}_1 = 1$, entonces $DW = 0$, lo cual quiere decir que, en el supuesto de máxima (perfecta) autocorrelación positiva, el estadístico DW valdrá 0.
- Si $\hat{\rho}_1 = -1$, entonces $DW = 4$, lo cual quiere decir que, en el supuesto de máxima (perfecta) autocorrelación negativa, el estadístico DW valdrá 4.
- Si $\hat{\rho}_1 = 0$, entonces $DW = 2$, por lo que, en el supuesto de ausencia de autocorrelación, el estadístico DW valdrá 2.

Podemos resumir el **rango de valores del estadístico DW** con el gráfico siguiente:



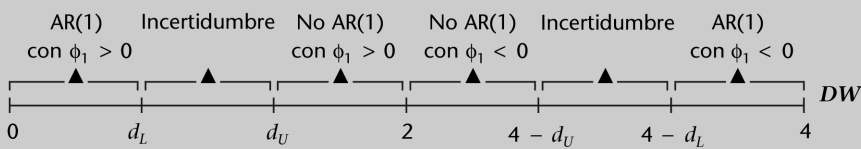
A pesar de lo que hemos comentado antes, dado un valor para el estadístico DW , necesitamos unos valores críticos que nos permitan afirmar si hay o no autocorrelación. En otras palabras, si, por ejemplo, el estadístico DW toma un valor de 2,3, ¿podemos afirmar que no hay autocorrelación según un esquema AR(1) con parámetro negativo? Y, de la misma manera, si toma un valor de 1,7, ¿podemos

Ved el coeficiente de autocorrelación simple en el subapartado 3.1.1 de este módulo didáctico.



afirmar que no hay autocorrelación según un esquema AR(1) con parámetro positivo? Para poder solucionar este problema Durbin y Watson tabularon dos cotas, d_U y d_L , que delimitan la zona de existencia y de no existencia de autocorrelación. Estas dos cotas dependen del tamaño muestral y del número de regresores del modelo de regresión que estamos estudiando si el término de perturbación presenta o no autocorrelación.

Entonces, para contrastar la hipótesis nula frente a la alternativa de autocorrelación positiva, emplearemos las **cotas de Durbin y Watson**, y para hacerlo frente a la alternativa de autocorrelación negativa utilizaremos $4 - d_U$ y $4 - d_L$, de acuerdo con el esquema siguiente:



A continuación, presentamos los inconvenientes del estadístico DW :

- Sólo es útil para contrastar si el término de perturbación está autocorrelacionado según un esquema AR(1). El problema es que, aunque no los hemos estudiado, hay otros esquemas de autocorrelación que no podríamos detectar con el uso de este test.
- Las cotas tabuladas por Durbin y Watson, d_U y d_L , sólo son estrictamente válidas si en el modelo de regresión inicial hay un término independiente.
- Presenta unas zonas de incertidumbre, de manera que, si el estadístico DW se encuentra entre d_L y d_U , o entre $4 - d_U$ y $4 - d_L$, no podremos tomar una decisión con respecto al contraste.
- Sólo es válido si todos los regresores del modelo son deterministas. Así pues, si entre los regresores hay alguna variable aleatoria (como, por ejemplo, la variable endógena retardada), el estadístico DW no es aplicable, ya que está sesgado hacia dos, con lo que tenderíamos a no rechazar la hipótesis nula cuando, en realidad, quizá sí deberíamos hacerlo.

Ejemplo de contraste por el test de Durbin-Watson

Supongamos que nos interesa estimar el siguiente modelo de regresión con datos anuales correspondientes al periodo 1960-1997:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \forall t = 1960, \dots, 1997.$$

En el cuadro de la página siguiente presentamos los resultados de la estimación por MCO de este modelo.

Estimación por MCO del modelo inicial

Variable dependiente: Y				
Número de observaciones: 38				
VARIABLE	COEFICIENTE	ERROR ESTAD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	3.0951487	0.6945175	4.4565453	0.0001
X2	2.9031776	0.9525440	3.0478147	0.0044
X3	0.8457306	0.7980491	1.0597476	0.2965
R-squared	0.217421	Mean of dependent var	4.892262	
Adjusted R-squared	0.172702	S.D. of dependent var	1.639200	
S.E. of regression	1.490950	Sum of squared resid	77.80257	
Log likelihood	-67.53484	F-statistic	4.861960	
Durbin-Watson stat	0.546676	Prob(F-statistic)	0.013701	

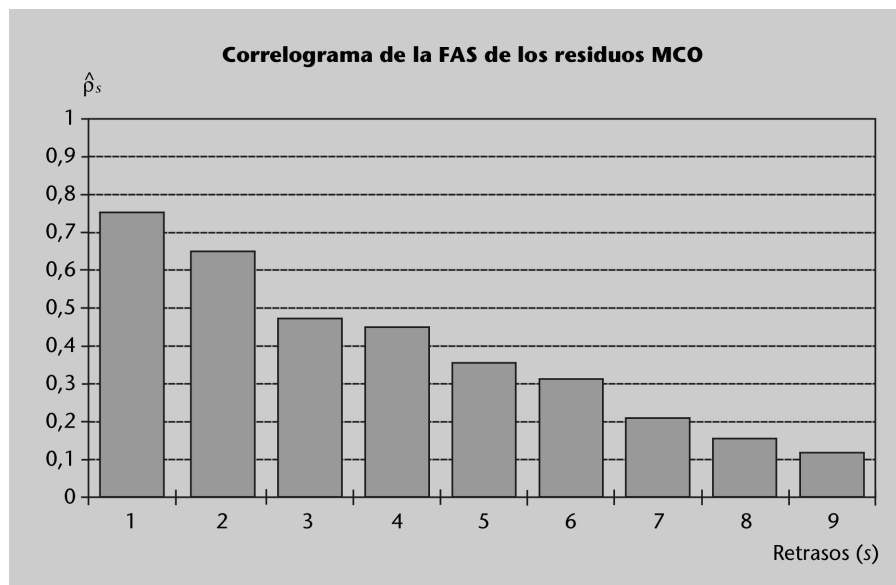
Como podemos ver, el estadístico Durbin-Watson toma un valor de 0,546676, que nos hace pensar que el término de perturbación del modelo está autocorrelacionado según un esquema autorregresivo de primer orden con parámetro positivo. Con el fin de confirmar esta sospecha, calculamos las cotas d_U y d_L tabuladas por Durbin y Watson, teniendo en cuenta que el número de observaciones es 38 y que hay dos regresores. Así pues, para un nivel de significación del 5%, estas cotas son 1,59 y 1,37, respectivamente. De acuerdo con lo que hemos comentado antes, y, dado que el estadístico de prueba (0,546676) es menor que la cota inferior ($d_L = 1,37$), podemos concluir que, en efecto, el término de perturbación sigue un proceso AR(1) con parámetro ϕ_1 positivo.

Además, si observamos el correlograma de la FAS (ved el gráfico 1 siguiente) asociada a los residuos de la estimación por MCO del modelo, vemos que presenta un comportamiento decreciente hacia cero, con todos los coeficientes de autocorrelación simple positivos. Esta situación se corresponde claramente con la que hemos presentado en el desarrollo teórico anterior.

Por otra parte, teniendo en cuenta el resultado registrado en 3.11, podemos calcular el valor del coeficiente de autocorrelación simple:

$$DW \cong 2(1 - \hat{\rho}_1) \Rightarrow \hat{\rho}_1 \cong 1 - \frac{DW}{2} = 1 - \frac{0,546676}{2} = 0,726662,$$

Gráfico 1



Lectura complementaria

Podéis encontrar las cotas de Durbin y Watson tabuladas en varios libros de los recomendados en la bibliografía, por ejemplo, en la siguiente obra:

A. Novales (1993).
Econometría (2a ed.).
Madrid: McGraw-Hill.

y, teniendo en cuenta que en un proceso AR(1) se cumple que* $\rho_1 = \phi_1$, podemos escribir que el esquema autorregresivo que sigue el término de perturbación es:

$$u_t = 0,726662u_{t-1} + \varepsilon_t.$$

*** Recordad que en un proceso AR(1) siempre se cumple que $\rho_s = \phi_1^s$.**

Imaginamos, sin embargo, que en lugar de especificar el modelo inicial especificamos un segundo en el que incorporamos entre los regresores la variable endógena retardada un periodo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 Y_{t-1} + v_t \quad \forall t = 1961, \dots, 1997,$$

y también la estimamos por MCO utilizando las mismas observaciones que en un principio (notamos, sin embargo, que en este segundo modelo empiezan en $t = 1961$ y no en $t = 1960$, puesto que, al estar la variable endógena retardada un periodo, se pierde la primera observación). Los resultados obtenidos los presentamos en el siguiente cuadro:

Gráfico 2

Estimación por MCO del modelo modificado				
Variable dependiente: Y				
Número de observaciones: 37				
VARIABLE	COEFICIENTE	ERROR ESTAD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	0.3251263	0.6803246	0.4778988	0.6359
X2	2.2437926	0.6928999	3.2382639	0.0027
X3	0.7166143	0.5665370	1.2649028	0.2148
Y (-1)	0.6468532	0.1070941	6.0400435	0.0000
R-squared	0.623803	Mean of dependent var	4.925158	
Adjusted R-squared	0.589603	S.D. of dependent var	1.649045	
S.E. of regression	1.056415	Sum of squared resid	36.82844	
Log likelihood	-52.41475	F-statistic	18.23998	
Durbin-Watson stat	2.072369	Prob(F-statistic)	0.000000	

Nota: observad que en el cuadro se ha perdido la primera observación al incluir como explicativa la variable endógena retardada un periodo.

Podemos observar cómo el estadístico DW toma un valor muy próximo a 2 (en concreto 2,072369). Este resultado nos llevaría, sin duda, a la conclusión de que el término de perturbación no presenta ningún esquema de autocorrelación según un proceso AR(1), y que sabemos que es absolutamente falsa, como hemos podido comprobar antes.

¿Por qué llegamos a esta conclusión errónea? La causa la hemos apuntado antes cuando hablábamos de los inconvenientes del test de Durbin-Watson. En concreto, nos referimos al último de los inconvenientes señalados: es la presencia de la variable endógena retardada la que provoca el sesgo hacia el valor 2 del estadístico de prueba DW.

Test de la h de Durbin

Con el fin de solucionar el inconveniente del test de Durbin-Watson, Durbin propuso un test que permite contrastar la presencia de autocorrelación según un esquema AR(1) en el término de perturbación de un modelo de regresión, cuando en éste se encuentra entre los regresores la variable endógena retardada (uno o más periodos). El mencionado contraste se denomina **contraste h de Durbin**. Así, dado el siguiente modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_r Y_{t-r} + u_t \quad (3.12)$$

para contrastar la hipótesis nula, que especificamos como se sigue:

$$H_0 : u_t \text{ no presenta autocorrelación según un esquema AR(1),}$$

frente a la alternativa:

$$H_A : u_t \text{ presenta autocorrelación según un esquema AR}(1),$$

estimamos por MCO en el modelo 3.12 y, acto seguido, obtenemos el estadístico de prueba que denotaremos por h y que está determinado por la siguiente expresión:

$$h = \hat{\phi}_1 \sqrt{\frac{T}{1 - TV\hat{A}R[\hat{\alpha}_1]}}, \quad (3.13)$$


en el que T es el tamaño de la muestra y $V\hat{A}R(\hat{\alpha}_1)$, la varianza estimada del estimador asociado al primer retardo de la variable endógena.

Si la muestra es grande en la medida suficiente, con la hipótesis nula, el estadístico 3.13 se distribuye según una normal con valor esperado cero y varianza unitaria:

$$h \sim N[0, 1] \text{ si el tamaño de la muestra es grande (es decir, asintóticamente).}$$

De acuerdo con lo que hemos visto antes, la regla de decisión por el **test de la h de Durbin** es la siguiente:

- Si el valor del estadístico de prueba es mayor o igual que el valor de la tabla de la $N[0,1]$ para un nivel de significación prefijado (habitualmente del 5%), rechazaremos la hipótesis nula, lo cual quiere decir que el término de perturbación presenta autocorrelación según un esquema AR(1).
- En caso contrario, concluiremos que no presenta autocorrelación según un esquema AR(1).

Tened en cuenta que, como en el caso del test de Durbin-Watson, el contraste se puede hacer en una cola según el signo de contraste y, por lo tanto, en caso de rechazar la hipótesis nula, podíamos saber si el parámetro del esquema autorregresivo era positivo o negativo. 


Ejemplo de contraste por el test de la h de Durbin

Reanudando el ejemplo anterior, calcularemos el estadístico h de Durbin con el fin de comprobar si el término de perturbación está autocorrelacionado según un esquema AR(1):

$$\begin{aligned} h &= \hat{\phi}_1 \sqrt{\frac{T}{1 - TV\hat{A}R[\hat{\alpha}_1]}} = -0,03618 \sqrt{\frac{37}{1 - 37(0,1070941)^2}} = \\ &= -0,03618 \sqrt{\frac{37}{0,575641588}} = -0,29. \end{aligned}$$

Notad que en la expresión 3.13,...

... en lugar de $\hat{\phi}_1$ podríamos considerar $\hat{\rho}_1$, teniendo en cuenta que en el esquema AR(1), tal y como hemos visto, siempre se cumple que $\hat{\rho}_1 = \hat{\phi}_1$. Por otra parte, notad, también, que en el estadístico de prueba sólo se considera la varianza estimada asociada al estimador del primer retardo de la variable endógena, aunque entre los regresores del modelo aparezcan más retardos (ved el modelo especificado en 3.12 y cómo hemos definido el estadístico de prueba 3.13).

Ved el gráfico 2 en "Ejemplo de contraste por el test de Durbin-Watson" de este mismo subapartado. 

El resultado obtenido por el estadístico de prueba $(-0,29)$ muestra que el término de perturbación no sigue un proceso autorregresivo de orden 1, debido a que el valor crítico en tablas en el 5% de significación es 1,64.

A continuación, enunciaremos los inconvenientes del test de la h de Durbin: 

- a) El test sólo es estrictamente válido asintóticamente.
- b) En el supuesto de que $TV\hat{AR}[\hat{\alpha}_1]$ sea mayor que 1, entonces en el estadístico de prueba 3.13 tendríamos la raíz cuadrada de un número negativo, con lo que el valor del estadístico de prueba h sería un número complejo y, en consecuencia, no podríamos compararlo con los valores de la tabla de la $N[0,1]$; por lo tanto, no podríamos llegar a ninguna conclusión.

Para solucionar este problema, se ha propuesto un procedimiento equivalente que consta de cuatro etapas:

- Estimar por MCO el modelo 3.12.
- Obtener el vector de residuos MCO asociados a la estimación efectuada en la etapa anterior, e .
- Especificar un modelo de regresión auxiliar en el que la variable endógena es el vector de residuos MCO obtenido en la etapa anterior y las variables explicativas son todas las del modelo 3.12, y, además, los residuos MCO retardos un periodo. En nuestro ejemplo esto es el siguiente modelo:

$$e_t = \delta_1 + \delta_2 X_{2t} + \delta_3 X_{3t} + \dots + \delta_k X_{kt} + \delta_{k+1} Y_{t-1} + \delta_{k+2} Y_{t-2} + \dots + \delta_{k+r} Y_{t-r} + \delta_{k+r+1} e_{t-1} + v_t.$$

- Estimar por MCO el modelo especificado en la etapa anterior y contrastar, de la manera habitual, la significación individual del parámetro asociado al residuo retardado un periodo, en nuestro caso δ_{k+r+1} . En caso de rechazar la hipótesis nula, es decir, si el parámetro δ_{k+r+1} es estadísticamente distinto de cero, podremos concluir que el término de perturbación del modelo 3.12 presenta autocorrelación según un esquema AR(1) y, en caso contrario, que no presenta autocorrelación.

Test de Breusch-Godfrey

Este test, a diferencia de los dos que hemos estudiado hasta ahora, es más general puesto que nos permite contrastar cualquier esquema de autocorrelación. De forma intuitiva, la idea que subyace en este test es la siguiente: si el término de perturbación no presenta ningún tipo de autocorrelación, entonces todos los coeficientes de autocorrelación simple*, $\rho_1, \rho_2, \rho_3, \dots$, serán cero. En consecuencia, se trata de contrastar la siguiente hipótesis nula:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0 \quad (\Rightarrow \text{no autocorrelación de orden } p),$$

* Los coeficientes de autocorrelación simple son los elementos de fuera de la diagonal principal de la matriz 3.7.

donde p es el orden de la autocorrelación que deseamos contrastar, frente a la hipótesis alternativa:

$$H_A : H_0 (\Rightarrow \text{autocorrelación de orden } p \text{ o inferior}).$$

Así pues, una vez especificadas las hipótesis nula y alternativa, las etapas que debemos seguir para llevar a cabo el contraste de Breusch-Godfrey son las cinco que hallaremos apuntadas a continuación:

- a) Estimar por MCO el modelo de regresión del que queremos averiguar si el término de perturbación presenta o no algún tipo de autocorrelación.
- b) Obtener el vector de residuos asociado a la estimación anterior, e .
- c) Especificar un modelo de regresión auxiliar en el que la variable endógena son los residuos obtenidos en la etapa anterior, e , y las variables explicativas son, por una parte, todas las variables explicativas que hay en el modelo original* y, por otra, los residuos MCO obtenidos en la etapa anterior retardados tantas veces como el orden de la autocorrelación que deseamos contrastar (tantos como los que se consideran en la hipótesis nula).
- d) Estimar por MCO el modelo especificado en la etapa anterior y obtener el coeficiente de determinación, R^2 , asociado a esta estimación.
- e) Calcular el estadístico de prueba, que se define así:

$$T' \cdot R^2$$

donde T' es el número de observaciones con el que hemos realizado la estimación del modelo de regresión auxiliar en la etapa anterior (el tamaño de la muestra de la que disponíamos en un principio menos el número de retardos de los residuos, p , considerados en el modelo auxiliar: $T - p$), y R^2 es el coeficiente de determinación obtenido en la etapa anterior. Este estadístico de prueba se distribuye según una χ^2 -cuadrado con tantos grados de libertad como coeficientes de autocorrelación simple hayamos considerado en la hipótesis nula:

$$T' \cdot R^2 \sim \chi_p^2. \quad (3.14)$$

Entonces, la regla de decisión en el **test de Breusch-Godfrey** es la siguiente:

- Si $T' \cdot R^2 \geq \chi_{p;\alpha}^2 \Rightarrow$ rechazamos la $H_0 \Rightarrow u_t$ presenta autocorrelación de orden p o inferior.
- Si $T' \cdot R^2 < \chi_{p;\alpha}^2 \Rightarrow$ no rechazamos la $H_0 \Rightarrow u_t$ no presenta autocorrelación de orden p o inferior.

* El modelo original es el modelo del que nos interesa contrastar la autocorrelación del término de perturbación.

Ejemplo de contraste por el test de Breusch-Godfrey

Siguiendo con los mismos datos y el mismo modelo que en los ejemplos anteriores, a continuación mostraremos los resultados de aplicar el test de Breusch-Godfrey. En concreto, en este caso, las hipótesis nulas y alternativas son:

$$H_0 : \rho_1 = 0 \ (\Rightarrow \text{no autocorrelación de primer orden}),$$

$$H_A : \rho_1 \neq 0 \ (\Rightarrow \text{autocorrelación de primer orden}).$$

En el siguiente cuadro presentamos los resultados de la estimación de la regresión auxiliar definida en la etapa c.

Ved "Ejemplo de contraste por el test de Durbin-Watson" y "Ejemplo de contraste por el test de Durbin" de este subapartado.


Estimación por MCO del modelo auxiliar				
Variable dependiente: Residuo				
Número de observaciones: 37				
VARIABLE	COEFICIENTE	ERROR ESTAD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	0.0045234	0.5601059	0.0080760	0.9936
X2	0.0286632	0.7682217	0.0373111	0.9705
X3	-0.1691057	0.6447201	-0.2622933	0.7947
Residuo (-1)	0.5310378	0.1192996	4.4512952	0.0001
R-squared	0.368194	Mean of dependent var	2.45E-09	
Adjusted R-squared	0.312447	S.D. of dependent var	1.450094	
S.E. of regression	1.202401	Sum of squared resid	49.15609	
Log likelihood	-58.81054	F-statistic	6.604676	
Durbin-Watson stat	1.706759	Prob(F-statistic)	0.001229	

Nota: observad que en el cuadro, al incluir como variable explicativa el primer retardos de los residuos MCO, se ha perdido la primera observación.

A partir de los resultados que aparecen en el cuadro, podemos calcular el estadístico de prueba:

$$T^n \cdot R^2 = 37 \cdot 0,368194 = 13,623178.$$

Teniendo en cuenta que el valor crítico de una χ^2 -cuadrado con un grado de libertad para un nivel de significación del 5% es 3,84, rechazamos la hipótesis nula, hecho que confirma que el término de perturbación presenta autocorrelación de primer orden.

Para acabar con el test de Breusch-Godfrey, nos queda por señalar dos cuestiones: por una parte, que sólo es válido asintóticamente y, por la otra, que podemos aplicarlo con independencia de si en el modelo del que queremos contrastar la posible presencia de autocorrelación en el término de perturbación hay regresores aleatorios o no. En cualquier caso, sin embargo, en el supuesto de que todos los regresores sean deterministas, es demostrable que el estadístico de prueba 3.14 se puede aproximar de la siguiente forma: 

$$T^n \cdot R^2 = T^n \sum_{s=1}^p \hat{\rho}_s^2 = T^n (\hat{\rho}_1^2 + \hat{\rho}_2^2 + \dots + \hat{\rho}_p^2), \tag{3.15}$$

donde $\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} = \frac{\text{COV}[e_t, e_{t-s}]}{\text{VAR}[e_t]}$.

Test de la Q de Box-Pierce y de Ljung-Box

El test Q propuesto por Box y Pierce permite contrastar si los primeros p coeficientes de autocorrelación simple del término de perturbación son nulos desde un punto de vista estadístico. En consecuencia, de la misma manera que el test

de Breusch-Godfrey, se trata de un test lo suficientemente general, ya que no se limita en los procesos autorregresivos.

Las hipótesis nulas y alternativa del test de la Q de Box-Pierce son:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0 \quad (\Rightarrow \text{no autocorrelación de orden } p \text{ o inferior}),$$

$$H_A : \text{no } H_0 \quad (\Rightarrow \text{autocorrelación de orden } p \text{ o inferior}),$$

y el estadístico de prueba, que se basa en los coeficientes de autocorrelación simple estimados, se define como vemos:

$$Q_p = T \sum_{s=1}^p \hat{\rho}_s^2,$$

$$\text{donde } \hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} = \frac{\text{COV}[e_t, e_{t-s}]}{\text{VAR}[e_t]}.$$

Con la hipótesis nula, el estadístico Q_p se distribuye, asintóticamente, según una χ -cuadrado con p grados de libertad.


La regla de decisión en el **test de la Q de Box-Pierce** es, por lo tanto, la que tenemos a continuación:

- Si $Q_p \geq \chi_{p;\alpha}^2 \Rightarrow$ rechazamos la $H_0 \Rightarrow u_t$ presenta autocorrelación de orden p .
- Si $Q_p < \chi_{p;\alpha}^2 \Rightarrow$ no rechazamos la $H_0 \Rightarrow u_t$ no presenta autocorrelación de orden p .

Nota: el test Q_p es equivalente al de Breusch-Godfrey, que aparece en la expresión 3.15.

Para finalizar, por lo que respecta al **test de la Q de Ljung-Box**, tan sólo nos interesa señalar que no es más que una corrección del test de la Q de Box-Pierce. En consecuencia, las hipótesis nulas y alternativas son las mismas. El estadístico de prueba se define de la siguiente forma:

$$Q'_p = T(T+2) \sum_{s=1}^p \frac{\hat{\rho}_s^2}{T-s},$$

donde T es el tamaño de la muestra y $\hat{\rho}_s = \frac{\hat{\gamma}_s}{\hat{\gamma}_0} = \frac{\text{COV}[e_t, e_{t-s}]}{\text{VAR}[e_t]}$. Como es lógico, este estadístico también se distribuye según una χ -cuadrado con p grados de libertad y la regla de decisión es la misma que en el caso del test de la Q de Box-Pierce. 

Ejemplos de contraste por los tests de Box-Pierce y de Ljung-Box

Con los mismos datos de los ejemplos que hemos desarrollado a lo largo de las páginas anteriores, hemos calculado el valor del estadístico Q de Box-Pierce y de Ljung-Box, teniendo en cuenta el primer coeficiente de autocorrelación simple estimado, que vale 0,726662. Los estadísticos de prueba obtenidos han sido los siguientes:

- $Q_p = T \sum_{s=1}^p \hat{\rho}_s^2 = 38 \cdot 0,726662^2 = 20,0654.$
- $Q'_p = T(T+2) \sum_{s=1}^p \frac{\hat{\rho}_s^2}{T-s} = 38(38+2) \frac{0,726662^2}{38-1} = 21,69236.$

Al comparar estos valores con el de una χ -cuadrado con un grado de libertad y un nivel de significación del 5% (3,84), llegamos a la conclusión, una vez más, de que el término de perturbación presenta autocorrelación de primer orden.

Si nos interesase contrastar la hipótesis nula:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_{12} = 0 \quad (\Rightarrow \text{no autocorrelación de orden 12 o inferior}),$$

frente a la hipótesis alternativa:

$$H_A : \text{no } H_0 \text{ autocorrelación de orden 12 o inferior),}$$

los estadísticos de prueba serían, para el caso del test de Box-Pierce, 74,52, y para el de Ljung-Box, 86,86. Al comparar estos valores con el de una χ -cuadrado con doce grados de libertad y un nivel de significación del 5% (que corresponde a 21,0), llegamos a la conclusión de que el término de perturbación presenta autocorrelación de orden 12 o inferior (en este caso sabemos, sin embargo, que es de orden 1).

3.4. Estimación por MCG

Como ya sabemos, si el término de perturbación presenta autocorrelación (y, en general, cuando es no esférico), los estimadores MCO cumplen las propiedades que mencionamos a continuación:

Podéis ver las propiedades de los estimadores MCO cuando el término de perturbación es no esférico en el subapartado 1.2 de este módulo didáctico.

a) Por lo que respecta a los parámetros **B**, son no sesgados, consistentes, pero no eficientes.

b) En cuanto a la varianza del término de perturbación, es sesgada.


Teniendo en cuenta lo que hemos dicho antes, y de acuerdo con lo que hemos estudiado en el primer apartado de este módulo didáctico, según las circunstancias en las que nos encontremos, tendremos que utilizar el método de MCG, debido a que éste nos permite superar las carencias que presenta el método de MCO, en el sentido de que para los parámetros **B** proporciona estimadores de varianza mínima (además de no sesgados y consistentes), y para la varianza del término de perturbación ofrece una estimación no sesgada.

Por otra parte, como también sabréis a estas alturas, para poder aplicar MCG tenemos que conocer la matriz Ω . Así, en el supuesto que hemos considerado a lo largo de este subapartado, si el término de perturbación presenta autocorrelación según un esquema AR(1), la matriz de varianzas y covarianzas viene determinada por:

$$\begin{aligned} \text{VAR}[U] &= \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)} \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix} = \\ &= \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)} \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{T-1} & \phi_1^{T-2} & \phi_1^{T-3} & \dots & 1 \end{bmatrix}; \end{aligned}$$

por lo que la matriz Ω es:

$$\Omega = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{bmatrix} = \begin{bmatrix} 1 & \phi_1 & \phi_1^2 & \dots & \phi_1^{T-1} \\ \phi_1 & 1 & \phi_1 & \dots & \phi_1^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_1^{T-1} & \phi_1^{T-2} & \phi_1^{T-3} & \dots & 1 \end{bmatrix}.$$

Así pues, en función del camino elegido para aplicar MCG, podemos demostrar que las matrices de interés Ω^{-1} y $T = P^{-1}$ (siendo $\Omega = PP'$) están determinadas por: 

$$\Omega^{-1} = \frac{1}{1 - \phi_1^2} \begin{bmatrix} 1 & -\phi_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -\phi_1 & 1 + \phi_1^2 & -\phi_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\phi_1 & 1 + \phi_1^2 & -\phi_1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\phi_1 & 1 + \phi_1^2 & -\phi_1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -\phi_1 & 1 \end{bmatrix},$$


$$T = \begin{bmatrix} \sqrt{1 - \phi_1^2} & 0 & 0 & 0 & \dots & 0 & 0 \\ -\phi_1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & -\phi_1 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & -\phi_1 & 1 \end{bmatrix}.$$

Apreciad, por lo tanto, que, al multiplicar la matriz X y los vectores Y y U por esta matriz de transformación T , transformamos las variables como mostramos a continuación:

$$Y_t^* = (Y_t - \phi_1 Y_{t-1}), X_t^* = (X_t - \phi_1 X_{t-1}), \text{ i } u_t^* = (u_t - \phi_1 u_{t-1}),$$

excepto Y_1^* i X_1^* . De acuerdo con la transformación anterior, el término de perturbación u_t^* se supone según una variable ruido blanco, es decir, no presenta autocorrelación.

Notad, sin embargo, que en el caso anterior suponemos que el parámetro del esquema autorregresivo, ϕ_1 (o, lo que es lo mismo, el coeficiente de autocorrelación simple de primer orden, ρ_1), es conocido. No obstante, ésta no es la situación más habitual en la práctica, sino todo lo contrario, por norma general este parámetro es desconocido. Así que, ante tal situación, tenemos que utilizar otros métodos de estimación que permitan superar este problema.

En este sentido, de entre los métodos que hay, estudiaremos dos: el método de Cochrane-Orcutt y el de Durbin. 

Para obtener los estimadores de MCG,...

... recordad que tenemos dos alternativas:

a) Podemos utilizar las siguientes expresiones:

$$\bullet \hat{B}_{MCG} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y),$$

$$\bullet \hat{\sigma}_{MCG}^2 = \frac{e' \Omega^{-1} e}{N - k}.$$

b) O, de forma alternativa, usar estas otras expresiones:

$$\bullet \hat{B}_{MCG} = (X^{*'} X^*)^{-1} (X^{*'} Y^*),$$

$$\bullet \hat{\sigma}_{MCG}^2 = \frac{e^{*'} e^*}{N - k},$$

donde $X^* = TX$, $Y^* = TY$ y $e^* = Te$.

Otros métodos de estimación...

... que podríamos utilizar son, por ejemplo, el de máxima verosimilitud y el de mínimos cuadrados no lineales; pero, como exceden el contenido de este curso, no los estudiaremos.


3.5. Estimación por los métodos Cochrane-Orcutt y Durbin

Estos dos métodos de estimación están basados en la misma idea. De forma intuitiva, consisten en transformar el modelo inicial con perturbaciones autocorrelacionadas según un esquema AR(1):

$$Y = XB + U, \text{ donde } U \sim \text{AR}(1),$$


en otro modelo con perturbaciones no autocorrelacionadas:

$$Y^* = X^*B + U^* \text{ donde } U^* \sim \text{ruido blanco.}$$

Apreciad, entonces, que, de hecho, estos dos métodos de estimación no son más que una variante del método de MCG (o MCP). 

Así, suponemos que el modelo que deseamos estimar es el siguiente:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + u_t \quad \forall t = 1, 2, \dots, T, \quad (3.16)$$

donde el término de perturbación sigue un proceso autorregresivo de orden 1: $u_t \sim \text{AR}(1)$. En ambos métodos tenemos que seguir los siguientes pasos: 

1) El primer paso consiste en retardar un periodo el modelo y multiplicar las expresiones por el coeficiente de autocorrelación simple:

$$\phi_1 Y_{t-1} = \phi_1 \beta_1 + \phi_1 \beta_2 X_{2t-1} + \dots + \phi_1 \beta_k X_{kt-1} + \phi_1 u_{t-1} \quad \forall t = 2, 3, \dots, T. \quad (3.17)$$

2) Acto seguido, calculamos la diferencia entre los dos modelos anteriores:

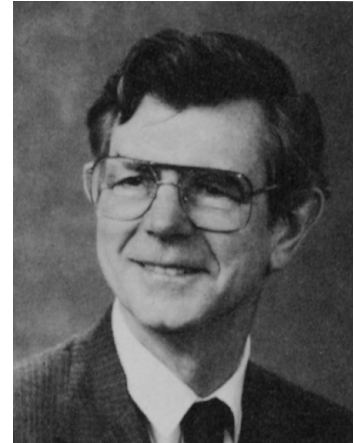
$$Y_t - \phi_1 Y_{t-1} = (1 - \phi_1)\beta_1 + \beta_2(X_{2t} - \phi_1 X_{2t-1}) + \dots + \beta_k(X_{kt} - \phi_1 X_{kt-1}) + (u_t - \phi_1 u_{t-1}) \quad \forall t = 2, 3, \dots, T. \quad (3.18)$$

De esta manera, llegamos al modelo 3.18, en el que el término de perturbación, $u_t - \phi_1 u_{t-1}$, es una variable aleatoria ruido blanco y, en consecuencia, no presenta autocorrelación.

Así pues, en lo que concierne al término de perturbación, no hay nada que impida que, al estimar el modelo 3.18 por MCO, obtengamos estimadores que cumplan todas las propiedades deseables. De todos modos, sin embargo, aunque lo anterior en efecto sea cierto, si observamos el modelo mencionado, vemos que es un modelo no lineal. Para ver este hecho con mayor claridad únicamente tenemos que desarrollar el modelo 3.18:

$$Y_t = \phi_1 Y_{t-1} + \beta_1 - \beta_1 \phi_1 + \beta_2 X_{2t} - \beta_2 \phi_1 X_{2t-1} + \dots + \beta_k X_{kt} - \beta_k \phi_1 X_{kt-1} + (u_t - \phi_1 u_{t-1}) \quad \forall t = 2, 3, \dots, T. \quad (3.19)$$

En consecuencia, no podemos estimar el modelo 3.19 (o 3.18) por MCO directamente, puesto que antes de proceder a ello tenemos que solucionar el proble-



Guy Orcutt

Guy Orcutt nació en Detroit en el año 1917. Se graduó en física y matemáticas por la Universidad de Michigan. En el año 1939 se interesó por la economía. Se matriculó en Económicas y se graduó. En 1949 publicó, junto con Donald Cochrane, de la Universidad de Cambridge, el método de estimación que había desarrollado para modelos con autocorrelación.

Para comprobar que...


... el término de perturbación del modelo 3.18 es una variable aleatoria ruido blanco, sólo tenemos que recordar la expresión de un término de perturbación que sigue un esquema AR(1):

$$u_t = \phi_1 u_{t-1} + \varepsilon_t$$

Si en esta expresión aislamos la variable ruido blanco, obtenemos:

$$\begin{aligned} \varepsilon_t &= u_t - \phi_1 u_{t-1} = \\ &= u_t - \rho_1 u_{t-1}, \end{aligned}$$

que es precisamente el término de perturbación que aparece en el modelo 3.18.

ma de la no linealidad de algunos de los parámetros que aparecen en éste. En este punto será donde, por una parte Cochrane y Orcutt, y, por la otra Durbin, proponen distintas soluciones. 

En concreto, el **método de Cochrane-Orcutt** propone seguir la estrategia que podéis ver a continuación:

1) Estimar por MCO el modelo inicial 3.16 sin tener en cuenta el hecho de que el término de perturbación sigue un esquema AR(1).

2) Obtener el vector de residuos asociados a la estimación efectuada en la etapa anterior y, a partir de estos residuos, obtener una estimación del coeficiente de autocorrelación simple de primer orden, $\hat{\phi}_1$. Para hacer esto, contamos con tres alternativas posibles, y son éstas:

a) Obtener $\hat{\phi}_1$ como el coeficiente de correlación entre e_t y e_{t-1} :

$$\hat{\phi}_1 = \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=1}^T e_t^2}.$$

b) Obtener $\hat{\phi}_1$ a partir del estadístico *DW* asociado a la estimación efectuada en la primera etapa:

$$DW = 2(1 - \hat{\phi}_1) \Rightarrow \hat{\phi}_1 = 1 - \frac{DW}{2}. \text{ Por lo tanto, } \hat{\phi}_1 = \hat{\rho}_1.$$

c) Obtener $\hat{\phi}_1$ como el estimador MCO de la siguiente regresión:

$$e_t = \phi_1 e_{t-1} + v_t \Rightarrow \hat{\phi}_{1\text{MCO}} = \frac{\sum_{t=2}^T e_t e_{t-1}}{\sum_{t=2}^T e_{t-1}^2}.$$

Podemos ver que la estimación del parámetro ϕ_1 , siguiendo esta tercera opción, es tanto más válida cuanto mayor sea el tamaño de la muestra de que disponemos, dado que, si T es grande en la medida suficiente, entonces se cumple la expresión que mostramos ahora:

$$\sum_{t=2}^T e_{t-1}^2 \cong \sum_{t=1}^T e_t^2.$$

3) Sustituir, en el modelo 3.18, el parámetro ϕ_1 por la estimación encontrada en la etapa anterior. De esta manera, llegamos al siguiente modelo, que ya podemos estimar por MCO:

$$Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + \dots + \beta_k X_{kt}^* + \hat{\varepsilon}_t \quad \forall t = 2, 3, \dots, T, \quad (3.20)$$

donde $Y_t^* = Y_t - \hat{\phi}_1 Y_{t-1}$, $\beta_1^* = \beta_1(1 - \hat{\phi}_1)$, $X_{2t}^* = X_{2t} - \hat{\phi}_1 X_{2t-1}$, $X_{kt}^* = X_{kt} - \hat{\phi}_1 X_{kt-1}$, y $\hat{\varepsilon}_t = u_t - \hat{\phi}_1 u_{t-1}$.

Por su parte, el **método de Durbin** propone la siguiente estrategia:

1) Reparametrizar el modelo 3.19 como se sigue:


$$Y_t = \phi_1 Y_{t-1} + \alpha_1 + \alpha_{21} X_{2t} + \alpha_{22} X_{2t-1} + \dots + \alpha_{k1} X_{kt} + \alpha_{k2} X_{kt-1} + \varepsilon_t$$

$$\forall t = 2, 3, \dots, T, \quad (3.21)$$

donde $\alpha_1 = \beta_1(1 - \phi_1)$, $\alpha_{21} = \beta_2$, $\alpha_{22} = \beta_2\phi_1$, $\alpha_{k1} = \beta_k$, y $\alpha_{k2} = \beta_k\phi_1$. De esta manera, podemos considerar que el modelo 3.21 es lineal, aunque en realidad no lo es, y, por lo tanto, podemos estimarlo por MCO, y de los estimadores obtenidos tan sólo quedarnos con el del parámetro ϕ_1 .

2) Sustituir la estimación encontrada en la etapa anterior por el parámetro ϕ_1 en el modelo 3.18, con lo que llegaremos al modelo 3.20, que ya podemos estimar por MCO.

No podemos dejar de tener en cuenta que, como se deduce de todo lo anterior, de hecho, la única diferencia que hay entre los métodos propuestos por Cochrane y Orcutt y por Durbin es la forma de estimar previamente el coeficiente del esquema autorregresivo de primer orden ($\rho_1 = \phi_1$) con el fin de introducirlo, más tarde, en el modelo.

En cualquier caso, podemos demostrar que, si en lugar de hacer una única iteración, como hemos explicado aquí, hacemos más, los resultados que obtendremos con ambos procedimientos convergen en un mismo valor. 

Ejemplo de estimación por el método de Cochrane-Orcutt

Con los mismos datos que hemos utilizado a lo largo de los ejemplos anteriores, hemos estimado por el método de Cochrane-Orcutt el modelo de regresión con el que trabajamos. El cuadro que hallamos a continuación presenta los resultados de la estimación. Veámoslos:

Resultados de la estimación por el método de Cochrane-Orcutt

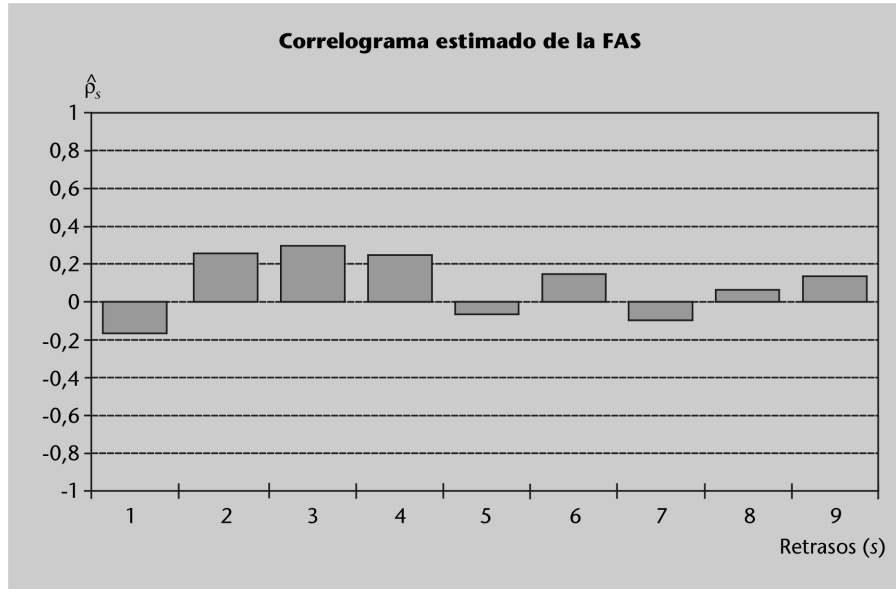
Variable dependiente: Y				
Número de observaciones: 37				
Convergencia lograda tras cuatro iteraciones				
VARIABLE	COEFICIENTE	ERROR ESTAD.	ESTADÍSTICO T	SIGNIFICACIÓN
C	3.5408558	0.7834544	4.5195428	0.0001
X2	1.8740922	0.5872368	3.1913738	0.0031
X3	1.0380681	0.4657797	2.2286678	0.0328
AR(1)	0.7594713	0.1123693	6.7587058	0.0000
R-squared	0.657669	Mean of dependent var	4.925158	
Adjusted R-squared	0.626548	S.D. of dependent var	1.649045	
S.E. of regression	1.007743	Sum of squared resid	33.51302	
Log likelihood	-50.66952	F-statistic	21.13267	
Durbin-Watson stat	2.269272	Prob(F-statistic)	0.000000	

En este cuadro podemos ver que el valor estimado por el parámetro $\rho_1 (= \phi_1)$ es 0,7594713 (que se encuentra muy cerca del que habíamos encontrado en el “Ejemplo de contraste por el test de Durbin-Watson” y que es estadísticamente significativo.

Por otra parte, comprobamos que el valor estimado para la constante (3,5408558) no es el del término independiente del modelo inicial (3.16), sino que es el que corresponde al modelo 3.20. Así pues, para encontrar la estimación del término independiente del modelo inicial, tenemos que realizar la siguiente operación:

$$\hat{\beta}_1^* = \hat{\beta}_1(1 - \hat{\phi}_1) \Rightarrow \hat{\beta}_1 = \frac{\hat{\beta}_1^*}{(1 - \hat{\phi}_1)} = \frac{3,5408558}{(1 - 0,7594713)} = 14,7211364.$$

Además, en el siguiente gráfico presentamos el correlograma estimado de la FAS obtenido a partir de los residuos de la estimación anterior, en el que podemos observar que no hay ningún comportamiento de tipo autorregresivo de orden 1.



De todos modos, para comprobar si estadísticamente estos coeficientes de autocorrelación simple estimados obtenidos a partir de la serie de residuos asociada a la estimación del cuadro de resultados son cero o distintos de cero, podemos calcular lo que se conoce como **bandas de confianza**, que establecen el límite de significación, en el sentido de que, si los coeficientes mencionados las superan, son estadísticamente distintos de cero, y al revés, en caso contrario. En concreto, la expresión para calcular estas bandas de confianza para un nivel de significación del 5% es:

$$\pm 1,96 \frac{1}{\sqrt{T}} \left[1 + 2 \sum_{s=1}^m \hat{\beta}_s^2 \right]^{0,5},$$

donde se han utilizado de manera relajada resultados obtenidos de Barlett (1946).

Así pues, en nuestro ejemplo tenemos que los primeros nueve coeficientes de autocorrelación simple estimados son:

Por lo tanto, obtendremos las bandas de confianza uniendo los siguientes puntos:

s	Coefficiente de autocorrelación simple estimado
1	-0,148
2	0,245
3	0,27
4	0,249
5	-0,039
6	0,168
7	-0,105
8	0,076
9	0,132

Lecturas complementarias

Para calcular las bandas de confianza obtenidas, ved la obra:

M.S. Barlett (1946). "On Theoretical Specification and Sampling Properties of Autocorrelated Time Series". *Journal of Royal Statistical Society* (núm. 8, pág. 27-41).

Y, para tener más detalles, podéis consultar, por ejemplo, la siguiente obra:

A. Aznar; F.J. Trivez (1993). *Métodos de predicción en economía. Análisis de variables temporales* (vol. II, pág. 160-166). Barcelona: Ariel Economía.

- Para $s = 1 \rightarrow \pm 1,96 \frac{1}{\sqrt{37}} [1 + 2(-0,148^2)]^{0,5} = \pm 0,329204315$.
- Para $s = 2 \rightarrow \pm 1,96 \frac{1}{\sqrt{37}} [1 + 2(-0,148^2 + 0,245^2)]^{0,5} = \pm 0,347620361$.
- Para $s = 3 \rightarrow \pm 1,96 \frac{1}{\sqrt{37}} [1 + 2(-0,148^2 + 0,245^2 + 0,27^2)]^{0,5} = \pm 0,420611443$.

Como podemos comprobar, los valores sucesivos de las bandas de confianza van aumentando a medida que vamos incorporando nuevos coeficientes de autocorrelación simple estimados. Este hecho nos permite concluir, sin necesidad de continuar calculando las bandas de confianza para retardos superiores, que ningún coeficiente de autocorrelación simple las supera, lo cual nos permite afirmar que son estadísticamente cero; es decir, podemos considerar que los residuos asociados a la estimación del cuadro de resultados son un ruido blanco*.

*** Cuando hay ruido blanco, se dice que tenemos un espectro plano.**

Glosario

autocovarianza de orden s

Covarianza entre una variable (como puede ser, por ejemplo, el término de perturbación) en un instante del tiempo t y la misma variable en otro instante del tiempo $t + s$ (o $t - s$).

coeficiente de apuntamiento (*kurtosis*, b_1)

Coefficiente que permite detectar el grado de puntiagudez de una distribución determinada. En el supuesto de que la distribución sea mesocúrtica, el coeficiente de apuntamiento vale tres, si es platicúrtica, su valor es inferior a tres y, si es leptocúrtica, toma un valor superior a tres.

coeficiente de asimetría (*skewness*, b_2)

Coefficiente que nos permite detectar si una distribución determinada es simétrica o no con respecto al valor esperado. En el supuesto de simetría perfecta, el coeficiente de asimetría vale cero.

coeficiente de autocorrelación simple de orden s

Cociente entre la autocovarianza de orden s y la varianza (si la variable es estacionaria). Dicho coeficiente informa acerca de la correlación entre un término de perturbación y este mismo término separado por s periodos de tiempo.

contraste de Breush-Godfrey

Contraste que permite comprobar la existencia de autocorrelación en el término de perturbación de un modelo de regresión. El estadístico de prueba, que podemos obtener a partir de una regresión auxiliar, se distribuye según una χ -cuadrado.

contraste de Breusch-Pagan

Test muy general que permite especificar un gran espectro de esquemas sobre la evolución de la variabilidad del término de perturbación, en el que podemos especificar un conjunto de variables como causantes de la existencia de heteroscedasticidad.

contraste de Durbin-Watson

Contraste útil para comprobar la existencia de autocorrelación según un esquema autorregresivo de orden 1 en el término de perturbación. El estadístico de prueba, DW , está acotado entre cero (autocorrelación perfecta positiva) y cuatro (autocorrelación perfecta negativa).

contraste de Glesjer

Test muy similar al propuesto por Breusch y Pagan, sin embargo, en este caso, sólo podemos especificar una única variable como la causante de la heteroscedasticidad.

contraste de Goldfeld-Quandt

Test que permite contrastar la presencia de heteroscedasticidad, cuando ésta viene causada por una única variable. El estadístico de prueba se distribuye según una F de Snedecor.

contraste de Jarque-Bera

Test que permite contrastar la hipótesis de normalidad del término de perturbación. El estadístico de prueba, que se distribuye según una χ -cuadrado con dos grados de libertad, se basa en los coeficientes de asimetría y de apuntamiento de los residuos del ajuste MCO del modelo.

contraste de White

Test que permite comprobar si todos los regresores de un modelo son los causantes de la existencia de heteroscedasticidad.

contraste h de Durbin

Contraste cuya utilidad es la de comprobar la existencia de autocorrelación en el término de perturbación según un esquema autorregresivo de orden uno, cuando en el modelo figura la variable endógena retardada entre las variables explicativas. El estadístico de prueba se distribuye, asintóticamente, según una ley normal con valor esperado cero y varianza unitaria.

contraste Q de Box-Pierce

Contraste para comprobar la existencia de autocorrelación en el término de perturbación de un modelo de regresión. El estadístico de prueba, que se obtiene a partir de los coeficientes de autocorrelación simple estimados, se distribuye según una χ -cuadrado.

contraste Q de Ljung-Box

Contraste útil para comprobar la existencia de autocorrelación en el término de perturbación de un modelo de regresión. El estadístico de prueba, que es una corrección del estadístico de prueba del test Q de Box-Pierce, se distribuye según una χ -cuadrado.

correlograma

Gráfico en el que se representan los coeficientes de autocorrelación.

esquema de autocorrelación autorregresivo de orden 1, AR(1)

Uno de los esquemas de autocorrelación posibles que puede presentar el término de perturbación de un modelo de regresión. En este caso, el término de perturbación es una combinación lineal de su primer retardo y de una variable aleatoria que recibe el nombre de *ruido blanco*.

función de autocorrelación simple (FAS)

Conjunto formado por todos los coeficientes de autocorrelación simple.

función de autocovarianzas

Conjunto formado por todas las autocovarianzas.

heteroscedasticidad

Ved *homoscedasticidad*.

homoscedasticidad

Hipótesis que postula que la varianza de los términos de perturbación es constante e idéntica para todos los individuos o para todos los periodos de tiempo. Esto se traduce en el hecho de que los elementos de la diagonal principal de la matriz de varianzas y covarianzas del término de perturbación son siempre los mismos. El incumplimiento de esta hipótesis se conoce con el nombre de **heteroscedasticidad**.

matriz Ω

Parte de la matriz de varianzas y covarianzas de los términos de perturbación cuando éstos son no esféricos: $\Omega_N \neq I_N$.

matriz P

Matriz que permite descomponer la matriz de varianzas y covarianzas del término de perturbación, $PP' = \Omega$.

matriz T

Matriz por la que tenemos que multiplicar el modelo de regresión con perturbaciones no esféricas para transformarlo en otro modelo con perturbaciones esféricas. Se calcula a partir de la inversa de la matriz P.

métodos de estimación de Cochrane-Orcutt y de Durbin

Métodos de estimación adecuados cuando el término de perturbación del modelo presenta autocorrelación según un esquema autorregresivo de orden 1 con parámetro desconocido. Se diferencian el uno del otro en la forma previa de estimar este parámetro desconocido.

mínimos cuadrados generalizados (MCG)

Método de estimación que proporciona estimadores eficientes, además de no sesgados y consistentes, cuando el término de perturbación del modelo es no esférico.

mínimos cuadrados ponderados (MCP)

Método de estimación equivalente al de mínimos cuadrados generalizados basado en la transformación de las variables.

modelo de regresión lineal múltiple generalizado (MRLMG)

Modelo de regresión en el que el término de perturbación es no esférico.

no autocorrelación

Hipótesis que postula que los términos de perturbación de los diferentes individuos (o de los diferentes periodos de tiempo) se encuentran incorrelacionados entre sí, lo cual se traduce en el hecho de que los elementos de fuera de la diagonal principal de la matriz de varianzas y covarianzas del término de perturbación son todos cero.

normalidad

Hipótesis en torno al término de perturbación que postula que se distribuye de acuerdo con una ley normal. El cumplimiento de esta hipótesis es fundamental, entre otras cuestiones, para los contrastes de la *t* de Student y de la *F* de Snedecor.

término de perturbación esférico

Término de perturbación cuando cumple las hipótesis de homoscedasticidad y de no autocorrelación. En este caso, la matriz de varianzas y covarianzas se puede descomponer como el producto de un escalar (la varianza) por una matriz identidad.

término de perturbación no esférico

Término de perturbación cuando no cumple la hipótesis de homoscedasticidad y/o de no autocorrelación. En este caso, la matriz de varianzas y covarianzas no es una matriz escalar.

VAR

Varianza o matriz de varianzas y covarianzas, según si hace referencia a una variable o a un vector de variables.

variable aleatoria ruido blanco

Variable que se caracteriza porque tiene un valor esperado cero y una matriz de varianzas y covarianzas escalar.

Bibliografía

Aznar, A.; Trivez, F. J. (1993). *Métodos de predicción en economía. Análisis de series temporales* (vol. II, pág. 160-166). Barcelona: Ariel Economía.

Bartlett, M. S. (1946). "On Theoretical Specification and Sampling Properties of Autocorrelated Time Series". *Journal of Royal Statistical Society, serie B* (núm. 8, pág. 27-41).

Gujarati, D. N. (1993). *Econometría* (cap. 10). Méjico: McGraw-Hill.

Jarque, C. M.; Bera, A. K. (1980). "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals". *Economic Letters* (núm. 6, pág. 255-259).

Johnston, J. (1987). *Métodos de econometría* (cap. 8). Barcelona: Vicens Vives.

Novalés, A. (1993). *Econometría* (2.^a edición, cap. 5, 6 y 7). Madrid: McGraw-Hill.

Uriel, E.; Contreras, D.; Moltó, M. L.; Peiró, A. (1990). *Econometría. El modelo lineal* (cap. 11). Madrid: Editorial AC.

