Ejercicios Resueltos (de examen)

Román Salmerón Gómez

1. Modelo lineal uniecuacional múltiple

1. **Usando los siguientes datos, consumo nacional (C_t) y renta nacional (R_t) en España para el periodo 1995-2005 a precios corrientes (10^9 euros), obtenga las estimaciones por MCO, así como las sumas de cuadrados total, explicada y residual, y el coeficiente de determinación, para el modelo de regresión $C_t = \beta_1 + \beta_2 R_t + u_t$.**

Año	C_t	R_t
1995	**349**	**388**
1996	**368**	**408**
1997	**388**	**433**
1998	**414**	**465**
1999	**444**	**498**
2000	**484**	**538**
2001	**518**	**574**
2002	**550**	**614**
2003	**586**	**656**
2004	**635**	**699**
2005	**686**	**748**

A partir de la información muestral se tiene que:

$$X^t X = \begin{pmatrix} 11 & 6021 \\ 6021 & 3443083 \end{pmatrix}, \qquad X^t C = \begin{pmatrix} 5422 \\ 3104015 \end{pmatrix},$$

por lo que la estimación del modelo por MCO se obtiene a partir de:

$$\widehat{\beta} = \left(X^t X\right)^{-1} X^t C = \begin{pmatrix} 2'1234 & -0'00371 \\ -0'00371 & 0'00000678 \end{pmatrix} \cdot \begin{pmatrix} 5422 \\ 3104015 \end{pmatrix} = \begin{pmatrix} -12'8761 \\ 0'924 \end{pmatrix}.$$

Por tanto, el modelo estimado queda: $\widehat{C}_t = -12'8761 + 0'924 R_t$.
La suma de cuadrados explicada se obtiene a partir de la expresión:

$$SCE = \widehat{\beta}^t \cdot X^t C - n \cdot \overline{C}^2 = 2798415'824 - 11 \cdot \left(\frac{5422}{11}\right)^2 = 125862'7334,$$

mientras que la de los residuos:

$$SCR = \frac{e^t e}{n-k} = \frac{182'1756}{11-2} = 20'2417,$$

donde se ha usado que:

$$e = C - \widehat{C} = \begin{pmatrix} 349 \\ 368 \\ 388 \\ 414 \\ 444 \\ 484 \\ 518 \\ 550 \\ 586 \\ 635 \\ 686 \end{pmatrix} - \begin{pmatrix} 345'6509 \\ 364'1317 \\ 387'2327 \\ 416'8019 \\ 447'2952 \\ 484'2567 \\ 517'5221 \\ 554'4837 \\ 593'2933 \\ 633'0270 \\ 678'3049 \end{pmatrix} = \begin{pmatrix} 3'3491 \\ 3'8683 \\ 0'7673 \\ -2'8019 \\ -3'2952 \\ -0'2567 \\ 0'4779 \\ -4'4837 \\ -7'2933 \\ 1'9730 \\ 7'6951 \end{pmatrix},$$

y por tanto, $e^t e = 182'1756$.

Por otro lado, la suma de cuadrados totales será $SCT = SCE + SCR = 125862'7334 + 182'1756 = 126044'909$.

Finalmente, el coeficiente de determinación es:

$$R^2 = \frac{SCE}{SCT} = \frac{125862'7334}{126044'909} = 0'998554.$$

2. **Para el modelo $Y_t = \beta_1 + \beta_2 v_t + \beta_3 w_t + u_t$ se tienen los siguientes datos:**

$$n = 12, \qquad SCT = 104'9167,$$

$$\left(X^t X\right)^{-1} = \begin{pmatrix} 0'6477 & -0'041 & -0'0639 \\ -0'041 & 0'0071 & -0'0011 \\ -0'0639 & -0'0011 & 0'0152 \end{pmatrix}, \qquad X^t Y = \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix}.$$

Se pide:

a) **Ajustar el modelo por el método de MCO y calcular el coeficiente de determinación.**

b) **Contraste de significación para $\beta_2 + \beta_3 = 1$.**

c) **Intervalo de predicción para $E[Y]$ sabiendo que $v_0 = 2'5$ y $w_0 = -0'3$.**

La estimación del modelo por MCO se obtiene a partir de:

$$\widehat{\beta} = \left(X^t X\right)^{-1} X^t Y = \begin{pmatrix} 0'6477 & -0'041 & -0'0639 \\ -0'041 & 0'0071 & -0'0011 \\ -0'0639 & -0'0011 & 0'0152 \end{pmatrix} \cdot \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix} = \begin{pmatrix} 1'6545 \\ 0'7391 \\ 0'2258 \end{pmatrix}.$$

Por tanto, el modelo estimado queda: $\widehat{Y}_t = 1'6545 + 0'7391 v_t + 0'2258 w_t$.

Para calcular el coeficiente de determinación tendremos en cuenta que:

$$SCE = \widehat{\beta}^t \cdot X^t Y - n \cdot \overline{Y}^2 = 768'3488 - 690'083 = 78'2654,$$

donde se ha usado que:

$$\widehat{\beta}^t \cdot X^t Y = (1'6545\ 0'7391\ 0'2258) \cdot \begin{pmatrix} 91 \\ 699 \\ 448 \end{pmatrix} = 768'3488,$$

y

$$\sum Y_t = 91 \to \overline{Y} = \frac{91}{12} = 7'583 \to \overline{Y}^2 = 57'5069.$$

Además, como $SCT = 104'9167$, el coeficiente de determinación será:

$$R^2 = \frac{SCE}{SCT} = \frac{78'26547}{104'9167} = 0'7459.$$

Es decir, el ajuste realizado explica aproximadamente un 74'59 % de la variabilidad de Y.

Para el contraste de significación de la restricción $H_0 : \beta_2 + \beta_3 = 1$, tendremos en cuenta que se rechaza la hipótesis nula si:

$$F_{exp} = \frac{\left(R\widehat{\beta} - r\right)^t \cdot \left[R\left(X^tX\right)^{-1}R^t\right]^{-1} \cdot \left(R\widehat{\beta} - r\right)}{q \cdot \widehat{\sigma}^2} > F_{q,n-k}(1-\alpha).$$

De la restricción $\beta_2 + \beta_3 = 1$ se obtiene que $R = (0\ 1\ 1)$, $r = 1$ y $q = 1$, por lo que:

$$R\widehat{\beta} - r = (0\ 1\ 1) \cdot \begin{pmatrix} 1'6545 \\ 0'7391 \\ 0'2258 \end{pmatrix} - 1 = -0'0351,$$

$$\begin{aligned} R \cdot \left(X^tX\right)^{-1} \cdot R^t &= (0\ 1\ 1) \cdot \begin{pmatrix} 0'6477 & -0'041 & -0'0639 \\ -0'041 & 0'0071 & -0'0011 \\ -0'0639 & -0'0011 & 0'0152 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \\ &= (-0'1049\ 0'006\ 0'0141) \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 0'0201. \end{aligned}$$

Y como:

$$\widehat{\sigma}^2 = \frac{SCR}{n-k} = \frac{SCT - SCE}{n-k} = \frac{104'9167 - 78'2654}{12-3} = \frac{26'6512}{9} = 2'962,$$

se tiene que:

$$F_{exp} = \frac{0'0351^2}{2'962 \cdot 0'0201} = \frac{0'0012}{0'0595} = 0'0207 \not> 5'117 = F_{1,9}(0'95),$$

por lo que no se rechaza la hipótesis nula.

Finalmente, el intervalo de predicción para $E[Y]$ es:

$$x_0^t\widehat{\beta} \pm t_{n-k}\left(1 - \frac{\alpha}{2}\right) \cdot \widehat{\sigma} \cdot \sqrt{x_0^t \cdot \left(X^tX\right)^{-1} \cdot x_0}.$$

Como:

$$x_0^t\widehat{\beta} = 1'6445 + 0'7391 \cdot 2'5 + 0'2258 \cdot (-0'3) = 3'43451,$$

$$\begin{aligned} x_0^t \cdot \left(X^tX\right)^{-1} \cdot x_0 &= (1\ 2'5\ \ -0'3) \cdot \begin{pmatrix} 0'6477 & -0'041 & -0'0639 \\ -0'041 & 0'0071 & -0'0011 \\ -0'0639 & -0'0011 & 0'0152 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2'5 \\ -0'3 \end{pmatrix} \\ &= (0'56437\ \ -0'02292\ \ -0'07121) \cdot \begin{pmatrix} 1 \\ 2'5 \\ -0'3 \end{pmatrix} = 0'528433, \end{aligned}$$

el intervalo de confianza al 95 % es:

$$3'43451 \pm 2'2622 \cdot 1'721 \cdot 0'727 = (0'60451,\ 6'26451).$$

3. **En un estudio de los determinantes de la inversión se usaron 20 datos anuales, correspondientes a las siguientes variables: inversión anual en billones de pesetas (Y), tipo de interés en porcentaje (X_1) y variación anual de PIB en billones de pesetas (X_2). Se dispone de la siguiente información:**

$$\begin{array}{lll} \sum X_{1t} = 100 & \sum X_{2t} = 24 & \sum Y_t = 5 \\ \sum X_{1t}Y_t = -255 & \sum X_{2t}Y_t = 146 & \sum X_{1t}X_{2t} = 100 \\ \sum X_{1t}^2 = 680 & \sum X_{2t}^2 = 48'8 & \sum \left(Y_t - \overline{Y}\right)^2 = 1200 \end{array}$$

Se pide:

a) **Obtenga las estimaciones por MCO del modelo $Y_t = \alpha + \beta X_{1t} + \delta X_{2t} + u_2$.**

b) **Contraste la significación global del modelo a partir del porcentaje de evolución temporal de la inversión que puede explicarse por la influencia lineal del tipo de interés y la variación anual del PIB.**

c) **Contraste la hipótesis nula:**

$$\left.\begin{array}{l} \beta_1 = 1 \\ \beta_2 = 2 \end{array}\right\}.$$

A partir de la información proporcionada en el enunciado se tiene que:

$$X^tX = \begin{pmatrix} 20 & 100 & 24 \\ 100 & 680 & 100 \\ 24 & 100 & 48'8 \end{pmatrix}, \qquad X^tY = \begin{pmatrix} 5 \\ -255 \\ 146 \end{pmatrix}.$$

Luego las estimaciones por MCO de los coeficientes del modelo serán:

$$\widehat{\beta} = \left(X^tX\right)^{-1} X^tY = \begin{pmatrix} 0'3623 & -0'0388 & -0'0988 \\ -0'0388 & 0'0063 & 0'0063 \\ -0'0988 & 0'0063 & 0'0562 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ -255 \\ 146 \end{pmatrix} = \begin{pmatrix} -2'725 \\ -0'875 \\ 6'125 \end{pmatrix}.$$

Por tanto, el modelo estimado queda: $\widehat{Y}_t = -2'725 - 0'875X_{1t} + 6'125X_{2t}$.

Para el modelo sea significativo a partir del coeficiente de determinación se ha de verificar que:

$$R^2 > \frac{\frac{k-1}{n-k} \cdot F_{k-1,n-k}(1-\alpha)}{1 + \frac{k-1}{n-k} \cdot F_{k-1,n-k}(1-\alpha)} = R^2_{sig}.$$

Puesto que:

$$\widehat{\beta}^t \cdot X^tY = (-2'725 \;\; -0'875 \;\, 6'125) \cdot \begin{pmatrix} 5 \\ -255 \\ 146 \end{pmatrix} = 1103'7,$$

y

$$\sum Y_t = 5 \rightarrow \overline{Y} = \frac{5}{20} = 0'25 \rightarrow n \cdot \overline{Y} = 1'25,$$

es claro que $SCE = \widehat{\beta}^t \cdot X^tY - n \cdot \overline{Y} = 1103'7 - 1'25 = 1102'45$. Y como el enunciado nos proporciona que $SCT = \sum \left(Y_t - \overline{Y}\right)^2 = 1200$, se obtiene que:

$$R^2 = \frac{1102'45}{1200} = 0'9187.$$

Por otro lado:

$$R^2_{sig} = \frac{\frac{2}{17} \cdot 3'59}{1 + \frac{2}{17} \cdot 3'59} = \frac{0'4224}{1'4224} = 0'2969.$$

Por tanto, como $R^2 > R^2_{sig}$, podemos afirmar que el modelo es significativo.

Finalmente, para contrastar la hipótesis

$$H_0 : \left\{ \begin{array}{l} \beta_1 = 1 \\ \beta_2 = 2 \end{array} \right. ,$$

se tiene que

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad r = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad q = 2.$$

En tal caso:

$$R\widehat{\beta} - r = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -2'725 \\ -0'875 \\ 6'125 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -0'875 \\ 6'125 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1'875 \\ 4'125 \end{pmatrix},$$

$$\begin{aligned} R \cdot \left(X^t X\right)^{-1} \cdot R^t &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0'3623 & -0'0388 & -0'0988 \\ -0'0388 & 0'0063 & 0'0063 \\ -0'0988 & 0'0063 & 0'0562 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -0'0063 & 0'0063 \\ 0'0063 & 0'0562 \end{pmatrix}. \end{aligned}$$

Por otro lado:

$$SCR = SCT - SCE = 1200 - 1102'45 = 97'55 \to \widehat{\sigma}^2 = \frac{97'55}{17} = 5'7382,$$

y entonces:

$$F_{exp} = \frac{(-1'875 \; 4'125) \cdot \begin{pmatrix} 178'7702 & -20'0401 \\ -20'0401 & 20'0401 \end{pmatrix} \cdot \begin{pmatrix} -1'875 \\ 4'125 \end{pmatrix}}{2 \cdot 5'7382} = \frac{1279'5}{11'4764} = 111'4897.$$

Como $F_{exp} > 3'59 = F_{2,17}(0'95)$, se rechaza la hipótesis nula.

4. **Se desea estudiar la influencia que sobre la demanda de carne de vacuno ha tenido el precio de la carne de cerdo (X_1) y de la ternera (X_2). Para ello se han tomado datos anuales desde 1979 a 2001 (ambos inclusive), obteniéndose los siguientes resultados:**

$$\widehat{Y}_t = 2'1 + 0'7X_{1t} - 1'5X_{2t}, \quad R^2 = 0'9, \quad SCE = 126.$$

¿Se podría afirmar, para un nivel de confianza del 95 %, que los precios no influyen sobre la demanda de ternera?

Para saber si los precios de la carne de cerdo y de ternera influyen en la demanda de la carne estudiaremos la significación conjunta del modelo. Puesto que:

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{0'9/2}{0'1/20} = \frac{0'45}{0'005} = 90 > 3'49 = F_{2,20}(0'95) = F_{k-1,n-k}(1-\alpha),$$

concluimos que se rechaza la hipótesis nula de que todos los coeficientes de las variables explicativas son nulos de forma simultánea, por lo que los precios de la carne influyen sobre la demanda.

5. **Para estimar el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ se ha obtenido una muestra de la cual ha resultado:**

$$X^t X = \begin{pmatrix} 14 & 7 & 14 \\ 7 & 4'5 & 7 \\ 14 & 7 & 15 \end{pmatrix}, \qquad X^t Y = \begin{pmatrix} 10 \\ 6 \\ 12 \end{pmatrix}, \qquad Y^t Y = 14.$$

Se pide:

a) **Estimar los coeficientes del modelo por MCO.**

b) **Estudiar la significación del modelo.**

c) **Contrastar la hipótesis $\beta_2 + 1 = \beta_3$.**

d) **Calcular el intervalo de predicción al 95 % para Y sabiendo que $X_2 = 5$ y $X_3 = 7$.**

Las estimaciones por MCO de los coeficientes del modelo serán:

$$\widehat{\beta} = \left(X^t X\right)^{-1} X^t Y = \begin{pmatrix} 1'3214 & -0'5 & -1 \\ -0'5 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 10 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} -1'7857 \\ 1 \\ 2 \end{pmatrix}.$$

Por tanto, el modelo estimado queda: $\widehat{Y}_t = -1'7857 + X_{2t} + 2X_{3t}$.

Para estudiar la significación global del modelo recurriremos al contraste ANOVA, de manera que el modelo será significativo si

$$F_{exp} = \frac{SCE/k-1}{SCR/n-k} > F_{k-1,n-k}(1-\alpha).$$

Teniendo en cuenta que:

$$\widehat{\beta}^t \cdot X^t Y = (-1'7857\ 1\ 2) \cdot \begin{pmatrix} 10 \\ 6 \\ 12 \end{pmatrix} = 12'1429,$$

y[1]

$$\sum Y_t = 10 \to \overline{Y} = \frac{10}{14} = 0'7143,$$

se tiene que $SCE = \widehat{\beta}^t \cdot X^t Y - n \cdot \overline{Y}^2 = 12'1429 - 14 \cdot 0'7143^2 = 4'9998$. Además, $SCT = Y^t Y - n \cdot \overline{Y}^2 = 14 - 14 \cdot 0'7143^2 = 6'8569$. Por tanto, $SCR = SCT - SCE = 1'8571$.

Con todo ello:

$$F_{exp} = \frac{4'9998/2}{1'8571/11} = 14'8074 > 3'98 = F_{2,1}(0'95).$$

Esto es, el modelo es significativo.

Para contrastar la hipótesis $H_0 : \beta_2 + 1 = \beta_3$, tendremos que tener en cuenta que $R = (0\ 1\ -1)$, $r = -1$ y $q = 1$. En tal caso:

$$R\widehat{\beta} - r = (0\ 1\ -1) \cdot \begin{pmatrix} -1'7857 \\ 1 \\ 2 \end{pmatrix} + 1 = 0,$$

lo cual conduce a que $F_{exp} = 0 \not> 4'84 = F_{1,11}(0'95) = F_{q,n-k}(1-\alpha)$. Por tanto, no se rechaza la hipótesis nula.

[1] El primer elemento de la matriz $X^t X$ indica que $n = 14$.

Finalmente, el intervalo de predicción para Y es:

$$x_0^t\widehat{\beta} \pm t_{n-k}\left(1-\frac{\alpha}{2}\right)\cdot\widehat{\sigma}\cdot\sqrt{1+x_0^t\cdot\left(X^tX\right)^{-1}\cdot x_0}.$$

Como:

$$x_0^t\widehat{\beta} = (1\ 5\ 7)\cdot\begin{pmatrix} -1'7857 \\ 1 \\ 2 \end{pmatrix} = 17'2143,$$

$$\widehat{\sigma}^2 = \frac{SCR}{n-k} = \frac{1'8571}{11} = 0'1688 \rightarrow \widehat{\sigma} = 0'4109,$$

$$\begin{aligned} x_0^t\cdot\left(X^tX\right)^{-1}\cdot x_0 &= (1\ 5\ 7)\cdot\begin{pmatrix} 1'3214 & -0'5 & -1 \\ -0'5 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\cdot\begin{pmatrix} 1 \\ 5 \\ 7 \end{pmatrix} \\ &= (-8'1786\ 4'5\ 6)\cdot\begin{pmatrix} 1 \\ 5 \\ 7 \end{pmatrix} = 56'3214, \end{aligned}$$

el intervalo de confianza al 95 % es:

$$17'2143 \pm 2'201\cdot 0'4109\cdot 7'5711 = (10'3671,\ 24'0615).$$

6. **Al objeto de determinar si existen o no diferencias en las calificaciones obtenidas por hombres y mujeres en una determinada asignatura, a partir de 20 observaciones se estimó el modelo:**

$$nota_t = \beta_0 + \beta_1 notamediaBUP_t + \beta_2 genero_t + u_t,$$

donde la variable *genero* **toma el valor 1 si se trata de una mujer y 0 para un varón. Los resultados de la estimación fueron los siguientes:**

$$\widehat{nota}_t = \underset{(\mathbf{4'5})}{\mathbf{25}} + \underset{(\mathbf{7'1})}{\mathbf{0'75}}\ notamediaBUP_t + \underset{(\mathbf{2'3})}{\mathbf{20'5}}\ genero_t \qquad R^2 = \mathbf{0'72}$$

¿Puede decirse que los resultados de unos y otros son distintos?

Teniendo en cuenta que la nota esperada para un varón y una mujer son, respectivamente:

$$E\left[nota_t/genero_t = 0\right] = \beta_0 + \beta_1 notamediaBUP_t,$$

$$E\left[nota_t/genero_t = 1\right] = \beta_0 + \beta_1 notamediaBUP_t + \beta_2,$$

se tiene que, para una misma nota media en BUP, la diferencia esperada entre la nota de una mujer y un hombre viene determinada por

$$E\left[nota_t/genero_t = 1\right] - E\left[nota_t/genero_t = 0\right] = \beta_2.$$

Como el contraste de significación individual para dicho parámetro es significativo:

$$t_{exp} = \left|\frac{20'5}{2'3}\right| = 8'913 > 2'0003 = t_{60}(0'975),$$

se tiene que dicho parámetro es distinto de cero. Por tanto, puede afirmarse que los resultados de unos y otros son distintos.

Además, como la estimación de dicho parámetro es positiva, la nota esperada para una mujer es mayor que la de un hombre (siempre y cuando tengan la misma nota media en BUP).

7. **Con información muestral relativa a 14 observaciones, se pretende estimar el modelo de regresión:**

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t,$$

a partir de:

$$X^t X = \begin{pmatrix} 14 & & & \\ 85 & 631 & & \\ 532 & 3126 & 20666 & \\ 2094 & 13132 & 78683 & 317950 \end{pmatrix}, \qquad X^t Y = \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix}.$$

Se pide:

a) **Calcular las estimaciones de los parámetros del modelo por MCO.**

b) **Estimar** $Var\left(\widehat{\beta}\right)$**.**

c) **¿Influyen las variaciones de** X_{2t} **en la variable dependiente?**

d) **Calcular el coeficiente de determinación corregido.**

e) **Calcular un intervalo de confianza del 95 % para la varianza del término de perturbación.**

f) **Contrastar la significación global del modelo al 95 %.**

Las estimaciones de los parámetros del modelo son:

$$\begin{aligned} \widehat{\beta} &= \left(X^t X\right)^{-1} X^t Y \\ &= \begin{pmatrix} 20'164 & 0'015065 & -0'23145 & -0'7617 \\ 0'015065 & 0'013204 & 0'001194 & -0'00094 \\ -0'23145 & 0'001194 & 0'003635 & 0'000575 \\ -0'7617 & -0'00094 & 0'000575 & 0'000401 \end{pmatrix} \cdot \begin{pmatrix} 248 \\ 1622 \\ 9202 \\ 37592 \end{pmatrix} \\ &= \begin{pmatrix} 32'891 \\ 0'80371 \\ -0'3982 \\ -0'03713 \end{pmatrix}. \end{aligned}$$

Por otro lado, teniendo en cuenta (por construcción del vector $X^t Y$) que:

$$\sum Y_t = 248 \rightarrow \overline{Y} = \frac{248}{14} = 17'714,$$

y que:

$$\widehat{\beta}^t \cdot X^t Y = 4552'552,$$

se tiene que $SCE = \widehat{\beta}^t \cdot X^t Y - n \cdot \overline{Y} = 4552'552 - 14 \cdot 17'714^2 = 159'551.$

Entonces, puesto que en el enunciado se nos indica que $SCT = 226'86$, es inmediato que $SCR = SCT - SCE = 67'308$ y, por tanto:

$$\widehat{\sigma}^2 = \frac{SCR}{n-k} = \frac{67'308}{14-4} = 6'7308.$$

Luego, la estimación de $Var\left(\widehat{\beta}\right)$ corresponde a:

$$\widehat{Var\left(\widehat{\beta}\right)} = \widehat{\sigma}^2 \cdot \left(X^t X\right)^{-1} = \begin{pmatrix} 1'3575 & 0'00101 & -0'0155 & -0'00512 \\ 0'00101 & 0'000888 & 0'00008 & -0'000063 \\ -0'0155 & 0'00008 & 0'00024 & 0'000038 \\ -0'00512 & -0'000063 & 0'000038 & 0'000027 \end{pmatrix}.$$

A partir de ambas estimaciones podremos determinar si las variaciones de X_{2t} influyen en Y_t:

$$t_{exp} = \left| \frac{-0'3982}{\sqrt{0'00024}} \right| = 25'704 > 2'228 = t_{10}(0'975).$$

Evidentemente se rechaza que $\beta_2 = 0$, por lo que X_{2t} influye en la variable dependiente.

Para calcular el coeficiente de determinación corregido tendremos en cuenta la siguiente expresión:

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-k}.$$

Puesto que $R^2 = \frac{159'551}{226'86} = 0'7033$ es claro que:

$$\overline{R}^2 = 1 - (1 - 0'7033) \cdot \frac{13}{10} = 0'6143.$$

Podemos observar que al eliminar la influencia de las variables explicativas el coeficiente de determinación ha disminuido alrededor del 9 %.

El intervalo de confianza para σ^2 es:

$$\left(\frac{(n-k) \cdot \widehat{\sigma}^2}{\chi^2_{n-k,1-\frac{\alpha}{2}}}, \frac{(n-k) \cdot \widehat{\sigma}^2}{\chi^2_{n-k,\frac{\alpha}{2}}} \right) = \left(\frac{10 \cdot 6'7308}{20'483}, \frac{10 \cdot 6'7308}{3'247} \right) = (3'286, 20'73).$$

Finalmente, para contrastar la significación conjunta del modelo construiremos la tabla ANOVA:

Fuente Variación	Suma de Cuadrados	Grados de Libertad	Medias
Explicada	159'551	3	53'1836
Residual	67'308	10	6'7308
Total	226'86		7'9015

Como $F_{exp} = 7'9015 > 3'71 = F_{3,10}(0'95)$, se rechaza la hipótesis nula de que todos los coeficientes son nulos de forma simultánea, por tanto, el modelo es significativo en su conjunto.

2. Multicolinealidad

1. **Dadas las siguientes matrices:**

$$X = \begin{pmatrix} 1 & 4 & 13 \\ 1 & 1 & 4 \\ 1 & 5 & 16 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & -5 & 3 \\ 1 & 3 & 44 \\ 1 & 0 & 6 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & -2 & 3'0001 \\ 1 & -3 & 4 \\ 1 & 5 & -3'9999 \end{pmatrix},$$

analice en cada caso la posible existencia de multicolinealidad.

En el caso de la primera matriz, puesto que la tercera columna se obtiene sumando a la primera el triple de la segunda, estamos ante un claro ejemplo de multicolinealidad perfecta (su determinante es cero), mientras que las columnas de la segunda matriz son linealmente independientes (determinante igual a -181), por lo que no hay multicolinealidad en este caso. Finalmente, en la tercera matriz se muestra un caso de multicolinealidad aproximada, ya que la primera fila menos la tercera es aproximadamente igual a la segunda (determinante igual a 0'0007).

2. **Si en el modelo $Y_t = \alpha + \beta X_t + \delta Z_t + u_t$ se cumple que $\dfrac{X_t}{Z_t} = \lambda$ constante. ¿Qué parámetros se pueden estimar?**

A partir de $\dfrac{X_t}{Z_t} = \lambda$ se tiene que $X_t = \lambda Z_t$, de forma que sin más que sustituir dicha expresión en la ecuación del modelo:

$$Y_t = \alpha + \beta\lambda Z_t + \delta Z_t + u_t = \alpha + (\beta\lambda + \delta)\, Z_t + u_t,$$

se pueden estimar α y $\beta\lambda + \delta$. Luego, a no ser que se tenga información a priori, no se podrían estimar los parámetros originales.

Si se opta por la opción $Z_t = \dfrac{X_t}{\lambda}$, se podrá estimar α y $\frac{\delta}{\lambda}$.

3. **Dado el modelo $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$, utilizando una muestra de 20 datos, se procedió a su estimación, obteniéndose:**

$$\begin{array}{ccccccccc} \widehat{Y}_t & = & \mathbf{8'34} & + & \mathbf{0'7}\ X_{2t} & \textbf{-} & \mathbf{0'4}\ X_{3t} & + & \mathbf{0'1}\ X_{4t} \qquad R^2 = \mathbf{0'96} \\ & & & & \mathbf{(0'56)} & & \mathbf{(0'7)} & & \mathbf{(0'5)} \end{array}$$

Se pide:

a) Analice el posible problema de multicolinealidad.

b) Si hay algún problema, indique la forma más adecuada de solucionarlo.

Atendiendo a los contrastes de significación individual, ningún coeficiente es significativo, ya que:

$$t_{exp} = \left| \frac{0'7}{0'56} \right| = 1'25 \not> 2'12 = t_{16}(0'975),$$

$$t_{exp} = \left| \frac{-0'4}{0'7} \right| = 0'5714 \not> 2'12 = t_{16}(0'975),$$

$$t_{exp} = \left| \frac{0'1}{0'5} \right| = 0'2 \not> 2'12 = t_{16}(0'975).$$

Además, el coeficiente de determinación es bastante alto y el modelo es conjuntamente significativo:

$$F_{exp} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} = \frac{0'93/3}{0'04/16} = \frac{0'32}{0'0025} = 128 > 3'24 = F_{3,16}(0'95) = F_{k-1,n-k}(1-\alpha).$$

Todo esto nos hace pensar en la posible existencia de multicolinealidad en el modelo.

La principal solución para eliminar la relación lineal entre las variables independientes consiste en eliminar del modelo la variable que causa la multicolinealidad.

4. **En el modelo de regresión lineal múltiple $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$ se verifica que $X_{2t} = 3X_{4t}$. Indique qué parámetros son estimables:**

a) cuando no se dispone de información a priori sobre los coeficientes, y

b) cuando se sabe que $\beta_4 = 2$.

Sustituyendo $X_{2t} = 3X_{4t}$ en la ecuación del modelo se obtiene que:

$$\begin{aligned} Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \\ &= \beta_1 + 3\beta_2 X_{4t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \\ &= \beta_1 + \beta_3 X_{3t} + (3\beta_2 + \beta_4)\, X_{4t} + u_t. \end{aligned}$$

Por tanto, son estimables los parámetros β_1, β_3 y la combinación lineal $3\beta_2 + \beta_4$. Por tanto, a no ser que haya información a priori no será posible obtener las estimaciones de β_2 y β_4.

Si se sabe que $\beta_4 = 2$, entonces la ecuación del modelo quedaría:

$$Y_t = \beta_1 + \beta_3 X_{3t} + (3\beta_2 + 2)\, X_{4t} + u_t,$$

por lo que se podrían estimar los parámetros β_1, β_2 y β_3.

3. Heteroscedasticidad

1. **Dado un modelo $Y_t = \beta_0 + \beta_1 X_t + u_t$, $t = 1, \ldots, n$, donde $Var(u_t) = \sigma^2 X_t^2$, obtener las expresiones de las variables transformadas de tal forma que las estimaciones por MCG de β puedan calcularse estimando por MCO.**

   Puesto que $Var(u_t) = \sigma^2 X_t^2$ es claro que la matriz de transformación en este caso es:

$$P = \begin{pmatrix} \frac{1}{X_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{X_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{X_n} \end{pmatrix},$$

   por lo que el nuevo modelo transformado, $Y_t^* = \beta_0 X_{1t}^* + \beta_1 X_{2t}^* + u_t^*$, vendrá determinado por

$$Y_t^* = \frac{Y_t}{X_t}, \quad X_{1t}^* = \frac{1}{X_t}, \quad X_{2t}^* = \frac{X_t}{X_t} = 1, \quad u_t^* = \frac{u_t}{X_t}, \quad \forall t.$$

   Además, como la perturbación aleatoria de este modelo transformado verifica, para cualquier valor de t, que:

$$\begin{aligned} E\left[u_t^*\right] &= E\left[\frac{u_t}{X_t}\right] = \frac{1}{X_t} \cdot E\left[u_t\right] = 0, \\ Var\left(u_t^*\right) &= Var\left(\frac{u_t}{X_t}\right) = \frac{1}{X_t^2} \cdot Var\left(u_t\right) = \frac{\sigma^2 \cdot X_t^2}{X_t^2} = \sigma^2, \\ Cov\left(u_t^*, u_{t-k}^*\right) &= E\left[u_t^* \cdot u_{t-k}^*\right] = E\left[\frac{u_t}{X_t} \cdot \frac{u_{t-k}}{X_{t-k}}\right] = \frac{E\left[u_t \cdot u_{t-k}\right]}{X_t \cdot X_{t-k}} = 0, \end{aligned}$$

   al estimarlo por MCO se obtendrán las estimaciones por MCG, que son lineales, insesgadas y óptimas. Adviértase que se supone que u verifica que tiene media cero y está incorrelada.

2. **En el modelo $Y_t = \beta_0 + \beta_1 X_t + u_t$, $t = 1, ..., n$, cuyas perturbaciones no están autocorrelacionadas, pero son tales que $Var(u_t) = E(u_t^2) = \dfrac{\sigma^2}{X_t^2}$, ¿cómo obtendría el estimador más adecuado de β?.**

   Puesto que se verifica que $Var(u_t) = E(u_t^2) = \dfrac{\sigma^2}{X_t^2}$, para obtener un estimador lineal, insesgado y óptimo de β transformaría el modelo original mediante la siguiente matriz

$$P = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_n \end{pmatrix},$$

   de manera que el nuevo modelo transformado, $Y_t^* = \beta_0 X_{1t}^* + \beta_1 X_{2t}^* + u_t^*$, determinado por

$$Y_t^* = Y_t \cdot X_t, \quad X_{1t}^* = 1 \cdot X_t = X_t, \quad X_{2t}^* = X_t \cdot X_t = X_t^2, \quad u_t^* = u_t \cdot X_t, \quad \forall t,$$

   es un modelo con perturbaciones esféricas, ya que, para cualquier valor de t, verifica:

$$\begin{aligned} E\left[u_t^*\right] &= E\left[u_t \cdot X_t\right] = X_t \cdot E\left[u_t\right] = 0, \\ Var\left(u_t^*\right) &= Var\left(u_t \cdot X_t\right) = X_t^2 \cdot Var\left(u_t\right) = X_t^2 \cdot \frac{\sigma^2}{X_t^2} = \sigma^2, \\ Cov\left(u_t^*, u_{t-k}^*\right) &= E\left[u_t^* \cdot u_{t-k}^*\right] = E\left[u_t X_t \cdot u_{t-k} X_{t-k}\right] = X_t \cdot X_{t-k} \cdot E\left[u_t \cdot u_{t-k}\right] = 0. \end{aligned}$$

   En tal caso, al estimarlo por MCO obtendremos estimaciones lineales, insesgadas y óptimas para β.

3. **Dado el modelo:** $Y_t = \beta_0 + \beta_1 X_t$ **con los siguientes datos:**

Y	**2**	**3**	**7**	**6**	**15**	**8**	**22**
X	**-3**	**-2**	**-1**	**0**	**1**	**2**	**3**
e	**1'37**	**-0'42**	**0'79**	**-3**	**3'21**	**-6'58**	**4'63**

Utilizar el contraste de Goldfeld-Quandt y el contraste de Glesjer para la detección de heteroscedasticidad.

Para detectar la heteroscedasticidad a partir del test de Goldfeld-Quant hay que ordenar las observaciones de menor a mayor respecto de la variable que se considera provoca la heteroscedasticidad. En este caso, puesto que sólo hay una variable independiente, X_t, hay que ordenar en función de sus valores. Como se puede observar, los datos ya están ordenados de forma adecuada.

A continuación hay que eliminar m observaciones centrales (normalmente un tercio de la muestra). Como $m = \frac{7}{3} = 2'3333$, en este caso deberíamos eliminar 2 observaciones centrales. Pero esta elección supondría que uno de los dos subgrupos estuviese formado por dos puntos, lo que conduce a un ajuste perfecto y, entonces, su suma de cuadrados de los residuos sería cero. Para evitar este hecho vamos a eliminar una única observación central, por lo que nos quedarían los subgrupos:

$$y_1 = \begin{pmatrix} 2 \\ 3 \\ 7 \end{pmatrix}, \quad X_1 = \begin{pmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \end{pmatrix} \quad \text{e} \quad y_2 = \begin{pmatrix} 15 \\ 8 \\ 22 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}.$$

De estos dos subgrupos lo único que nos interesa es su suma de cuadrados de los residuos. Así, para el primer subgrupo se tiene que:

$$SCR_1 = y_1^t y_1 - \widehat{\beta}_1^t X_1^t y_1 = 1'5,$$

donde se ha usado que:

$$\widehat{\beta}_1 = \left(X_1^t X_1\right)^{-1} X_1^t y_1 = \begin{pmatrix} 3 & -6 \\ -6 & 14 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 12 \\ -19 \end{pmatrix} = \begin{pmatrix} 2'3334 & 1 \\ 1 & 0'5 \end{pmatrix} \cdot \begin{pmatrix} 12 \\ -19 \end{pmatrix} = \begin{pmatrix} 9 \\ 2'5 \end{pmatrix}.$$

Mientras que para el segundo:

$$SCR_2 = y_2^t y_2 - \widehat{\beta}_2^t X_2^t y_2 = 73'5,$$

donde se ha usado que:

$$\widehat{\beta}_2 = \left(X_2^t X_2\right)^{-1} X_2^t y_2 = \begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 45 \\ 97 \end{pmatrix} = \begin{pmatrix} 2'3334 & -1 \\ -1 & 0'5 \end{pmatrix} \cdot \begin{pmatrix} 45 \\ 97 \end{pmatrix} = \begin{pmatrix} 8 \\ 3'5 \end{pmatrix}.$$

En tal caso, puesto que

$$F_{exp} = \frac{SCR_2}{SCR_1} = \frac{73'5}{1'5} = 49 > 9'28 = F_{3,3}(0'95) = F_{\frac{n-m}{2},\frac{n-m}{2}}(0'95),$$

se rechaza la hipótesis nula de homocedasticidad. Es decir, la perturbación aleatoria del modelo considerado es heterocedástica.

Para aplicar el test de Glesjer, hay que plantear la regresión auxiliar

$$|e_t| = \alpha + \beta X_t^h + v_t,$$

donde los valores más comunes para h son $\pm 2, \pm 1, \pm 1/2$. En este caso, debido a la naturaleza de las observaciones, sólo es posible estudiar los casos en los que $h = 1, 2$.

Cuadro 1: Datos regresión auxiliar

$\lvert e_t \rvert$	X_t^2	X_t
1’37	9	-3
0’42	4	-2
0’79	1	-1
3	0	0
3’21	1	1
6’58	4	2
4’63	9	3

Atendiendo a la información de la tabla 1 se tiene que para $h = 1$:

$$\begin{aligned}
\widehat{\beta} &= \left(X^t X\right)^{-1} X^t y = \begin{pmatrix} 2'8571 \\ 0'8757 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{y^t y - \widehat{\beta}^t X^t y}{n-k} = \frac{9'0994}{7-2} = 1'6199, \\
\widehat{Var\left(\widehat{\beta}\right)} &= \widehat{\sigma}^2 \cdot \left(X^t X\right)^{-1} = \begin{pmatrix} 0'2314 & 0 \\ 0 & 0'0579 \end{pmatrix},
\end{aligned}$$

donde

$$X = \begin{pmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \qquad y = \begin{pmatrix} 1'37 \\ 0'42 \\ 0'79 \\ 3 \\ 3'21 \\ 6'58 \\ 4'63 \end{pmatrix}.$$

En tal caso, puesto que

$$t_{exp} = \frac{0'8757}{\sqrt{0'0579}} = 3'6393 > 2'571 = t_5(0'975) = t_{n-k}(0'975),$$

se tiene que se rechaza la hipótesis nula de que la pendiente sea cero.

Para $h = 2$:

$$\begin{aligned}
\widehat{\beta} &= \left(X^t X\right)^{-1} X^t y = \begin{pmatrix} 2'5714 \\ 0'0714 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{y^t y - \widehat{\beta}^t X^t y}{n-k} = \frac{29'1434}{5} = 5'8287, \\
\widehat{Var\left(\widehat{\beta}\right)} &= \widehat{\sigma}^2 \cdot \left(X^t X\right)^{-1} = \begin{pmatrix} 1'9429 & -0'2776 \\ -0'2776 & 0'0694 \end{pmatrix},
\end{aligned}$$

donde

$$X = \begin{pmatrix} 1 & 9 \\ 1 & 4 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 4 \\ 1 & 9 \end{pmatrix}, \qquad y = \begin{pmatrix} 1'37 \\ 0'42 \\ 0'79 \\ 3 \\ 3'21 \\ 6'58 \\ 4'63 \end{pmatrix}.$$

En tal caso, puesto que

$$t_{exp} = \frac{0'0714}{\sqrt{0'0694}} = 0'271 \not> 2'571 = t_5(0'975) = t_{n-k}(0'975),$$

se tiene que no se rechaza la hipótesis nula de que la pendiente sea cero.

Por tanto, atendiendo a los resultados obtenidos, podemos decir que hay heteroscedasticidad en la perturbación aleatoria del modelo. Además, podemos suponer[2] que $E\left[u_t^2\right] = \sigma^2 X_t$. Luego, para eliminar la heteroscedasticidad habría que transformar el modelo mediante la matriz:

$$P = \begin{pmatrix} \frac{1}{\sqrt{X_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{X_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{X_7}} \end{pmatrix}.$$

4. **En un estudio econométrico se ha relacionado linealmente el desempleo con las demandas de trabajo y el IPC en 50 provincias españolas. Para analizar la posible presencia de heteroscedasticidad provocada por el IPC, se ha procedido a ordenar las observaciones de menor a mayor respecto de dicha variable, se han eliminado 14 datos centrales y a partir de los dos subgrupos restantes se han obtenido los siguientes resultados:** $SCR_1 = 65432$ **y** $SCR_2 = 97548$

   a) **Detectar la posible presencia de heteroscedasticidad.**

   b) **Indique cuales serían los efectos que tendría la presencia de heteroscedasticidad sobre los estimadores por MCO y cómo resolvería dichos efectos suponiendo que en este caso la varianza de las perturbaciones depende proporcionalmente del IPC.**

   Para detectar la posible presencia de heteroscedasticidad en el modelo usaremos el test de Goldfeld-Quant, de tal manera que puesto que

$$F_{exp} = \frac{SCR_2}{SCR_1} = \frac{97548}{65432} = 1'4908 \not> 2'2172 = F_{18,18}(0'95) = F_{\frac{n-m}{2},\frac{n-m}{2}}(0'95),$$

   no se rechaza la hipótesis nula de homocedasticidad.

   En el caso de que hubiese existido heteroscedasticidad en el modelo, los estimadores por MCO no serían óptimos. Para resolver esta situación habría que transformar el modelo en uno con perturbaciones esféricas, de tal forma que al aplicarle MCO se obtuvieran estimadores lineales, insesgados y óptimos.

   Así por ejemplo, si se verifica que $Var(u_t) = E\left[u_t^2\right] = \sigma^2 IPC_t$, la matriz para transformar el modelo sería:

$$P = \begin{pmatrix} \frac{1}{\sqrt{IPC_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{IPC_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{IPC_{50}}} \end{pmatrix},$$

   obteniéndose el nuevo modelo transformado, $Y_t^* = \beta_0 X_{1t}^* + \beta_1 X_{2t}^* + u_t^*$, determinado por

$$Y_t^* = \frac{D_t}{\sqrt{IPC_t}}, \quad X_{1t}^* = \frac{1}{\sqrt{IPC_t}}, \quad X_{2t}^* = \frac{DT_t}{\sqrt{IPC_t}}, \quad u_t^* = \frac{u_t}{\sqrt{IPC_t}}, \quad \forall t,$$

[2] Se tiene que para $h = 1$ la $SCR = 9'0994$ y para $h = 2$ la $SCR = 29'1434$. Como en ambos casos la SCT es la misma, el coeficiente de determinación será mayor para $h = 1$.

donde D denota al desempleo y DT a la demanda de trabajo. Además, como la perturbación aleatoria de este modelo transformado verifica, para cualquier valor de t, que:

$$\begin{aligned}
E\left[u_t^*\right] &= E\left[\frac{u_t}{\sqrt{IPC_t}}\right] = \frac{1}{\sqrt{IPC_t}} \cdot E\left[u_t\right] = 0, \\
E\left[{u_t^*}^2\right] &= E\left[\frac{u_t^2}{IPC_t}\right] = \frac{1}{IPC_t} \cdot E\left[u_t^2\right] = \frac{\sigma^2 \cdot IPC_t}{IPC_t} = \sigma^2, \\
Cov\left(u_t^*, u_{t-k}^*\right) &= E\left[u_t^* \cdot u_{t-k}^*\right] = E\left[\frac{u_t}{\sqrt{IPC_t}} \cdot \frac{u_{t-k}}{\sqrt{IPC_{t-k}}}\right] = \frac{E\left[u_t \cdot u_{t-k}\right]}{\sqrt{IPC_t} \cdot \sqrt{IPC_{t-k}}} = 0,
\end{aligned}$$

estamos ante un modelo con perturbaciones esféricas, por lo que al estimarlo por MCO se obtendrán las estimaciones por MCG, que son lineales, insesgadas y óptimas.

5. **Dada la siguiente muestra de gastos en viajes (GV_i) y renta (R_i) correspondiente a diez familias:**

Obs	GV	R
1	12	325
2	20	600
3	10	410
4	20	550
5	15	370
6	8	250
7	10	580
8	15	650
9	25	630
10	10	420

Estime un modelo lineal para explicar los gastos en viajes y detecte la posible existencia de heteroscedasticidad mediante los siguientes métodos:

a) **Método gráfico**

b) **Test de Glesjer**

c) **Test de Goldfeld y Quandt**

d) **Test de White**

e) **Test de Breusch-Pagan**

La estimación del modelo lineal que explica los gastos en viajes mediante la renta a partir de las observaciones consideradas es:

$$\underset{}{\widehat{GV}_t} = \underset{(5'234)}{1'96707} + \underset{(0'0105)}{0'0261921} \cdot R_t, \quad R^2 = 0'4358.$$

Puesto que este tipo de regresiones de sección cruzada suelen presentar heteroscedasticidad en la perturbación aleatoria vamos a estudiar a continuación esta posibilidad.

En primer lugar consideraremos los métodos gráficos. Así, en la figura (1) se tiene el gráfico de dispersión de los residuos frente a cada observación, donde se puede observar que los grupos de observaciones 1-6 y 7-10 tienen distinta varianza. Además, en la figura (2), gráfico de dispersión de los residuos frente a la variable que se sospecha provoca la heteroscedasticidad en el modelo (que en este caso no puede ser otra sino la renta), se observa que la variabilidad de los residuos aumenta conforme lo hace la renta. Todo esto nos hace sospechar que hay heteroscedasticidad en la perturbación aleatoria del modelo y que ésta viene determinada por la variable renta.

Para confirmar esta sospecha recurriremos a los distintos métodos analíticos estudiados.

Figura 1: Gráfico de los residuos

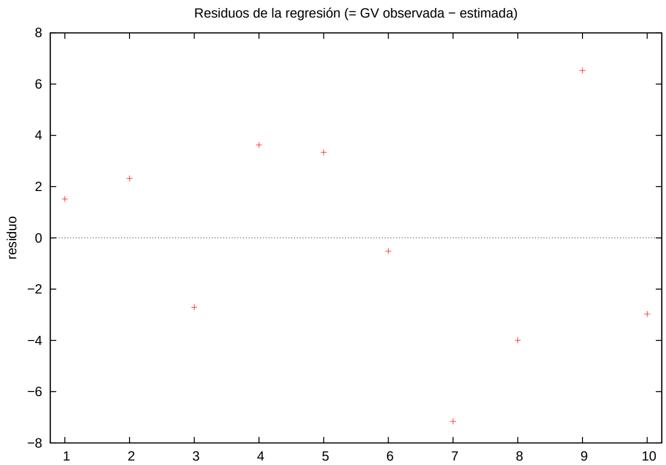
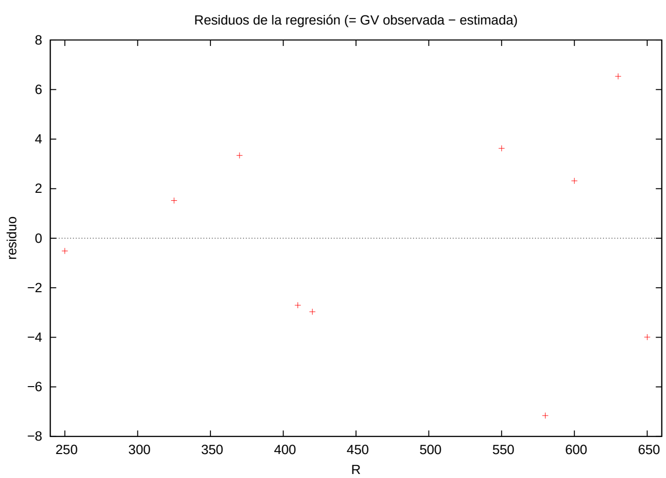


Figura 2: Gráfico de dispersión



Cuadro 2: Información regresiones auxiliares

e_t	GV_t	R_t	$R_t^{1/2}$	$R_t^{-1/2}$	R_t^{-1}
1'5205	12	325	18'0278	0'0555	0'0031
2'3177	20	600	24'4949	0'0408	0'0017
-2'7058	10	410	20'2485	0'0494	0'0024
3'6273	20	550	23'4521	0'0426	0'0018
3'3418	15	370	19'2354	0'052	0'0027
-0'5151	8	250	15'8114	0'0632	0'004
-7'1585	10	580	24'0832	0'0415	0'0017
-3'9919	15	650	25'4951	0'0392	0'0015
6'5319	25	630	25'0998	0'0398	0'0016
-2'9678	10	420	20'4939	0'0488	0'0024

a) Así por ejemplo, empezando por el test de Glesjer, habrá que tener en cuenta la información de la tabla 2 para realizar las regresiones auxiliares:

$$|e_t| = \alpha_0 + \alpha_1 R_t^h + v_t, \quad h = \pm 1, \pm 1/2.$$

En dicha tabla los residuos se han calculado mediante la expresión

$$e = y - X \cdot \widehat{\beta},$$

donde

$$y = \begin{pmatrix} 12 \\ 20 \\ 10 \\ 20 \\ 15 \\ 8 \\ 10 \\ 15 \\ 25 \\ 10 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 325 \\ 1 & 600 \\ 1 & 410 \\ 1 & 550 \\ 1 & 370 \\ 1 & 250 \\ 1 & 580 \\ 1 & 650 \\ 1 & 630 \\ 1 & 420 \end{pmatrix},$$

$$\widehat{\beta} = \begin{pmatrix} 10 & 4785 \\ 4785 & 2467825 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 145 \\ 74050 \end{pmatrix} = \begin{pmatrix} 1'3848 & -0'0027 \\ -0'0027 & 0'00001 \end{pmatrix} \cdot \begin{pmatrix} 145 \\ 74050 \end{pmatrix} = \begin{pmatrix} 1'9671 \\ 0'0262 \end{pmatrix}.$$

Entonces, para $h = 1$:

$$\begin{aligned}
\widehat{\alpha} &= \begin{pmatrix} 1'3848 & -0'0027 \\ -0'0027 & 0'00001 \end{pmatrix} \cdot \begin{pmatrix} 35 \\ 18463 \end{pmatrix} = \begin{pmatrix} -1'5509 \\ 0'0105 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{158'2483 - 139'8624}{10 - 2} = \frac{18'3959}{8} = 2'2982, \\
\widehat{Var}(\widehat{\alpha}) &= \begin{pmatrix} 3'1827 & -0'0062 \\ -0'0062 & 0'0000129 \end{pmatrix}.
\end{aligned}$$

Esto es:

$$\begin{array}{c} |\widehat{e}_t| = -1'5509 + 0'0105 \cdot R_t. \\ (1'784) \quad (0'0036) \end{array}$$

De manera que en el contraste de significación individual de la pendiente se rechaza la hipótesis nula ya que:

$$t_{exp} = \frac{0'0105}{0'0036} = 2'9167 > 2'31 = t_8(0'975).$$

Para $h = 1/2$:

$$\begin{aligned}
\widehat{\alpha} &= \begin{pmatrix} 10 & 216'4 \\ 216'4 & 4785 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 34'6783 \\ 795'4088 \end{pmatrix} \\
&= \begin{pmatrix} 4'7712 & -0'2158 \\ -0'2158 & 0'01 \end{pmatrix} \cdot \begin{pmatrix} 34'6783 \\ 795'4088 \end{pmatrix} = \begin{pmatrix} -6'2063 \\ 0'447 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{158'2483 - 140'2934}{8} = 2'2444, \\
\widehat{Var}(\widehat{\alpha}) &= \begin{pmatrix} 10'7084 & -0'4844 \\ -0'4844 & 0'0224 \end{pmatrix}.
\end{aligned}$$

Esto es:

$$\underset{}{|\widehat{e}_t|} = \underset{(3'273)}{-6'2063} + \underset{(0'149)}{0'447} \cdot R_t^{1/2}.$$

De manera que en el contraste de significación individual de la pendiente se rechaza la hipótesis nula ya que:

$$t_{exp} = \frac{0'447}{0'149} = 2'986 > 2'31 = t_8(0'975).$$

Para $h = -1/2$:

$$\begin{aligned}
\widehat{\alpha} &= \begin{pmatrix} 10 & 0'4729 \\ 0'4729 & 0'0229 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 34'6783 \\ 1'5324 \end{pmatrix} \\
&= \begin{pmatrix} 4 & -83'3 \\ -83'3 & 1762'2 \end{pmatrix} \cdot \begin{pmatrix} 34'6783 \\ 1'5324 \end{pmatrix} = \begin{pmatrix} 12'4377 \\ -189'6627 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{158'2483 - 140'6711}{8} = 2'1971, \\
\widehat{Var}(\widehat{\alpha}) &= \begin{pmatrix} 8'9 & -183'1 \\ -183'1 & 3871'9 \end{pmatrix}.
\end{aligned}$$

Esto es:

$$|\widehat{e}_t| = \underset{(2'983)}{12'4377} - \underset{(62'2245)}{189'6627} \cdot R_t^{-1/2}.$$

De manera que en el contraste de significación individual de la pendiente se rechaza la hipótesis nula ya que:

$$|t_{exp}| = \frac{189'6627}{62'2245} = 3'048 > 2'31 = t_8(0'975).$$

Para $h = -1$:

$$\begin{aligned}
\widehat{\alpha} &= \begin{pmatrix} 10 & 0'0229 \\ 0'0229 & 0'0001 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 34'6783 \\ 0'0687 \end{pmatrix} \\
&= \begin{pmatrix} 1'0173 & -400 \\ -400 & 17440 \end{pmatrix} \cdot \begin{pmatrix} 34'6783 \\ 0'0687 \end{pmatrix} = \begin{pmatrix} 7'8 \\ -1881'1 \end{pmatrix}, \\
\widehat{\sigma}^2 &= \frac{158'2483 - 140'5488}{8} = 2'2124, \\
\widehat{Var}(\widehat{\alpha}) &= \begin{pmatrix} 2'2507 & -884'9156 \\ -884'9156 & 385850 \end{pmatrix}.
\end{aligned}$$

Esto es:

Cuadro 3: Reordenación de las observaciones en función de la renta

GV_t	8	12	15	10	10	20	10	20	25	15
R_t	250	325	370	410	420	550	580	600	630	650

$$|\widehat{e}_t| = \underset{(1'501)}{7'8} - \underset{(621'166)}{1881'1} \cdot R_t^{-1}.$$

De manera que en el contraste de significación individual de la pendiente se rechaza la hipótesis nula ya que:

$$|t_{exp}| = \frac{1881'1}{621'166} = 3'0283 > 2'31 = t_8(0'975).$$

Como en todos los casos se rechaza la hipótesis de que la pendiente sea nula, habrá heteroscedasticidad en el modelo.

b) Para aplicar el test de Goldfeld-Quant reordenamos las observaciones de menor a mayor de acuerdo a la variable renta (obteniendo la tabla 3) y obtenemos las sumas de cuadrados de los residuos de las regresiones de los subgrupos que surgen al eliminar las 4 observaciones centrales.

Así, para el primer subgrupo se tiene

$$y_1 = \begin{pmatrix} 8 \\ 12 \\ 15 \end{pmatrix}, \quad X_1 = \begin{pmatrix} 1 & 250 \\ 1 & 325 \\ 1 & 370 \end{pmatrix},$$

obteniéndose

$$\widehat{\beta}_1 = \begin{pmatrix} -6'5476 \\ 0'0578 \end{pmatrix}, \quad SCR_1 = y_1^t y_1 - \widehat{\beta}_1^t X_1^t y_1 = 433 - 432'9082 = 0'0918.$$

Mientras que para el segundo

$$y_2 = \begin{pmatrix} 20 \\ 25 \\ 15 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 600 \\ 1 & 630 \\ 1 & 650 \end{pmatrix},$$

obteniéndose

$$\widehat{\beta}_2 = \begin{pmatrix} 69'4737 \\ -0'0789 \end{pmatrix}, \quad SCR_2 = y_2^t y_2 - \widehat{\beta}_2^t X_2^t y_2 = 1250 - 1207'9 = 42'1053.$$

Entonces:

$$F_{exp} = \frac{42'1053}{0'0918} = 458'6634 > 9'27663 = F_{3,3}(0'95),$$

y en tal caso, se rechaza la hipótesis nula de homocedasticidad. Es decir, la perturbación aleatoria del modelo es heterocedástica.

c) En el caso del test de White habrá que calcular el coeficiente de determinación de la regresión auxiliar

$$e_t^2 = \alpha_0 + \alpha_1 \cdot R_t + \alpha_2 \cdot R_t^2 + v_t.$$

Atendiendo a la información proporcionada por la tabla 4 se tiene que

$$\widehat{\alpha} = \begin{pmatrix} -21'3268 \\ 0'0811 \\ -0'0000067 \end{pmatrix}, \quad SCR = 5163'5 - 3504'1 = 1659'3,$$

Cuadro 4: Información regresiones auxiliares

e_t^2	R_t	R_t^2
2’3119	325	105625
5’3715	600	360000
7’3216	410	168100
13’1570	550	302500
11’1679	370	136900
0’2653	250	62500
51’2441	580	336400
15’9356	650	422500
42’6657	630	396900
8’8076	420	176400

$$\overline{e^2} = 15'8248, \quad SCT = 5163'5 - 10 \cdot 15'8248 = 2659'3.$$

Y en tal caso:

$$R^2 = 1 - \frac{1659'3}{2659'3} = 0'376.$$

Por tanto, como

$$\chi^2_{exp} = 10 \cdot 0'376 = 3'76 \not> 5'991 = \chi^2_2(0'95),$$

no se rechaza la hipótesis nula de homocedasticidad.

d) Para el test de Breusch-Pagan la regresión auxiliar es

$$e_t^2 = \alpha_0 + \alpha_1 \cdot R_t + v_t.$$

Atendiendo a la información proporcionada por la tabla 4 se tiene que

$$\widehat{\alpha} = \begin{pmatrix} -20'0151 \\ 0'0749 \end{pmatrix}, \quad SCR = 5163'5 - 3504 = 1659'6,$$

$$\overline{e^2} = 15'8248, \quad SCT = 5163'5 - 10 \cdot 15'8248 = 2659'3.$$

Y en tal caso:

$$R^2 = 1 - \frac{1659'6}{2659'3} = 0'3759.$$

Por tanto, como

$$\chi^2_{exp} = 10 \cdot 0'3759 = 3'759 \not> 3'8414 = \chi^2_1(0'95),$$

no se rechaza la hipótesis nula de homocedasticidad.

Podemos observar que a partir de los dos primeros test concluiríamos que hay heteroscedasticidad en el modelo mientras que a partir de los dos últimos no rechazaríamos la hipótesis nula de homocedasticidad. Esta contradicción se puede deber a que los dos últimos test se deben usar cuando la muestra es grande y no tan pequeña como en este caso. Por tanto, como los dos primeros test son idóneos para situaciones donde la muestra es pequeña y una variable es la causante de la heteroscedasticidad (situación de este modelo), consideraremos que la perturbación aleatoria del modelo que estudia los gastos en viajes a partir de la renta es heterocedástica.

En tal caso, a partir de las sumas de cuadrados de los residuos[3] de las regresiones auxiliares (ver tabla 5) calculadas al aplicar el test de Glesjer (en la estimación de σ^2), podemos concluir que la perturbación aleatoria es proporcional a la inversa de la raíz cuadrada de la renta, esto es, $Var\,(u_t) = \frac{\sigma^2}{\sqrt{R_t}} = \frac{\sigma^2}{R_t^{1/2}}$.

[3]Puesto que la suma de cuadrados totales es la misma en todas las regresiones, el mayor coeficiente de determinación corresponderá a la menor suma de cuadrados de los residuos.

Cuadro 5: Suma de cuadrados de los residuos de las regresiones auxiliares del test de Glesjer

h	1	1/2	-1/2	-1
SCR	18'3959	17'9549	17'5772	17'6995

Cuadro 6: Observaciones del modelo transformado

Y_t^*	X_{1t}^*	X_{2t}^*
50'9509	4'245911	1379'921
98'9846	4'949232	2969'539
44'9983	4'499829	1844'930
96'8547	4'842735	2663'504
65'7872	4'385816	1622'752
31'8108	3'976354	994'088
49'0746	4'907463	2846'328
75'7390	5'049267	3282'024
125'2493	5'009970	3156'281
45'2702	4'527019	1901'348

Entonces, la matriz para transformar el modelo sería:

$$P = \begin{pmatrix} R_1^{1/4} & 0 & \cdots & 0 \\ 0 & R_2^{1/4} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{10}^{1/4} \end{pmatrix},$$

obteniéndose el nuevo modelo transformado, $Y_t^* = \beta_0 X_{1t}^* + \beta_1 X_{2t}^* + u_t^*$, determinado por

$$Y_t^* = GV_t \cdot R_t^{1/4}, \quad X_{1t}^* = 1 \cdot R_t^{1/4} = R_t^{1/4}, \quad X_{2t}^* = R_t \cdot R_t^{1/4} = R_t^{5/4}, \quad u_t^* = u_t \cdot R_t^{1/4}, \quad \forall t.$$

En la tabla 6 se tienen las observaciones del modelo transformado.

Además, se comprueba fácilmente que el nuevo modelo es un modelo con perturbaciones esféricas. En efecto, para cualquier valor de t, se verifica que:

$$\begin{aligned} E\left[u_t^*\right] &= E\left[u_t \cdot R_t^{1/4}\right] = R_t^{1/4} \cdot E\left[u_t\right] = 0, \\ E\left[{u_t^*}^2\right] &= E\left[u_t^2 \cdot R_t^{1/2}\right] = R_t^{1/2} \cdot E\left[u_t^2\right] = R_t^{1/2} \cdot \frac{\sigma^2}{R_t^{1/2}} = \sigma^2, \\ Cov\left(u_t^*, u_{t-k}^*\right) &= E\left[u_t^* \cdot u_{t-k}^*\right] = E\left[u_t \cdot R_t^{1/4} \cdot u_{t-k} \cdot R_{t-k}^{1/4}\right] = R_t^{1/4} \cdot R_{t-k}^{1/4} \cdot E\left[u_t \cdot u_{t-k}\right] = 0. \end{aligned}$$

Por tanto, al estimarlo por MCO se obtendrán las estimaciones por MCG, que son lineales, insesgadas y óptimas:

$$\underset{}{\widehat{GV}_t} = \underset{(5'85113)}{1'84598} + \underset{(0'0113727)}{0'0264311} \cdot R_t, \quad R^2 = 0'932014.$$

6. **Dado el siguiente modelo estimado por MCO:**

$$\widehat{PIB}_t = \underset{(182'9)}{-187'7} + \underset{(0'19)}{3'76} \cdot OCU_t, \quad t = 1, \ldots, 18$$

Se han realizado contrastes para analizar la existencia de homocedasticidad, rechazándose en todo caso esta hipótesis. A la vista de las siguientes regresiones, ¿cómo eliminaría la heteroscedasticidad?

$$|\widehat{e}_t| = -68600'1 + \underset{(123'4)}{419'7} \cdot OCU_t, \quad R^2 = 0'38$$

$$|\widehat{e}_t| = 77998'8 + \underset{(0'06)}{0'16} \cdot OCU_t^2, \quad R^2 = 0'3$$

¿Qué consecuencias conlleva la no existencia de homocedasticidad en el modelo?

Teniendo en cuenta que:

$$t_{exp} = \frac{419'7}{123'4} = 3'4011 > 2'12 = t_{16}(0'975),$$
$$t_{exp} = \frac{0'16}{0'06} = 2'6667 > 2'12 = t_{16}(0'975),$$

en ambos casos se rechaza la hipótesis nula de que la pendiente sea cero, por lo que aplicando el test de Glesjer concluimos que efectivamente hay heteroscedasticidad en el modelo.

Por otro lado, puesto que la primera regresión auxiliar tiene un mayor coeficiente de determinación consideraremos que la varianza de la perturbación aleatoria es proporcional a la variable OCU, es decir, $Var(u_t) = \sigma^2 \cdot OCU_t$. Por tanto, para eliminar la heteroscedasticidad del modelo lo transformaría mediante la matriz

$$P = \begin{pmatrix} \frac{1}{\sqrt{OCU_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{OCU_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{OCU_{18}}} \end{pmatrix}.$$

Finalmente, la no existencia de homocedasticidad en el modelo hace que los estimadores obtenidos por MCO no sean óptimos.

7. **Utilizando una muestra de 25 observaciones anuales se ha estimado el siguiente modelo:**

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t.$$

Utilizando sólo las primeras 10 observaciones se obtiene la siguiente ecuación estimada:

$$\widehat{Y}_t = 80'5 + 0'93 X_{2t} - 0'87 X_{3t}, \qquad SCR = 125'7,$$

mientras que para las últimas 10 observaciones se obtiene la siguiente estimación:

$$\widehat{Y}_t = 20'61 + 0'53 X_{2t} - 0'105 X_{3t}, \qquad SCR = 498'94.$$

Además, se dispone de la siguiente información:

$$|e_t| = 6'81 - 625'17 \cdot \frac{1}{X_{2t}}, \quad R^2 = 0'43,$$

$$|e_t| = 10'23 - 89'54 \cdot \frac{1}{\sqrt{X_{2t}}}, \quad R^2 = 0'33.$$

Se pide:

a) **Detectar si hay presencia de heteroscedasticidad en el modelo y, en tal caso, indicar qué variable la induce.**

b) **Especificar cuál sería la matriz de transformación más adecuada para solucionar la heteroscedasticidad, en caso de que la hubiera.**

Para detectar si hay presencia de heteroscedasticidad en el modelo usaremos el test de Goldfeld-Quant. Puesto que:

$$F_{exp} = \frac{SCR_2}{SCR_1} = \frac{498'94}{125'7} = 3'97 > 2'98 = F_{10,10}(0'95) = F_{\frac{n-m}{2},\frac{n-m}{2}}(0'95),$$

se rechaza la hipótesis nula de homocedasticidad (adviértase que hemos tenido en cuenta que se disponen de 25 observaciones, $n = 25$, de las cuales se eliminan 5 centrales, $m = 5$). Esto es, hay heteroscedasticidad en el modelo.

Por otro lado, atendiendo a las regresiones en las que la variable dependiente es el valor absoluto de los residuos, $|e_t|$, la heteroscedasticidad la induce la variable X_{2t} y consideraremos que $Var(u_t) = \sigma^2 \cdot \frac{1}{X_{2t}}$. Esta elección se debe a que de los dos modelos proporcionados es el primero el que ofrece un mejor ajuste (mayor coeficiente de determinación). En tal caso, la matriz de transformación del modelo será:

$$P = \begin{pmatrix} \sqrt{X_{21}} & 0 & \cdots & 0 \\ 0 & \sqrt{X_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{X_{2\ 25}} \end{pmatrix}.$$

8. **Suponiendo que existe heteroscedasticidad en un modelo lineal, $V_t = \beta_0 + \beta_1 P_t + u_t$, que estudia las ventas en función del precio y que la relación entre la varianza de las perturbaciones aleatorias y el precio es cuadrática, indique como transformaría el modelo para que sea homocedástico.**

   Puesto que la relación entre la varianza de las perturbaciones aleatorias y el precio es cuadrática entonces se verifica que $Var(u_t) = E(u_t^2) = \sigma^2 \cdot P_t^2$. En tal caso, para que el modelo sea homocedástico lo transformaría mediante la siguiente matriz:

$$P = \begin{pmatrix} \frac{1}{P_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{P_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{P_n} \end{pmatrix}.$$

   De esta forma, el nuevo modelo transformado es un modelo con perturbaciones esféricas (por tanto, homocedástico) y al estimarlo por MCO obtendremos estimaciones lineales, insesgadas y óptimas.

9. **En un modelo lineal para estimar el consumo (C) en función de la renta (R) se han ordenado los datos de menor a mayor y se han obtenido las siguientes regresiones:**

$$\widehat{C}_t = 1'1 + 1'5 \cdot R_t, \quad \sum_{t=1}^{10} e_t^2 = 230,$$

$$\widehat{C}_t = 20'1 + 0'5 \cdot R_t, \quad \sum_{t=21}^{30} e_t^2 = 43230.$$

   Analice la presencia de heteroscedasticidad en el modelo.

   Para detectar si hay presencia de heteroscedasticidad en el modelo usaremos el test de Goldfeld-Quant. Puesto que:

$$F_{exp} = \frac{SCR_2}{SCR_1} = \frac{43230}{230} = 187'956 > 2'98 = F_{10,10}(0'95) = F_{\frac{n-m}{2},\frac{n-m}{2}}(0'95),$$

se rechaza la hipótesis nula de homocedasticidad (adviértase que hemos tenido en cuenta que se disponen de 30 observaciones, $n = 30$, de las cuales se eliminan 10 centrales, $m = 10$). Luego, hay heteroscedasticidad en el modelo.

10. **Haciendo uso de los datos per cápita sobre el gasto público en servicios sociales (GPSS) y el PIB de 34 países en 2009, se ha obtenido por MCO la siguiente estimación:**

$$\widehat{GPSS}_i = -0'1245 + 0'0731 \cdot PIB_i.$$

 a) **Conociendo la siguiente regresión auxiliar**

 $$\widehat{e}_i^2 = 0'0176 - 0'0052 \cdot PIB_i + 0'0004 \cdot PIB_i^2,$$

 con $R^2 = 0'9582$, contraste si hay heteroscedasticidad en el modelo.

 b) **Suponiendo que existe heteroscedasticidad y bajo el supuesto $E[u_i^2] = \dfrac{\sigma^2}{PIB_i^2}$, especifique la matriz de transformación que corrija dicho problema.**

 Aplicando el test de White hay heteroscedasticidad en el modelo ya que:

 $$n \cdot R_{aux}^2 = 34 \cdot 0'9582 = 32'5788 > 5'991 = \chi_2^2(0'95).$$

 Además, como $E[u_i^2] = \frac{\sigma^2}{PIB_i^2}$, es claro que la matriz de transformación que corrije la heteroscedasticidad es:

 $$P = \begin{pmatrix} PIB_1 & 0 & \cdots & 0 \\ 0 & PIB_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & PIB_{34} \end{pmatrix}.$$

11. **Supongamos que tras estimar el modelo original, $Y_t = \beta_1 + \beta_2 X_t + u_t$, por MCO se ha realizado la siguiente regresión auxiliar:**

$$|\widehat{e}_t| = \alpha_0 + \frac{\alpha_1}{X_t} + v_t,$$

 rechazándose la hipótesis $H_0 : \alpha_1 = 0$. ¿Qué conclusiones se pueden obtener? ¿Existe algún problema en el modelo? En caso afirmativo, describa detalladamente como resolverlo comprobando que se ha solucionado.

 Teniendo en cuenta el test de Glesjer se puede concluir que existe heteroscedasticidad en el modelo original (y que $Var(u_t) = \sigma^2 \cdot \frac{1}{X_t}$), el cual habría que transformar multiplicando por la siguiente matriz:

 $$P = \begin{pmatrix} \sqrt{X_1} & 0 & \cdots & 0 \\ 0 & \sqrt{X_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sqrt{X_n} \end{pmatrix},$$

 para obtener el modelo transformado:

 $$Y_t^* = \sqrt{X_t} \cdot Y_t, \quad X_t^* = \sqrt{X_t} \cdot X_t, \quad cte^* = \sqrt{X_t}, \quad u_t^* = \sqrt{X_t} \cdot u_t, \quad \forall t,$$

 el cual ya es homocedástico:

 $$Var(u_t^*) = Var\left(\sqrt{X_t} \cdot u_t\right) = X_t \cdot Var(u_t) = X_t \cdot \frac{\sigma^2}{X_t} = \sigma^2.$$

12. **Supongamos que para ilustrar la detección de la heteroscedasticidad el profesor de Econometría considera un modelo que trata de explicar el número de accidentes de tráfico, Y, en función de los años de conducción, X_1, y de la edad, X_2, a partir de 90 observaciones. Tras realizar el análisis de los procedimientos gráficos, dicho profesor sospecha que la variable edad puede ser causa de heteroscedasticidad en el modelo, por lo que procede a ordenar la muestra de forma creciente en función de dicha variable. De las 90 observaciones omite 30 centrales y ajusta por MCO los dos grupos de observaciones restantes obteniéndose que $\widehat{\sigma}_1^2 = 30'7$ para el primero y que $\sum\limits_{t=61}^{90} e_t^2 = 1428'32$ para el segundo. ¿Existe heteroscedasticidad en el modelo?**

    Un grupo de alumnos descontentos con la explicación del profesor consideran que la heteroscedasticidad la provoca más de una variable, por lo que estiman la siguiente regresión auxiliar:

    $$\widehat{e}_t^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1^2 + \alpha_4 X_2^2 + \alpha_5 X_1 X_2 + v_t,$$

    obteniendo un coeficiente de determinación $R^2 = 0'63$. ¿Existe heteroscedasticidad en el modelo?

    ¿Hay algún tipo de contradicción entre las conclusiones obtenidas por el profesor y los alumnos? ¿Cómo resolverla?

    A partir del test de Goldfeldt-Quant no se rechaza la hipótesis nula de homocedasticidad ya que

    $$F_{exp} = \frac{1428'32}{828'9} = 1'723151 \not> 1'84087 = F_{30,30}(0'95),$$

    donde se ha usado que $SCR_1 = (n_1 - k_1) \cdot \widehat{\sigma}_1^2 = 27 \cdot 30'7 = 828'9$.

    Por otro lado, teniendo en cuenta el test de White si se rechaza la hipótesis nula de homocedasticidad ya que

    $$\chi^2_{exp} = 90 \cdot 0'63 = 56'7 > 11'07 = \chi^2_5(0'95).$$

    El conflicto radica en que el primer test se debe aplicar cuando la muestra es pequeña, cosa que no ocurre en este caso.

13. **Supongamos que se disponen de las siguientes observaciones para el modelo $Y_t = \alpha + \beta X_t + u_t$:**

Y_t	X_t
20	**11**
30	**16**
32	**18**

    Si se verifica que $Var(u_t) = \sigma^2 \cdot X_t$, ¿aplicar MCG al modelo anterior es equivalente a aplicar MCO al siguiente conjunto de datos? Razone su respuesta.

Y_t^*	X_t^*
6'0302	**3'3166**
7'5	**4**
7'5424	**4'2426**

    Para corregir la heteroscedasticidad hay que transformar los datos dividiendo entre $\sqrt{X_t}$ obteniéndose los datos que se indican. Luego es equivalente aplicar MCG al modelo original y aplicar MCO al modelo transformado.

4. Autocorrelación

1. **Para la estimación de la producción de gases contaminantes por una fábrica de papel se ha utilizado el siguiente modelo:**

$$\begin{aligned} C_t &= \beta_0 + \beta_1 V_t + \epsilon_t, \\ \epsilon_t &= \rho\epsilon_{t-1} + \nu_t, \end{aligned}$$

para $t = 1, \cdots, T$, y donde C denota la producción de gases contaminantes y V al volumen de papel fabricado. Comente qué supuesto del modelo lineal general se incumple en este caso si se intenta estimar la primera ecuación por mínimos cuadrados ordinarios.

En este caso, puesto que la perturbación aleatoria del modelo de regresión tiene una estructura autorregresiva de orden 1, la misma no será incorrelada. En efecto:

$$\begin{aligned} Cov\left(\epsilon_t, \epsilon_{t-1}\right) &= E\left[\epsilon_t \cdot \epsilon_{t-1}\right] = E\left[\left(\rho\epsilon_{t-1} + \nu_t\right) \cdot \epsilon_{t-1}\right] \\ &= E\left[\rho\epsilon_{t-1}^2 + \nu_t\epsilon_{t-1}\right] = \rho E\left[\epsilon_{t-1}^2\right] + E\left[\nu_t\epsilon_{t-1}\right] = \rho \cdot \sigma^2 \neq 0, \end{aligned}$$

donde se ha supuesto que $Var\left(\epsilon_t\right) = \sigma^2$, $\forall t$, y que los procesos ϵ y ν están incorrelados.

2. **A partir del modelo:**

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

y suponiendo que las perturbaciones tienen varianza constante y siguen el siguiente proceso $u_t = 0'5u_{t-1} + \nu_t$. Se pide obtener el estimador MCG del modelo sabiendo que disponemos de los siguientes datos:

Y	**4**	**6**	**10**
X	**2**	**3**	**6**

Evidentemente habrá autocorrelación en la perturbación aleatoria (ejercicio 1), de forma que, tal y como indica el enunciado, habrá que estimar el modelo por mínimos cuadrados generalizados. Sin embargo, para poder aplicar dicho método es necesario conocer la matriz Ω. Por tanto, el primer paso que hay que realizar es obtener dicha matriz.

En efecto, puesto que:

$$\begin{aligned} E\left[u_t \cdot u_{t-1}\right] &= E\left[\left(0'5u_{t-1} + \nu_t\right) \cdot u_{t-1}\right] \\ &= 0'5 \cdot E\left[u_{t-1}^2\right] + E\left[\nu_t \cdot u_{t-1}\right] = 0'5 \cdot \sigma^2, \\ E\left[u_t \cdot u_{t-2}\right] &= E\left[\left(0'5u_{t-1} + \nu_t\right) \cdot u_{t-2}\right] \\ &= 0'5 \cdot E\left[u_{t-1} \cdot u_{t-2}\right] + E\left[\nu_t \cdot u_{t-2}\right] = 0'5 \cdot 0'5 \cdot \sigma^2 = 0'25 \cdot \sigma^2, \end{aligned}$$

donde sa ha usado que ν_t y u_t están incorrelados, se tiene que:

$$\Sigma = \begin{pmatrix} \sigma^2 & 0'5\sigma^2 & 0'25\sigma^2 \\ 0'5\sigma^2 & \sigma^2 & 0'5\sigma^2 \\ 0'25\sigma^2 & 0'5\sigma^2 & \sigma^2 \end{pmatrix} = \sigma^2 \cdot \begin{pmatrix} 1 & 0'5 & 0'25 \\ 0'5 & 1 & 0'5 \\ 0'25 & 0'5 & 1 \end{pmatrix},$$

esto es:

$$\Omega = \begin{pmatrix} 1 & 0'5 & 0'25 \\ 0'5 & 1 & 0'5 \\ 0'25 & 0'5 & 1 \end{pmatrix}.$$

En tal caso:

$$\begin{aligned} \widehat{\beta} &= \left(X^t\Omega^{-1}X\right)^{-1} \cdot X^t\Omega^{-1}y = \begin{pmatrix} 1'6667 & 6'3334 \\ 6'3334 & 36'3334 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 11'3334 \\ 60'6667 \end{pmatrix} \\ &= \begin{pmatrix} 1'7772 & -0'3098 \\ -0'3098 & 0'0815 \end{pmatrix} \cdot \begin{pmatrix} 11'3334 \\ 60'6667 \end{pmatrix} = \begin{pmatrix} 1'3478 \\ 1'4348 \end{pmatrix}, \end{aligned}$$

donde se ha usado que

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}, \quad y = \begin{pmatrix} 4 \\ 6 \\ 10 \end{pmatrix}, \quad \Omega^{-1} = \begin{pmatrix} 1'3334 & -0'6667 & 0 \\ -0'6667 & 1'6667 & -0'6667 \\ 0 & -0'6667 & 1'3334 \end{pmatrix}.$$

Luego, la estimación por MCG será $\widehat{Y}_t = 1'3478 + 1'4348 \cdot X_t$.

3. **Se dispone de una serie de 25 datos que relacionan el salario nominal, X_t, y el empleo, Y_t. Una vez hecha la estimación los residuos obtenidos son:**

t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
e_t	**2'6**	**2'8**	**-2'3**	**0'6**	**-0'75**	**0'12**	**1'2**	**2'5**	**-3**	**-2'4**	**0'5**	**0'7**

t	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**
e_t	**0'8**	**-6**	**-1**	**-2**	**2'3**	**2'4**	**4**	**3'43**	**2'3**	**-1'1**	**-2'2**	**-2'55**	**?**

Además, se han realizado las regresiones sobre los 10 primeros datos y sobre los diez últimos obteniéndose $SCR_1 = 75'43$ y $SCR_2 = 91'4$. Contrastar la existencia de autocorrelación y heteroscedasticidad.

Para contrastar la presencia de heteroscedasticidad en el modelo usaremos el contraste de Godlfeld-Quant:

$$F_{exp} = \frac{91'4}{75'43} = 1'21172 \not> 2'98 = F_{10,10}(0'95).$$

Por tanto, no se rechaza la hipótesis nula de que la perturbación aleatoria sea homocedástica.

Por otro lado, para contrastar la presencia de autocorrelación calcularemos el estadístico de Durbin-Watson a partir de la información de la tabla 7:

$$d = \frac{186'9201}{151'0168} = 1'2377.$$

Como $d_L = 1'2879$ y $d_U = 1'4537$, entonces hay autocorrelación positiva en la perturbación aleatoria del modelo ya que $d < d_L$.

4. **Contrastar la existencia de autocorrelación de primer orden en el modelo:**

$$\widehat{P}_t = -2'884 + 0'462 S_t + 0'184 P_{t-1}$$

sabiendo que tenemos una muestra de quince observaciones y que:

$$\widehat{Var\left(\widehat{\beta}\right)} = \widehat{\sigma}^2 \left(X'X\right)^{-1} = \begin{pmatrix} 19'39 & -0'384 & -0'175 \\ -0'384 & 0'0153 & -0'01258 \\ -0'175 & -0'0125 & 0'03577 \end{pmatrix}$$

$$d = 1'75021$$

Puesto que como regresora aparece la variable dependiente retardada para estudiar la autocorrelación en este modelo hay que utilizar la h de Durbin. En tal caso, se rechaza la hipótesis nula de incorrelación si

$$|h| = \left| \rho \cdot \sqrt{\frac{n}{1 - n \cdot var}} \right| > Z_{1-\frac{\alpha}{2}},$$

donde var es la varianza estimada del coeficiente correspondiente a la variable retardada y $Z_{1-\frac{\alpha}{2}}$ es el punto de una distribución N(0,1) que deja a su izquierda una probabilidad $1 - \frac{\alpha}{2}$.

Cuadro 7: Cálculo del estadístico de Durbin-Watson

e_t	e_t^2	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
2’6	6’76			
2’8	7’84	2’6	0’2	0’04
-2’3	7’84	2’8	-5’1	26’01
0’6	0’36	-2’3	2’9	8’41
-0’75	0’5625	0’6	-1’35	1’8225
0’12	0’0144	-0’75	0’87	0’7569
1’2	1’44	0’12	1’08	1’1664
2’5	6’25	1’2	1’3	1’69
-3	9	2’5	-5’5	20’25
-2’4	5’76	-3	0’6	0’36
0’5	0’25	-2’4	2’9	8’41
0’7	0’49	0’5	0’2	0’04
0’8	0’64	0’7	0’1	0’01
-6	36	0’8	-6’8	46’24
-1	1	-6	5	25
-2	4	-1	-1	1
2’3	5’29	-2	4’3	18’49
2’4	5’76	2’3	0’1	0’01
4	16	2’4	1’6	2’56
3’43	11’7649	4	-0’57	0’3249
2’3	5’29	3’43	-1’13	1’2769
-1’1	1’21	2’3	-3’4	11’56
-2’2	4’84	-1’1	-1’1	1’21
-2’55	6’5025	-2’2	-0’35	0’1225
-2’95	8’7025	-2’55	-0’4	0’16
	151’0168			186’9201

Es evidente que $n = 15$ y $var = 0'03577$. Por otro lado, como $d = 1'75021$ se tiene que $\rho \simeq 1 - \frac{1'75021}{2} = 0'1249$.

Luego, sin más que sustituir:

$$|h| = \left| 0'1249 \cdot \sqrt{\frac{15}{1 - 15 \cdot 0'03577}} \right| = 0'71057 \not> 1'96 = Z_{0'975}.$$

Por tanto, no rechazo la hipótesis nula de incorrelación.

5. **Se ha recogido información de la economía española para el periodo 1985-1990 del consumo público y el PIB con objeto de estimar un modelo de regresión lineal que explique el consumo público. Se ha llegado al siguiente modelo estimado por MCO:**

$$\hat{C}_t = 2'1864 + 0'0796 PIB_t$$

con:

$$\sum_{t=2}^{6} (e_t - e_{t-1})^2 = 0'049327$$

$$\sum_{t=1}^{6} e_t^2 = 0'0161$$

Se pide contrastar la existencia de autocorrelación.

En este caso

$$d = \frac{0'049327}{0'0161} = 3'0638, \quad d_L = 0'6101, \quad d_U = 1'4002.$$

Como $2'5998 = 4 - d_U < d < 4 - d_L = 3'3898$, el contraste de Durbin-Watson no es concluyente.

6. **El número de pequeños accidentes ocurridos en las calles de una ciudad (Y) y el número de coches matriculados en la misma (X) durante 10 años han sido los siguientes:**

Y	**X**
25	**510**
27	**520**
28	**528**
32	**540**
33	**590**
36	**650**
38	**700**
40	**760**
41	**800**
45	**870**

Dado el modelo $Y_t = \beta_0 + \beta_1 X_t + u_t$, se pide:

a) **Estimar la recta que exprese el número de accidentes ocurridos en función del número de coches matriculados.**

b) **Calcular el estadístico Durbin-Watson y detectar la posible existencia de autocorrelación.**

c) **Aplicar mínimos cuadrados generalizados para solucionar el posible problema de autocorrelación comprobando que realmente ha sido resuelto.**

Para estimar la recta que exprese el número de accidentes ocurridos en función del número de coches matriculados tendremos encuenta que

$$y = \begin{pmatrix} 25 \\ 27 \\ 28 \\ 32 \\ 33 \\ 36 \\ 38 \\ 40 \\ 41 \\ 45 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 510 \\ 1 & 520 \\ 1 & 528 \\ 1 & 540 \\ 1 & 590 \\ 1 & 650 \\ 1 & 700 \\ 1 & 760 \\ 1 & 800 \\ 1 & 870 \end{pmatrix},$$

de forma que

$$\begin{aligned} \widehat{\beta} &= \left(X^t X\right)^{-1} X^t y = \begin{pmatrix} 10 & 6468 \\ 6468 & 4335984 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 345 \\ 230674 \end{pmatrix} \\ &= \begin{pmatrix} 2'8436 & -0'0042 \\ -0'0042 & 0'00001 \end{pmatrix} \cdot \begin{pmatrix} 345 \\ 230674 \end{pmatrix} = \begin{pmatrix} 2'5676 \\ 0'0494 \end{pmatrix}. \end{aligned}$$

Por tanto, la estimación buscada es

$$\widehat{Y_t} = 2'5676 + 0'0494 \cdot X_t.$$

Además, a partir de dicha estimación se calculan los residuos como:

$$e = y - \widehat{y} = y - X \cdot \widehat{\beta} = \begin{pmatrix} 25 \\ 27 \\ 28 \\ 32 \\ 33 \\ 36 \\ 38 \\ 40 \\ 41 \\ 45 \end{pmatrix} - \begin{pmatrix} 27'7462 \\ 28'2399 \\ 28'6349 \\ 29'2273 \\ 31'6958 \\ 34'658 \\ 37'1265 \\ 40'0887 \\ 42'0635 \\ 45'5194 \end{pmatrix} = \begin{pmatrix} -2'7462 \\ -1'2399 \\ -0'6349 \\ 2'7727 \\ 1'3042 \\ 1'342 \\ 0'8735 \\ -0'0887 \\ -1'0635 \\ -0'5194 \end{pmatrix}.$$

En las representaciones gráficas de las figuras 3 y 4 sobre los residuos anteriores podemos observar en la primera rachas de valores por encima y por debajo del cero y en la segunda una tendencia creciente. Ambas situaciones nos hace pensar en la presencia de autocorrelación positiva en la perturbación aleatoria del modelo.

En efecto, teniendo en cuenta los residuos se obtiene la información de la tabla 8 y a partir de éstos se tiene que:

$$d = \frac{18'796}{22'8435} = 0'8228.$$

En tal caso, como $d_L = 0'8791$ y $d_U = 1'3197$, entonces hay autocorrelación positiva ya que $d < d_L$.

Para resolver este problema de autocorrelación utilizaremos el método de Prais-Winsten para transformar los datos ya que se disponen de pocas observaciones. Entonces:

$$Y_t^* = \begin{cases} \sqrt{1-\rho^2} \cdot Y_1 \\ Y_t - \rho \cdot Y_{t-1}, \ t > 1 \end{cases}, \quad X_{it}^* = \begin{cases} \sqrt{1-\rho^2} \cdot X_{i1}, \ i = 1,2 \\ X_{it} - \rho \cdot X_{i\ t-1}, \ t > 1, \ i = 1,2 \end{cases},$$

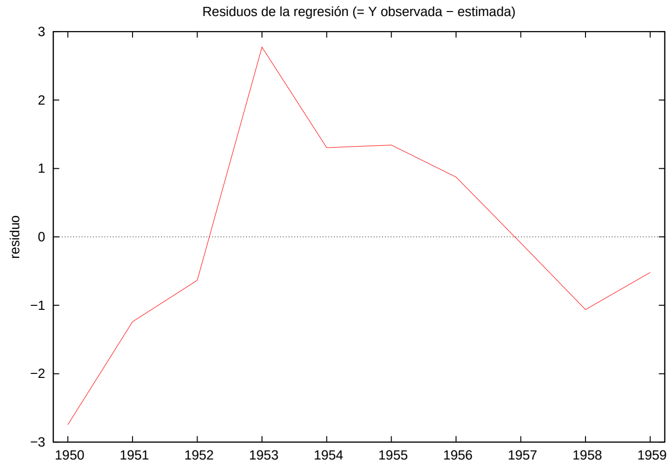


Figura 3: Gráfico temporal de los residuos

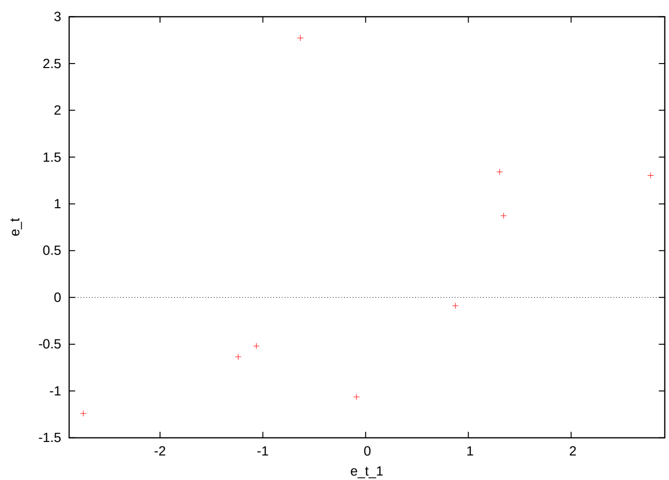


Figura 4: Gráfico de dispersión

Cuadro 8: Cálculo del estadístico de Durbin-Watson

e_t	e_t^2	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
-2'7462	7'5416			
-1'2399	1'5373	-2'7462	1'5063	2'2689
-0'6349	0'403	-1'2399	0'605	0'3661
2'7727	7'6879	-0'6349	3'4076	11'6115
1'3042	1'701	2'7727	-1'4685	2'1565
1'342	1'801	1'3042	0'0378	0'0014
0'8735	0'763	1'342	-0'4685	0'2195
-0'0887	0'0079	0'8735	-0'9622	0'9258
-1'0635	1'131	-0'0887	-0'9748	0'9502
-0'5194	0'2697	-1'0635	0'5441	0'2961
	22'8435			18'796

Cuadro 9: Observaciones del modelo transformado

Y_t^*	cte^*	X_t^*
20'2106	0'8084	412'2965
12'285	0'4114	219'814
12'1078	0'4114	221'928
15'5192	0'4114	229'2192
14'1648	0'4114	272'156
16'5762	0'4114	302'726
16'8104	0'4114	317'41
17'6332	0'4114	347'98
17'456	0'4114	352'664
20'8674	0'4114	399'12

Cuadro 10: Cálculo del estadístico de Durbin-Watson

y^*	$\widehat{y}^*$	e_t	e_t^2	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
20'2106	22'1014	-1'8908	3'5751			
12'285	11'7439	0'5411	0'2928	-1'8908	2'4319	5'9139
12'1078	11'8489	0'2589	0'067	0'5411	-0'2822	0'0796
15'5192	12'2111	3'3081	10'9435	0'2589	3'0492	9'2979
14'1648	14'3438	-0'179	0'032	3'3081	-3'4871	12'1599
16'5762	15'8622	0'714	0'5097	-0'179	0'893	0'7974
16'8104	16'5916	0'2188	0'0479	0'714	-0'4952	0'2452
17'6332	18'11	-0'4768	0'2274	0'2188	-0'6956	0'4839
17'456	18'3427	-0'8867	0'7862	-0'4768	-0'4099	0'168
20'8674	20'6502	0'2172	0'0472	-0'8867	1'1039	1'2186
			16'5288			30'3644

donde el valor de ρ se puede obtener a partir del estadístico de Durbin-Watson:

$$d = 0'8228 \Longrightarrow 0'8228 \simeq 2 \cdot (1-\rho) \Longrightarrow \rho \simeq 1 - \frac{0'8228}{2} = 0'5886.$$

Teniendo en cuenta este valor se obtienen los valores transformados de la tabla 9, a partir de los cuales se obtiene el estimador por MCG sin más que aplicarle MCO:

$$\begin{aligned}
\widehat{\beta} &= \left(X^t X\right)^{-1} X^t y = \begin{pmatrix} 0'00001 \cdot 10^5 & 0'0143 \cdot 10^5 \\ 0'0143 \cdot 10^5 & 9'9132 \cdot 10^5 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 0'0075 \cdot 10^5 \\ 5'2107 \cdot 10^5 \end{pmatrix} \\
&= \begin{pmatrix} 8'5305 & -0'0123 \\ -0'0123 & 0'00001 \end{pmatrix} \cdot \begin{pmatrix} 0'0075 \cdot 10^5 \\ 5'2107 \cdot 10^5 \end{pmatrix} = \begin{pmatrix} 2'0068 \\ 0'0497 \end{pmatrix}.
\end{aligned}$$

Esto es, $\widehat{Y}_t = 2'0068 + 0'0497 \cdot X_t$.

Además, a partir de la información de la tabla 10 se tiene que:

$$d = \frac{30'3644}{16'5288} = 1'8371.$$

Y como $1'3197 = d_U < d < 2'6803 = 4 - d_U$ la perturbación aleatoria del modelo es incorrelada.

7. **Se ha recogido información de la economía española para el periodo 1998-2006 del consumo público y del PIB con el objeto de estimar un modelo de regresión lineal que explique**

el consumo público. A través de la estimación de MCO se han obtenido los siguientes resultados:

$$R^2 = 0'85, \quad SCT = 5'08, \quad \sum_{t=2}^{9} e_t \cdot e_{t-1} = 0'22.$$

Detectar la posible presencia de autocorrelación a través del contraste de Durbin-Watson.

Teniendo en cuenta que

$$R^2 = 1 - \frac{SCR}{SCT} \Longrightarrow SCR = \left(1 - R^2\right) \cdot SCT = (1 - 0'85) \cdot 5'08 = 0'762,$$

es claro que

$$\rho = \frac{\sum\limits_{t=2}^{9} e_t \cdot e_{t-1}}{\sum\limits_{t=1}^{9} e_t^2} = \frac{0'22}{0'762} = 0'28872,$$

y entonces

$$d \simeq 2 \cdot (1 - 0'28872) = 1'42257218.$$

Como $d_L = 0'8243$ y $d_U = 1'31$, se verifica que $1'31 = d_U < d = 1'42 < 2'69 = 4 - d_U$. Esto es, el modelo es incorrelado.

8. **Al estimar por MCO un modelo lineal, a partir de 21 observaciones, se obtuvo:**

$$\underset{}{\widehat{Y}_t} = \underset{(0'3)}{1'3} + \underset{(0'18)}{0'97} \cdot Y_{t-1} + \underset{(0'41)}{2'31} \cdot X_t, \quad d = 1'21,$$

donde las cifras entre paréntesis son las desviaciones típicas. Contrastar la presencia de autocorrelación en la perturbación aleatoria.

Puesto que como regresora aparece la variable dependiente retardada para estudiar la autocorrelación en este modelo hay que utilizar la h de Durbin. En tal caso, se rechaza la hipótesis nula de incorrelación si

$$|h| = \left| \rho \cdot \sqrt{\frac{n}{1 - n \cdot var}} \right| > Z_{1-\frac{\alpha}{2}},$$

donde var es la varianza estimada del coeficiente correspondiente a la variable retardada y $Z_{1-\frac{\alpha}{2}}$ es el punto de una distribución N(0,1) que deja a su izquierda una probabilidad $1 - \frac{\alpha}{2}$.

Es evidente que $n = 21$ y $var = 0'18^2 = 0'0324$. Por otro lado, como $d = 1'21$ se tiene que $\rho \simeq 1 - \frac{1'21}{2} = 0'395$.

Luego, sin más que sustituir:

$$|h| = \left| 0'395 \cdot \sqrt{\frac{21}{1 - 21 \cdot 0'0324}} \right| = 3'201 > 1'96 = Z_{0'975}.$$

Por tanto, rechazo la hipótesis nula de incorrelación, es decir, hay autocorrelación en la perturbación aleatoria.

9. **Dado un modelo lineal de consumo en función del PIB con los siguientes datos:**

Y_t	**22**	**15**	**8**	**6**	**3**	**2**	**7**
X_t	**3**	**1**	**2**	**0**	**-2**	**-3**	**-1**

Contraste la existencia de autocorrelación sabiendo que la regresión del modelo original por MCO produce los siguientes residuos:

Cuadro 11: Cálculo del estadístico de Durbin-Watson

e_t	e_t^2	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
4'63	21'4369			
3'21	10'3041	4'63	-1'42	2'0164
-6'58	43'2964	3'21	9'79	95'8441
-3	9	-6'58	3'58	12'8164
-0'42	0'1764	-3	2'58	6'6564
1'37	1'8769	-0'42	1'79	3'2041
0'79	0'6241	1'37	-0'58	0'3364
	87'2148			120'8738

e_t	**4'63**	**3'21**	**-6'58**	**-3**	**-0'42**	**1'37**	**0'79**

Teniendo en cuenta la información de la tabla 11 se tiene que

$$d = \frac{120'8738}{87'2148} = 1'3859.$$

Como $d_L = 0'6996$ y $d_U = 1'3564$, se tiene que $d_U < d < 4 - d_U = 2'6436$. Luego, la perturbación aleatoria del modelo considerado está incorrelada.

10. **A partir de una muestra de 20 datos se ha estimado por MCO el siguiente modelo:**

$$\widehat{Y}_t = 4'9 + 2'2X_{2t} + 3'5X_{3t},$$

mientras que con los residuos del modelo anterior se ha realizado la siguiente regresión:

$$\widehat{e}_t = 0'75 \cdot e_{t-1}.$$

Se pide:

a) **Analizar la presencia de autocorrelación de primer orden utilizando el contraste de Durbin-Watson.**

b) **Suponiendo que las perturbaciones siguen un proceso autorregresivo de primer orden y que ha obtenido una estimación adecuada del coeficiente de dicho proceso, especifique la ecuación que usaría para obtener estimaciones eficientes de los parámetros del modelo.**

Puesto que $\rho = 0'75$, entonces $d \simeq 2 \cdot (1 - \rho) = 0'5$. Y como $d_L = 1'1004$ y $d_U = 1'5367$ hay autocorrelación positiva en la perturbación aleatoria del modelo ya que $d < d_L$.

Por otro lado, puesto que se disponen de pocas observaciones, se debería usar el procedimiento iterativo de Prais-Winsten con $\rho = 0'75$ para la obtención de estimaciones óptimas.

11. **Utilizando una muestra de 25 observaciones anuales se estima mediante MCO el siguiente modelo que estudia la demanda (D) en función del precio (P) y la renta (R):**

$$\underset{(322'08)}{\widehat{D}_t = 521'2} + \underset{(0'036)}{0'532 \cdot R_t} - \underset{(18'75)}{23'25 \cdot P_t} + \underset{(0'05)}{0'415 \cdot D_{t-1}}, \qquad d = 2'088.$$

¿Se puede decir que los estimadores por MCO son óptimos?

Puesto que como regresora aparece la variable dependiente retardada para estudiar la autocorrelación en este modelo hay que utilizar la h de Durbin. En tal caso, se rechaza la hipótesis nula de incorrelación si

$$|h| = \left|\rho \cdot \sqrt{\frac{n}{1 - n \cdot var}}\right| > Z_{1-\frac{\alpha}{2}},$$

donde var es la varianza estimada del coeficiente correspondiente a la variable retardada y $Z_{1-\frac{\alpha}{2}}$ es el punto de una distribución N(0,1) que deja a su izquierda una probabilidad $1-\frac{\alpha}{2}$.

Es evidente que $n = 25$ y $var = 0'05^2 = 0'0025$. Por otro lado, como $d = 2'088$ se tiene que $\rho \simeq 1 - \frac{2'088}{2} = -0'044$.

Luego, sin más que sustituir:

$$|h| = \left| -0'044 \cdot \sqrt{\frac{25}{1 - 25 \cdot 0'0025}} \right| = 0'2273 \not> 1'96 = Z_{0'975}.$$

En tal caso, no rechazo la hipótesis nula de incorrelación, y por tanto, los estimadores obtenidos por MCO son óptimos.

12. **Se ha estimado por MCO la formación bruta del capital, FBC, en función del producto interior bruto, PIB, para los años 1985 a 1997, obteniéndose que:**

$$\begin{array}{ccccccc} \widehat{FBC}_t & = & \mathbf{5144'144} & + & \mathbf{0'070929} \cdot PIB_t & & \rho = 0'7846 \\ & & \mathbf{(1039'264)} & & \mathbf{(0'018711)} & & \end{array}$$

¿Son óptimos los estimadores obtenidos?

Como $d \simeq 2 \cdot (1-\rho) = 2 \cdot (1 - 0'7846) = 2 \cdot 0'2154 = 0'4308$ y para $n = 13$ y $k' = 1$ se tiene al 5 % que $d_L = 1'01$ y $d_U = 1'34$, es evidente que hay autocorrelación positiva en el modelo ya que $d < d_L$ y en consecuencia los estimadores obtenidos por MCO no serán óptimos.

13. **A partir de una muestra de 95 observaciones se ha obtenido la siguiente estimación:**

$$\begin{array}{ccccccccccc} \widehat{Y}_t & = & \mathbf{8'34} & + & \mathbf{0'7} \cdot X_{2t} & \mathbf{-} & \mathbf{0'4} \cdot X_{3t} & + & \mathbf{0'1} \cdot X_{4t} & & d = 2'86 \\ & & \mathbf{(7'1)} & & \mathbf{(0'56)} & & \mathbf{(0'7)} & & \mathbf{(0'05)} & & \end{array}$$

Analice la posible presencia de autocorrelación en la perturbación aleatoria del modelo anterior y, en caso afirmativo, describa detalladamente la forma más adecuada de solucionarlo.

Para $n = 95$ y $k' = 3$ se tiene que $d_L = 1'602$ y $d_U = 1'732$, luego como $d > 4 - d_L = 2'398$ es evidente que hay autocorrelación negativa en el modelo.

Puesto que se dispone de un número alto de observaciones el método idóneao para transformar el modelo y eliminar así el problema de autocorrelación es el de Cochrane-Orcutt. En dicho proceso se transforman los datos según:

$$Y_t^* = Y_t - \rho \cdot Y_{t-1}, \qquad X_{it}^* = X_{it} - \rho \cdot X_{i\ t-1}, \quad t > 1, \quad i = 1, 2, 3, 4,$$

donde $\rho = 1 - \frac{d}{2} = 1 - \frac{2'86}{2} = 1 - 1'43 = -0'43$.

14. **¿Existe autocorrelación en el siguiente modelo estimado por MCO?**

$$\begin{array}{ccccccccc} \widehat{Y}_t & = & \mathbf{1'3} & + & \mathbf{0'97} \cdot Y_{t-1} & + & \mathbf{2'31} \cdot X_t & n = 21 & d = 1'21 \\ & & \mathbf{(0'8)} & & \mathbf{(0'07)} & & \mathbf{(1'2)} & & \end{array}$$

A partir de la h de Durbin:

$$h = 0'394 \cdot \sqrt{\frac{21}{1 - 21 \cdot 0'0049}} = 0'395 \cdot 4'83826 = 1'911113,$$

donde $\rho \simeq 1 - \frac{d}{2} = 1 - \frac{1'21}{2} = 0'395$ y $var = 0'07^2 = 0'0049$, se tiene que no se rechaza la hipótesis nula de incorrelación ya que $|h| \not> 1'96$. Luego no existe autocorrelación en el modelo.