

HISPALINK-ASTURIAS

Documentos de trabajo

**ECONOMETRÍA APLICADA Y PREDICCIÓN.
SUPUESTOS RESUELTOS CON *ECONOMETRIC VIEWS***

Ana Jesús López Menéndez
Rigoberto Pérez Suárez

DOCUMENTO DE TRABAJO 1/99 (Enero 1999)

Han participado en la elaboración de este documento de trabajo:

Ana Jesús López Menéndez
Rigoberto Pérez Suárez

D. Legal: AS/359-99

Presentación

Las técnicas econométricas constituyen herramientas cada día más imprescindibles para la investigación en Economía Aplicada. Esta afirmación, que en la actualidad suele ser asumida con generalidad, resulta sorprendente si tenemos en cuenta la juventud de la Econometría, disciplina que surgió como tal hace menos de 70 años.

Nuestra experiencia docente en Econometría, tanto en materias de licenciatura como de tercer ciclo (doctorado y postgrado) nos ha llevado a dos conclusiones fundamentales. En primer lugar, consideramos que el estudio de la que Econometría Teórica queda incompleto si el alumno no se adentra en la práctica de esta disciplina, que aumenta su motivación descubriéndole tanto las posibilidades como los problemas inherentes a los trabajos aplicados.

Por otra parte, nos parece innegable que la difusión de las técnicas econométricas se ha visto favorecida por el desarrollo de aplicaciones informáticas y, en especial, aquéllas -como es el caso de *Econometric Views*- que han hecho un esfuerzo por resultar accesibles al usuario.

Nuestro convencimiento sobre ambos aspectos nos ha llevado a incluir la realización de prácticas con soporte informático y la elaboración de trabajos aplicados como parte fundamental del aprendizaje de la Econometría.

Este documento de trabajo contiene un total de 10 supuestos en los que hemos pretendido combinar planteamientos y técnicas diversas, que incluyen desde el modelo lineal básico hasta algunos desarrollos recientes. A continuación recogemos un índice temático de estas aplicaciones, en el que se indica también el nombre del correspondiente archivo de trabajo en Eviews. Todos estos archivos (con extensión WF1) junto con algunos programas e instrucciones aparecen recogidos en el disquette adjunto.

INDICE DE CONTENIDOS

Supuesto	Archivo	Contenidos	Página
1	VENTAS.WF1	Modelo lineal, Estimación Mínimo Cuadrática, Autocorrelación, Heteroscedasticidad, Mínimos Cuadrados Ponderados, Normalidad, Cambio estructural.	2
2	CULTURA.WF1	Estimación, Contraste de restricciones, AR(1), Predicción.	12
3	GASTOS.WF1	Regresión con variables cualitativas. Estimación y predicción de modelos <i>logit</i> .	18
4	EMPRESAS.WF1	Modelo SUR. Estimación. Contrastes de ausencia de correlaciones y de igualdad de coeficientes	22
5	INDUSTRIA.WF1	Modelos de ecuaciones simultáneas. Identificación, Métodos de Estimación. Causalidad de Granger.	28
6	CARNE.WF1	Análisis del esquema de composición. Estacionalidad. Predicción basada en modelos de alisado. Modelización ARIMA	35
7	SIMULA.WF1	Simulaciones de series a partir de distintos procesos generadores de datos. Identificación y estimación de modelos.	42
8	REVISTA.WF1	Series temporales. Modelos ARIMA. Análisis de intervención. Predicción.	50
9	COOPERA.WF1	Datos de panel. Estimación de modelos con efectos fijos y aleatorios.	54
10	CONSUMO.WF1	Introducción a la cointegración. Análisis de estacionariedad. Test de Johansen, Estimación de Modelos de Corrección de Error.	59
Referencias Bibliográficas			64
ANEXOS			66

SUPUESTO 1

- ⇒ Unos grandes almacenes han recopilado durante la campaña navideña información relativa al volumen de ventas de 50 CD musicales (variable VENTAS, expresada en unidades), su precio (PRECIO, en euros por unidad) y el esfuerzo publicitario (PUBLICIDAD, obtenida como promedio de minutos de publicidad semanal en cadenas de radio y televisión).

La información aparece recogida en el archivo **VENTAS.WF1**:

Obs	Ventas	Precio	Publicidad	Obs	Ventas	Precio	Publicidad
1	1231	14.19	12.4	26	1420	16.88	16.8
2	1243	15.63	9.9	27	1242	15.44	8.8
3	893	12.63	2.4	28	984	15.50	3.2
4	1413	12.69	13.8	29	1148	14.00	5.4
5	1128	13.13	3.5	30	1425	11.56	17.3
6	1081	11.88	1.8	31	1226	14.25	11.2
7	1439	11.81	17.8	32	1277	16.38	11.2
8	1242	15.31	9.8	33	1130	12.56	7.9
9	1101	17.44	8.3	34	1442	13.00	17
10	1117	17.63	9.8	35	1092	12.13	3.3
11	1238	13.81	12.6	36	1068	16.50	7.1
12	1235	14.25	11.5	37	1400	13.81	15.3
13	1102	17.63	7.4	38	1240	14.13	12.7
14	1009	15.38	6.1	39	1067	16.81	6.1
15	1233	15.31	9.5	40	1532	15.50	19.6
16	1157	13.00	8.8	41	1201	15.00	6.3
17	1166	13.81	4.9	42	1193	14.00	9
18	1535	15.88	18.8	43	1506	15.44	18.7
19	1492	14.06	18.9	44	922	13.88	2.2
20	890	12.63	2.3	45	1450	15.25	16
21	1326	17.38	14.1	46	1125	13.19	4.5
22	975	15.50	2.9	47	1118	13.25	4.3
23	1061	16.75	5.9	48	1201	14.31	9.3
24	1153	13.13	7.6	49	1074	17.00	8.3
25	985	15.00	5.3	50	1370	15.31	15.4

- a) Estimar un modelo econométrico para las ventas de CD durante la campaña navideña.
- b) Contrastar las hipótesis habituales sobre las perturbaciones, proponiendo las modificaciones oportunas.
- c) Durante el mes de Enero han sido lanzados al mercado 12 nuevos CD, para los que se ha registrado la siguiente información:

Obs	Ventas	Precios	Publicidad	Obs	Ventas	Precios	Publicidad
51	1460	16.25	9.2	57	689	17.19	17.5
52	772	16.88	10.2	58	805	17.50	18.3
53	799	15.00	16	59	785	18.31	18.2
54	515	15.00	4	60	883	18.75	22.1
55	863	18.13	19.6	61	536	13.13	4.1
56	1200	17.50	10.2	62	1018	15.63	24.9

Estudiar si estas ventas siguen pautas comunes a las de la campaña navideña.

a) Partiendo de la propuesta de un modelo lineal para la variable VENTAS en función de PRECIO y PUBLICIDAD se obtiene la estimación mínimo cuadrática:

Dependent Variable: VENTAS

Method: Least Squares

Sample: 1 50

Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1064.253	72.61065	14.65698	0.0000
PRECIO	-10.82454	4.934838	-2.193494	0.0333
PUBLICIDAD	31.05061	1.564819	19.84294	0.0000
R-squared	0.893367	Mean dependent var	1205.960	
Adjusted R-squared	0.888829	S.D. dependent var	170.5741	
S.E. of regression	56.87330	Akaike info criterion	10.97765	
Sum squared resid	152024.9	Schwarz criterion	11.09237	
Log likelihood	-271.4413	F-statistic	196.8815	
Durbin-Watson stat	2.048885	Prob(F-statistic)	0.000000	

donde se observan los signos esperados para los coeficientes de las variables PRECIO y PUBLICIDAD y se obtienen resultados significativos en los contrastes individuales de significación (t de Student) y también en el test global F de Snedecor.

Relacionado con este último contraste se encuentra el coeficiente de determinación, R² cuyo valor supera el 89%, permitiendo calificar como aceptable la capacidad explicativa del modelo (un 89.3% de las variaciones en las compras navideñas de CD quedan explicadas por su precio y los gastos publicitarios).

Las expresiones de cálculo de estos resultados recogidos en el output de regresión mínimo cuadrática del programa Eviews son las siguientes:

R-squared (Coeficiente de determinación)	$R^2 = 1 - \frac{\hat{u}'\hat{u}}{y'y - n\bar{Y}^2}$	Mean dependent var (Valor medio de la variable dependiente)	$\bar{Y} = \frac{\sum Y_t}{n}$
Adjusted R-squared (Coeficiente de determinación ajustado)	$\bar{R}^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-k)}$	S.D. dependent var (Desviación estándar de la variable dependiente)	$S_Y = \sqrt{\frac{\sum (Y_t - \bar{Y})^2}{n-1}}$
S.E. of regression (Error estándar de la regresión)	$S = \sqrt{\frac{\hat{u}'\hat{u}}{n-k}}$	Akaike info criterion (Criterio de información de Akaike)	$AIC = \frac{2k}{n} + \text{Ln}\left(\frac{\hat{u}'\hat{u}}{n}\right)$
Sum squared resid (Suma de residuos cuadráticos)	$\hat{u}'\hat{u}$	Schwarz criterion (Criterio de Schwarz)	$SIC = \frac{k \text{Ln}(n)}{n} + \text{Ln}\left(\frac{\hat{u}'\hat{u}}{n}\right)$
Log likelihood (Logaritmo de máxima verosimilitud)	$\text{LnL} = \frac{-n}{2} (\text{Ln}(2\pi\hat{\sigma}^2) + 1)$	F-statistic	$F_{n-k}^{k-1} = \frac{R^2/k-1}{1-R^2/n-k}$
Durbin-Watson stat (Estadístico Durbin-Watson)	$d_{DW} = \frac{\sum_{t=2} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1} \hat{u}_t^2}$	Prob(F-statistic)	$P(F_{n-k}^{k-1} > F^*)$

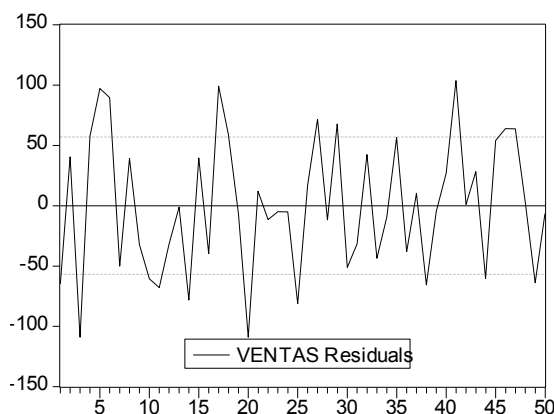
Además es posible recuperar algunos de los resultados de la regresión mediante los comandos correspondientes¹ y almacenarlos si se desea en un vector o en una tabla. Así, la discrepancia t para la variable PRECIO, que ocupa la segunda posición, se obtendría como $C(10)=C(2)/\text{SQR}@COVARIANCE(3,3)$ y el correspondiente nivel crítico como $C(11)=@TDIST(C(10), @REGOBS-@NCOEF)$

Si deseamos "reconstruir" todos los resultados de la regresión mínimo cuadrática podríamos convertir las series consideradas en matrices y vectores, utilizando operadores del programa *Eviews* tal y como se recoge en el Anexo A1.

b) Las hipótesis habituales sobre las perturbaciones aleatorias u son: esperanza nula, matriz de varianzas-covarianzas escalar (que engloba los supuestos de homoscedasticidad y no autocorrelación) y normalidad.

Dado que la nulidad de la esperanza no puede ser contrastada analizamos las hipótesis sobre la matriz de varianzas y covarianzas.

Una primera aproximación es la proporcionada por la gráfica de los residuos de la regresión, que en este caso no parece detectar síntomas de correlación serial:



¹ Estos comandos se encuentran en Ayudas (*Functions for Regression Statistics*) y los más habituales son:

@R2	R ²	@RBAR2	R ² Ajustado	@SE	Error estándar de regresión
@SSR	Suma de residuos cuadráticos	@DW	Estadístico Durbin-Watson	@F	Estadístico F
@LOGL	Logaritmo de verosimilitud	@MEANDEP	Media de la variable dependiente	@REGOBS	Número de observaciones de la regresión
@SDDEP	Desviación estándar de la variable dependiente	@NCOEF	Número de coeficientes de la regresión	@COVARIANCE(i,j)	Covarianza entre los coeficientes i y j

El estadístico habitual para contrastar la *ausencia de correlación serial de primer orden* es

el propuesto por Durbin y Watson, que viene dado por la expresión $d_{DW} = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}$.

Con nuestra información muestral (tal y como cabía esperar, al tratarse de datos de corte transversal) no se presentan problemas de autocorrelación ya que el estadístico Durbin-watson recogido en la salida de la regresión proporciona un resultado cercano a 2, situándose por tanto en la zona de no correlación².

Teniendo en cuenta que podrían existir correlaciones de orden superior a la unidad, convendría aplicar algún test más global, como los propuestos por Ljung y Box (1979) y Breusch y Godfrey (1978).

Ambos contrastes aparecen implementados en el programa Eviews y proporcionan idénticas conclusiones. La hipótesis nula es la inexistencia de correlaciones hasta un orden prefijado y el test Q de Ljung-Box, que se muestra junto al correlograma de los residuos, proporciona niveles críticos elevados.

Sample: 1 50
Included observations: 50

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.038	-0.038	0.0781	0.780
		2 0.016	0.015	0.0921	0.955
		3 -0.223	-0.222	2.8381	0.417
		4 -0.082	-0.104	3.2200	0.522
		5 -0.043	-0.050	3.3283	0.650
		6 0.084	0.033	3.7494	0.711
		7 -0.229	-0.281	6.9103	0.438
		8 -0.067	-0.144	7.1885	0.516
		9 0.001	0.001	7.1886	0.617
		10 0.014	-0.111	7.2005	0.706
		11 0.222	0.127	10.493	0.487
		12 0.028	-0.002	10.546	0.568
		13 0.180	0.223	12.834	0.461
		14 -0.013	0.052	12.845	0.539
		15 -0.166	-0.188	14.893	0.459
		16 -0.105	-0.008	15.735	0.472

La misma conclusión se obtiene aplicando el test Breusch-Godfrey, que se denomina LM por basarse en los multiplicadores de Lagrange. A modo de ilustración, proponiendo órdenes de correlación hasta 10, se obtienen niveles críticos superiores al 50%.

Breusch-Godfrey Serial Correlation LM Test:

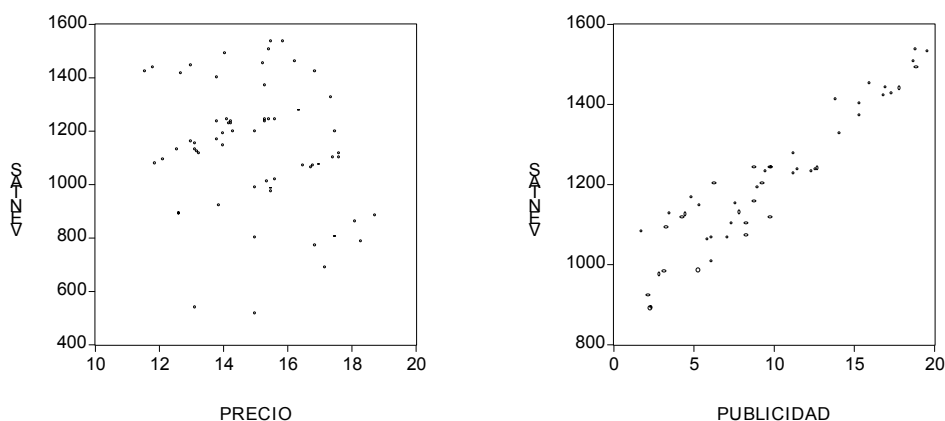
F-statistic	0.800421	Probability	0.629088
Obs*R-squared	8.892737	Probability	0.542313

² Para el tamaño muestral n=50 y dos regresores, los valores tabulados en las distribuciones auxiliares de Durbin-Watson proporcionan para un nivel de significación del 5% los resultados $d_L=1.462$ y $d_U=1.628$. Por tanto la zona de no autocorrelación $[d_U, 4-d_U]$ sería en este caso $[1.628, 2.372]$

Para contrastar ahora la hipótesis de *homoscedasticidad* de las perturbaciones podemos plantear -además del análisis gráfico- varios contrastes, siendo los más habituales los propuestos por Goldfeld y Quandt (1965) y White (1980)

El primero de ellos no aparece implementado en Eviews pero puede sin embargo ser construido con relativa sencillez. Conviene tener presente que el test de Goldfeld y Quandt propone, frente a la hipótesis de varianzas constantes de las perturbaciones, la alternativa de varianzas proporcionales a alguna de las variables explicativas, por lo cual resulta interesante estudiar qué variable puede aparecer asociada a la heteroscedasticidad.

El examen las nubes de puntos (SCAT Diagram) de VENTAS junto con las variables PRECIO y PUBLICIDAD permite apreciar una menor dispersión en ventas a medida que aumentan los gastos publicitarios.



Con esta información parece adecuado asumir para la dispersión una hipótesis

alternativa del tipo: $\sigma_i^2 = \frac{\sigma^2}{X_i}$ siendo X la variable publicidad. Así, pues, el test de Goldfeld

y Quandt -implementado en el programa *GQ.prg* recogido en el disquette adjunto y descrito en el Anexo A2- ordena las observaciones en orden creciente según la variable publicidad y compara los residuos cuadráticos asociados a las regresiones de dos submuestras, concluyendo que éstos son significativamente distintos³.

Por su parte, el test de White aparece implementado directamente en el programa Eviews y se basa en la realización de regresiones sobre los residuos cuadráticos asociados al modelo propuesto para las ventas.

³ Más concretamente, considerando dos submuestras de tamaño 28 y 29 respectivamente se obtienen sumas de residuos cuadráticos $\hat{u}'_1 \hat{u}_1=583626.68$ y $\hat{u}'_2 \hat{u}_2=1299119.382$ que conducen a un valor de la discrepancia F de Snedecor $F=2.225$, que lleva asociado un nivel crítico del 2.4%.

White Heteroskedasticity Test:

F-statistic	5.377019	Probability	0.000608
Obs*R-squared	18.96386	Probability	0.001952

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Sample: 1 50

Included observations: 50

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	26074.46	31609.73	0.824887	0.4139
PRECIO	-1365.173	4219.908	-0.323508	0.7478
PRECIO^2	1.127730	139.3722	0.008092	0.9936
PRECIO*PUBLICIDAD	80.79962	49.71005	1.625418	0.1112
PUBLICIDAD	-1694.023	677.0832	-2.501942	0.0161
PUBLICIDAD^2	13.01835	17.27164	0.753741	0.4550
R-squared	0.379277	Mean dependent var		3040.498
Adjusted R-squared	0.308740	S.D. dependent var		3275.128
S.E. of regression	2723.007	Akaike info criterion		18.76903
Sum squared resid	3.26E+08	Schwarz criterion		18.99847
Log likelihood	-463.2257	F-statistic		5.377019
Durbin-Watson stat	2.314205	Prob(F-statistic)		0.000608

Observando los bajos niveles críticos asociados a los estadísticos F y chi-cuadrado (inferiores al 1%) se concluye que el resultado es significativo para rechazar el supuesto de dispersión constante en las perturbaciones. Como consecuencia de la presencia de heteroscedasticidad, los estimadores mínimo cuadráticos no serían eficientes y los contrastes de significación obtenidos en la estimación planteada para las ventas no serían válidos.

Si a pesar de ello decidiéramos utilizar los estimadores Mínimo Cuadráticos (MC), deberíamos corregir su matriz de varianzas-covarianzas, teniendo en cuenta que la dispersión no es constante. El programa Eviews permite incorporar las correcciones de heteroscedasticidad propuestas por White (1980) y por Newey-West (1987), que proporcionan los mismos estimadores para los parámetros pero en cambio alteran sus varianzas y en consecuencia los contrastes de significación basados en la t de Student.

A modo de resumen, presentamos una tabla comparativa, que recoge las desviaciones estándar y los valores de la t asociados a la estimación inicial y a las correcciones de heteroscedasticidad contempladas en el programa.⁴

⁴ La corrección de White, proporciona estimaciones consistentes para la matriz de varianzas-covarianzas de los estimadores MC en presencia de heteroscedasticidad de estructura desconocida, y asumiendo perturbaciones incorreladas. Por su parte, Newey y West proponen una corrección más general, que

Variable	Coefficient	Std Error. MC	Std Error. White	Std Error. Newey-West	t Student. MC	t Student. White	t Student Newey-West
C	1064.253	72.61065	77.61859	72.94224	14.65698	13.71131	14.59035
PRECIO	-10.82454	4.934838	4.645503	4.334002	-2.193494	-2.330111	-2.497585
PUBLICIDAD	31.05061	1.564819	1.572735	1.542482	19.84294	19.74306	20.13029

Como puede apreciarse, estas correcciones no cambian las conclusiones respecto a la significación de las variables, obteniéndose valores t mayores en valor absoluto (y por consiguiente niveles críticos inferiores) para el coeficiente de PRECIO.

Las correcciones anteriormente comentadas solucionan el problema de la consistencia en la estimación de la matriz de varianzas-covarianzas de los estimadores. No obstante, una vez detectado el problema de la heteroscedasticidad y su posible estructura sería conveniente llevar a cabo una estimación del modelo por el método de Mínimos Cuadrados Ponderados (MCP).

Como ya hemos visto anteriormente, la dispersión parece encontrarse inversamente relacionada con el esfuerzo publicitario: $\sigma_i^2 = \frac{\sigma^2}{X_i}$, supuesto que viene apoyado por la realización de regresiones auxiliares sobre los residuos cuadráticos. Así, el coeficiente de determinación para la regresión de residuos cuadráticos frente a la inversa de la publicidad $\left(\frac{1}{X_i}\right)$ se acerca al 30%, superando a los R² asociados a otras alternativas $\left(\frac{1}{X_i^2}, X_i^2, X_i, \dots\right)$.

Con esta información parece adecuado transformar el modelo inicial mediante los pesos $w_i = \sqrt{X_i}$, que se generan mediante el comando GENR [PESO=@SQR\(PUBLICIDAD\)](#). La estimación por Mínimos Cuadrados Ponderados proporciona entonces:

Dependent Variable: VENTAS		Method: Least Squares		
Sample: 1 50		Included observations: 50		
Weighting series: PESO				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	978.6479	65.48158	14.94539	0.0000
PRECIO	-5.521181	4.179508	-1.321012	0.1929
PUBLICIDAD	31.66199	1.387717	22.81588	0.0000
Weighted Statistics				
R-squared	0.991358	Mean dependent var	1251.221	
Adjusted R-squared	0.990990	S.D. dependent var	524.9074	
S.E. of regression	49.82432	Akaike info criterion	10.71301	
Sum squared resid	116675.8	Schwarz criterion	10.82773	
Log likelihood	-264.8252	F-statistic	265.1156	
Durbin-Watson stat	1.835456	Prob(F-statistic)	0.000000	
Unweighted Statistics				
R-squared	0.890004	Mean dependent var	1205.960	
Adjusted R-squared	0.885323	S.D. dependent var	170.5741	
S.E. of regression	57.76310	Sum squared resid	156819.0	
Durbin-Watson stat	1.842205			

estima insesgadamente la matriz de varianzas-covarianzas en presencia de heteroscedasticidad y autocorrelación.

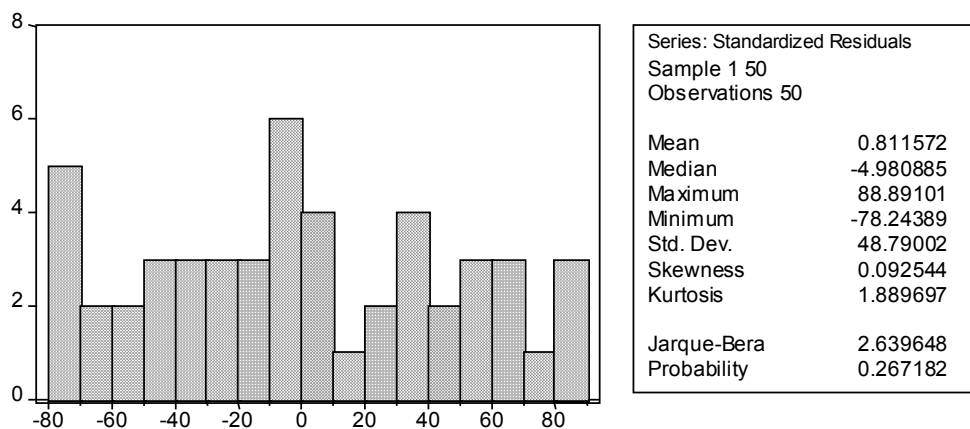
Como vemos, la salida del programa proporciona la estimación del modelo transformado (en la que los signos se mantienen pero la variable PRECIO pierde significación) junto a la información relativa a los residuos de la estimación MCP (*unweighted statistics*, basados en $\hat{u}_{MCP} = Y - X\hat{\beta}_{MCP}$) y también a los residuos del modelo ponderado (*weighted statistics*, a partir de las series $\hat{u}_{MCP}^* = Y^* - X^*\hat{\beta}_{MCP} = wY - wX\hat{\beta}_{MCP}$, teniendo en cuenta que los pesos w han sido normalizados respecto a su media)⁵.

Si sobre el modelo transformado efectuamos de nuevo el test de White se obtiene un nivel crítico elevado, que confirma que hemos solucionado el problema de la heteroscedasticidad.

White Heteroskedasticity Test:

F-statistic	0.417953	Probability	0.833711
Obs*R-squared	2.267057	Probability	0.811093

Por último, debemos efectuar el contraste de la **hipótesis de normalidad** para las perturbaciones. El programa proporciona el test de Jarque y Bera (1980), basado en la asimetría y el apuntamiento sobre el histograma de los residuos⁶, que en este caso proporciona un resultado no significativo para rechazar (nivel crítico del 26%):



⁵ Para llevar a cabo la estimación por MCP, el programa Eviews normaliza la serie de pesos dividiendo entre su media, transformación que no afecta a las estimaciones de los parámetros y proporciona residuos ponderados más comparables con los no ponderados. Estos últimos son los almacenados como residuos de la estimación (serie RESID), ya que los residuos ponderados sólo pueden ser utilizados sin problemas en situaciones donde la serie de pesos no tiene escala.

⁶ Denotando por g_1 y g_2 los coeficientes muestrales de asimetría y apuntamiento de Fisher respectivamente, la discrepancia de Jarque y Bera viene dada por la expresión $d_{JB} = \frac{n}{6} \left(g_1^2 + \frac{1}{4} g_2^2 \right)$ que bajo el supuesto de normalidad sigue distribución chi-cuadrado con 2 grados de libertad.

c) Para analizar ahora si las nuevas observaciones se adaptan al mismo esquema de comportamiento estimado para las ventas navideñas podríamos comenzar por estimar el modelo incluyendo las nuevas observaciones:

Dependent Variable: VENTAS

Method: Least Squares

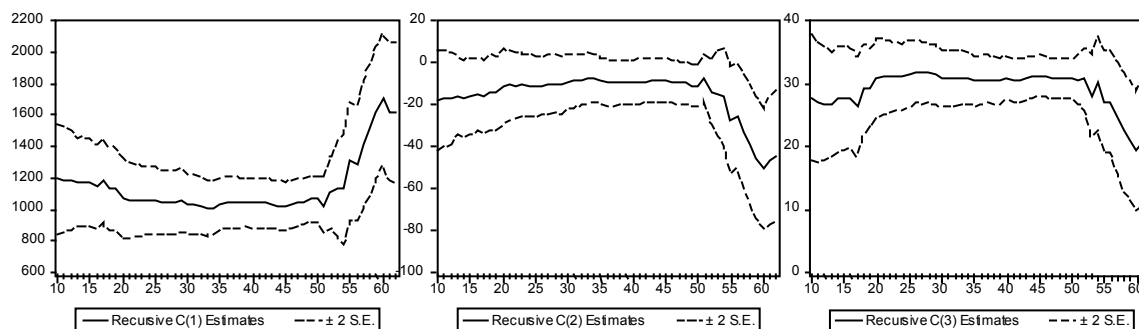
Sample: 1 62

Included observations: 62

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1619.138	207.0531	7.819918	0.0000
PRECIO	-44.77451	13.42741	-3.334561	0.0015
PUBLICIDAD	18.08329	5.297036	3.413850	0.0012
R-squared	0.222256	Mean dependent var		1139.081
Adjusted R-squared	0.195891	S.D. dependent var		234.5425
S.E. of regression	210.3192	Akaike info criterion		13.58231
Sum squared resid	2609817.	Schwarz criterion		13.68523
Log likelihood	-418.0515	F-statistic		8.430199
Durbin-Watson stat	1.116943	Prob(F-statistic)		0.000602

apreciándose algunos cambios en los coeficientes respecto a la estimación inicial, que aconsejan examinar su estabilidad⁷. Una posibilidad es efectuar el test de estimaciones recursivas de coeficientes, que proporciona los resultados siguientes⁸:



Dado que estas representaciones gráficas apoyan nuestras sospechas de cambios de estructura en el modelo tras la campaña navideña, conviene llevar a cabo el test de estabilidad de Chow (1960).

Así, proponiendo como dato de ruptura el 51, el programa proporciona la salida:

Chow Breakpoint Test: 51

F-statistic	38.30705	Probability	0.000000
Log likelihood ratio	69.18275	Probability	0.000000

⁷ Hemos efectuado la regresión por MCO con la corrección de heteroscedasticidad de White. No obstante, los resultados serían similares si partiésemos del modelo estimado por MCP.

⁸ Conclusiones similares se obtendrían mediante otros análisis gráficos de estabilidad (residuos recursivos, sumas acumulativas o CUSUM, etc).

resultado al que se llega efectuando dos regresiones separadas y comparando los residuos cuadráticos asociados a ambas con los del modelo global, tal y como recoge la tabla adjunta:

Muestra	Suma de residuos cuadráticos
Observaciones 1-50	$\hat{u}'_1 \hat{u}_1 = 152024.9$
Observaciones 51-62	$\hat{u}'_2 \hat{u}_2 = 703046.2$
Observaciones 1-62	$\hat{u}' \hat{u} = 2609817$
Discrepancia	
$\frac{\hat{u}' \hat{u} - (\hat{u}'_1 \hat{u}_1 + \hat{u}'_2 \hat{u}_2)}{\frac{k}{n_1 + n_2 - 2k}} \approx F_{n_1 + n_2 - 2k}^k$	F=38.31

A la vista de este resultado, con un nivel crítico prácticamente nulo, la conclusión es que debemos rechazar la existencia de un modelo único de ventas para todo el recorrido muestral.

Por tanto, sería recomendable introducir en el modelo una variable cualitativa asociada al cambio estructural, que podría afectar al término independiente y/o a los efectos marginales de las variables PRECIO y PUBLICIDAD:

Dependent Variable: VENTAS

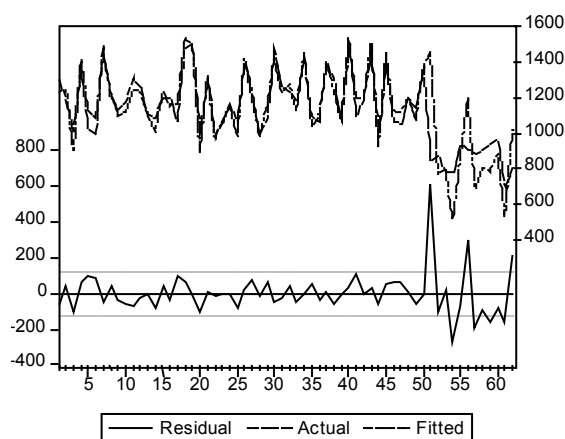
Method: Least Squares

Sample(adjusted): 1 62

Included observations: 62 after adjusting endpoints

White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1064.253	79.18292	13.44043	0.0000
PRECIO	-10.82454	4.739128	-2.284078	0.0262
PUBLICIDAD	31.05061	1.604432	19.35302	0.0000
D51	-1017.342	590.3790	-1.723202	0.0904
D51*PRECIO	60.00307	43.66710	1.374102	0.1749
D51*PUBLICIDAD	-31.26711	14.63784	-2.136047	0.0371
R-squared	0.745183	Mean dependent var	1139.081	
Adjusted R-squared	0.722431	S.D. dependent var	234.5425	
S.E. of regression	123.5683	Akaike info criterion	12.56323	
Sum squared resid	855071.1	Schwarz criterion	12.76908	
Log likelihood	-383.4602	F-statistic	32.75304	
Durbin-Watson stat	2.206906	Prob(F-statistic)	0.000000	



SUPUESTO 2

- ⇒ Una Fundación Cultural ha recogido información relativa a los últimos 28 años sobre las iniciativas que ha organizado anualmente (conciertos, obras teatrales y exposiciones) y la recaudación total obtenida (millones de pesetas), que aparece almacenada en el archivo **CULTURA.WF1**.

Año	RECAUDACION	CONCIERTO	TEATRO	EXPOSICION
1970	119.98	124	65	85
1971	90.65	84	49	84
1972	85.83	76	44	95
1973	88.69	92	45	90
1974	94.45	86	53	74
1975	90.25	80	45	69
1976	81.61	76	42	90
1977	93.26	76	50	95
1978	97.68	86	49	95
1979	99.77	88	49	114
1980	97.53	88	50	92
1981	93.24	78	45	95
1982	99.96	80	50	82
1983	90.90	78	48	117
1984	93.96	82	48	99
1985	103.33	88	54	85
1986	101.70	88	52	98
1987	99.70	86	51	97
1988	94.18	84	51	104
1989	96.45	88	52	97
1990	96.02	82	51	90
1991	97.71	88	54	81
1992	91.75	82	53	85
1993	94.29	88	56	107
1994	86.11	80	48	89
1995	83.86	74	46	93
1996	89.62	78	50	92
1997	92.00	84	55	95
1998	95.60	83	60	96

- a) Estimar un modelo econométrico para explicar la evolución de la recaudación durante el período 1970-1996.
- b) Un miembro del Patronato de la Fundación ha afirmado: "*El efecto marginal sobre la recaudación de las representaciones teatrales duplica al de los conciertos, mientras las exposiciones presentan contribución nula*". Justificar si la información disponible apoya este supuesto.
- c) Contrastar la hipótesis de ausencia de correlación serial entre perturbaciones. ¿Sería recomendable alguna modificación sobre el modelo inicial?
- d) Efectuar predicciones de la producción para 1997 y 1998, analizando su bondad.

- a) La estimación mínimo cuadrática del modelo lineal que incluye como variables explicativas las actividades organizadas por la Fundación: $\hat{Y} = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$ proporciona el siguiente resultado:

Dependent Variable: RECAUDACION
 Method: Least Squares
 Sample: 1970 1996
 Included observations: 27

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21.16284	10.97278	1.928668	0.0662
CONCIERTO	0.377441	0.129237	2.920530	0.0077
TEATRO	0.761535	0.265748	2.865630	0.0087
EXPOSICION	0.037065	0.070278	0.527413	0.6030
R-squared	0.769318	Mean dependent var	94.53617	
Adjusted R-squared	0.739229	S.D. dependent var	7.438609	
S.E. of regression	3.798583	Akaike info criterion	5.643087	
Sum squared resid	331.8723	Schwarz criterion	5.835062	
Log likelihood	-72.18167	F-statistic	25.56808	
Durbin-Watson stat	0.826037	Prob(F-statistic)	0.000000	

donde se observan los signos esperados para los coeficientes estimados, pero se obtiene un nivel crítico elevado asociado al coeficiente de la variable EXPOSICION.

- b) Las hipótesis planteadas pueden ser contrastadas mediante el test de Wald que, por tratarse de restricciones lineales, resulta equivalente al test F de Snedecor:

La salida del programa Eviews para este contraste contiene los resultados asociados a ambos contrastes, obtenidos mediante la comparación de las sumas de los residuos cuadráticos asociados a los modelos libre ($\hat{u}'\hat{u}$) y restringido ($\hat{u}'_R \hat{u}_R$).

Wald Test:
 Null Hypothesis: C(3)=2*C(2)
 C(4)=0

F-statistic	0.140254	Probability	0.869875
Chi-square	0.280508	Probability	0.869138

Para el modelo libre (sin restricciones) hemos obtenido anteriormente $\hat{u}'\hat{u} = 331.8723$ mientras que la estimación restringida⁹ proporciona:

Dependent Variable: RECAUDACION
 Method: Least Squares
 Sample: 1970 1996
 Included observations: 27
RECAUDACION=C(1)+C(2)*CONCIERTO+2*C(2)*TEATRO=
= C(1)+C(2)*(CONCIERTO+2*TEATRO)

	Coefficient	Std. Error	t-Statistic	Prob.
C	25.01784	7.705909	3.246579	0.0033
CONCIERTO+2*TEATRO	0.376907	0.041604	9.059472	0.0000
R-squared	0.766517	Mean dependent var	94.53617	
Adjusted R-squared	0.757177	S.D. dependent var	7.438609	
S.E. of regression	3.665526	Akaike info criterion	5.507008	
Sum squared resid	335.9021	Schwarz criterion	5.602996	
Log likelihood	-72.34461	F-statistic	82.07403	
Durbin-Watson stat	0.709739	Prob(F-statistic)	0.000000	

⁹ Obsérvese que el modelo puede ser expresado en este caso como: $Y = \beta_1 + \beta_2 X_2 + 2\beta_2 X_3$ o equivalentemente $Y = \beta_1 + \beta_2 (X_2 + 2X_3)$

Es decir, al imponer al modelo los supuestos de *efecto marginal de las representaciones teatrales doble al de los conciertos* ($C(3)=2*C(2)$) y *contribución nula de las exposiciones* ($C(4)=0$) los residuos cuadráticos se elevan hasta $\hat{u}'_R \hat{u}_R = 335.9021$.

Las expresiones de las discrepancias F y W son respectivamente:

$$F = \left(\frac{\hat{u}'_R \hat{u}_R - \hat{u}' \hat{u}}{\hat{u}' \hat{u}} \right) \left(\frac{n-k}{r} \right) \approx F_{n-k}^r \quad W = \frac{\hat{u}'_R \hat{u}_R - \hat{u}' \hat{u}}{\hat{u}' \hat{u} / (n-k)} \approx \chi_r^2$$

que, sustituyendo los residuos cuadráticos de las estimaciones anteriores, el tamaño muestral $n=27$, el número de parámetros $k=4$ y el número de restricciones $r=2$ conducen a los resultados muestrales $F=0.140254$ $W=0.2805058$ entre los que se aprecia la relación $W=rF$

Los valores anteriores llevan asociados niveles críticos elevados, que no conducen al rechazo de la hipótesis planteada por el Patronato de la Fundación. Por tanto, consideramos a partir de ahora el modelo restringido anteriormente estimado.

- c) La hipótesis de ausencia de correlación serial entre perturbaciones puede ser contrastada mediante el test de Durbin-Watson, que para el modelo restringido proporciona el resultado $D=0.7097$, situado en la zona de autocorrelación positiva (puede comprobarse que los valores tabulados para una significación del 5% son en este caso $d_L=1.328$, $d_U=1.476$)

Teniendo en cuenta que para tamaños muestrales elevados se cumple $d=2(1-r)$ obtendríamos una correlación muestral $r \approx 0.645$, resultado similar a los obtenidos a partir de los residuos de la regresión (la correlación muestral entre \hat{u}_t y \hat{u}_{t-1} es 0.6416 mientras la regresión de los residuos \hat{u}_t sobre sus valores retardados \hat{u}_{t-1} proporciona 0.6469)¹⁰.

A la vista de estos datos, resulta aconsejable la incorporación al modelo de un componente autorregresivo de primer orden AR(1), que conduce a la estimación:

Dependent Variable: RECAUDACION		Method: Least Squares		
Sample(adjusted): 1971 1996		Included observations: 26		
Convergence achieved after 6 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21.44099	7.661119	2.798675	0.0102
CONCIERTO+2*TEATRO	0.396665	0.042459	9.342334	0.0000
AR(1)	0.645443	0.163114	3.957005	0.0006
R-squared	0.745071	Mean dependent var	93.55754	
Adjusted R-squared	0.722904	S.D. dependent var	5.536565	
S.E. of regression	2.914445	Akaike info criterion	5.085403	
Sum squared resid	195.3618	Schwarz criterion	5.230568	
Log likelihood	-63.11024	F-statistic	33.61065	
Durbin-Watson stat	2.158538	Prob(F-statistic)	0.000000	
Inverted AR Roots	.65			

¹⁰ La no coincidencia de resultados se justifica porque el coeficiente de correlación muestral sería:

$$\hat{\rho} = \frac{\sum_{t=2} \hat{u}_t \hat{u}_{t-1}}{\sqrt{\sum_{t=2} \hat{u}_t^2} \sqrt{\sum_{t=2} \hat{u}_{t-1}^2}} \quad \text{mientras que la regresión } \hat{u}_t = \hat{\rho} \hat{u}_{t-1} + \varepsilon_t \text{ conduce a una estimación } \hat{\rho} = \frac{\sum_{t=2} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2} \hat{u}_{t-1}^2} .$$

donde se aprecia un coeficiente AR(1) estimado 0.645 y un coeficiente de Durbin-Watson cercano al valor de referencia 2.

La inexistencia de correlación en este modelo puede comprobarse a partir del correlograma de los residuos (que incorpora el test Q de Ljung-Box) o bien efectuando el contraste LM, que proporciona niveles críticos elevados.

Sample: 1971 1996

Included observations: 26

Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. * .	. * .	1	-0.086	-0.086	0.2133	
. * .	. * .	2	-0.077	-0.085	0.3934	0.531
. **	. **	3	0.244	0.233	2.2797	0.320
. *	. *	4	0.082	0.123	2.5031	0.475
. *	. *	5	0.126	0.193	3.0571	0.548
. ** .	. ** .	6	-0.212	-0.251	4.6930	0.454
. .	. .	7	0.031	-0.049	4.7291	0.579
. * .	. ** .	8	-0.117	-0.292	5.2844	0.625
. * .	. .	9	-0.058	0.007	5.4286	0.711
. .	. .	10	0.023	0.018	5.4532	0.793
. ** .	. * .	11	-0.249	-0.064	8.4535	0.585
. .	. .	12	-0.024	-0.056	8.4825	0.670

- d) Para la realización de predicciones partimos del modelo anterior (restringido y con componente autorregresivo de primer orden) aplicando el comando FORECAST, que proporciona las alternativas estática (predicción con horizonte 1 o *one-step-ahead forecast*) y dinámica (predicción con horizonte cambiante 1,...h o *multi-step forecast*).

En cualquiera de las opciones, el programa proporciona los valores de recaudación previstos, su representación gráfica que incluye bandas de confianza y varias medidas resumen de los errores cometidos¹¹ (raíz del error cuadrático medio, error absoluto medio, error absoluto porcentual medio e índice de Theil, para el que se desglosan las participaciones de sesgo, varianza y covarianza).

Debemos tener presente que el índice propuesto por H. Theil (1958) puede ser definido a partir de los valores previstos o de las correspondientes tasas de variación. En concreto, el

¹¹ Lógicamente estas medidas de bondad sólo se encuentran disponibles cuando, como en este caso, la predicción se realiza *ex post*, esto es, para períodos en los que se conocen los verdaderos valores de la variable.

programa Eviews utiliza la expresión¹²:
$$U = \frac{\sqrt{\frac{\sum_{t=1}^{t+h} (Y_t - \hat{Y}_t)^2}{h}}}{\sqrt{\frac{\sum_{t=1}^{t+h} Y_t^2}{h}} + \sqrt{\frac{\sum_{t=1}^{t+h} \hat{Y}_t^2}{h}}}$$
 cuyo valor se encuentra

acotado entre 0 y 1, correspondiéndose los índices nulos con el caso de predicciones perfectas.

Se verifica además la propiedad de descomponibilidad, según la cual se cumple

$$U^2 = U_S^2 + U_V^2 + U_C^2, \text{ donde } U_S^2 = \frac{(\bar{Y} - \bar{\hat{Y}})^2}{\frac{\sum_{t=1}^{t+h} (Y_t - \hat{Y}_t)^2}{h}}$$
 recoge el componente de sesgo,

$$U_V^2 = \frac{(S_Y - S_{\hat{Y}})^2}{\frac{\sum_{t=1}^{t+h} (Y_t - \hat{Y}_t)^2}{h}} \text{ es el componente de varianza y } U_C^2 = \frac{2(1-r)S_Y S_{\hat{Y}}}{\frac{\sum_{t=1}^{t+h} (Y_t - \hat{Y}_t)^2}{h}} \text{ es el}$$

componente de covarianza.

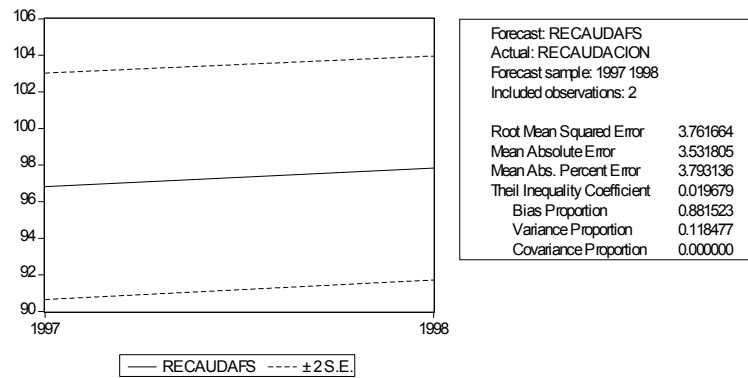
Las salidas proporcionadas por el programa Eviews aparecen resumidas a continuación, apreciándose una mayor fiabilidad para las predicciones estáticas, que llevan también asociado un menor componente de sesgo en el índice de Theil:

A la vista de estos resultados, concluimos que el modelo propuesto sobreestima las recaudaciones de la Fundación para los años 1997 y 1998, especialmente si se opta por la predicción dinámica (que ignoraría la información sobre recaudaciones registradas en el año 1997, conduciendo a un componente de sesgo del 99%).

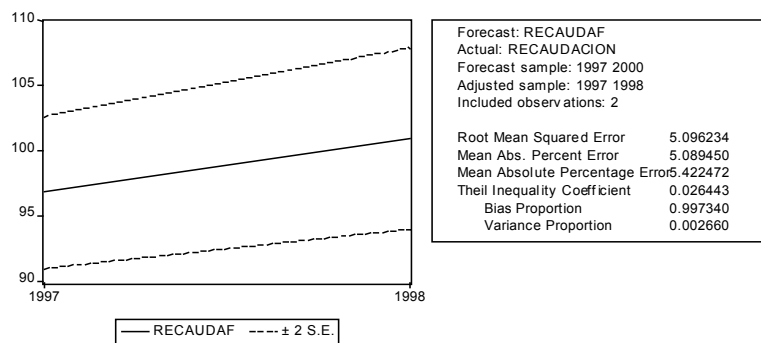
¹² Esta medida presenta el inconveniente de utilizar como referencia las predicciones (que aparecen en el

denominador de U), hecho que llevó a H.Theil (1966) a proponer el indicador
$$U = \frac{\sqrt{\frac{\sum_{t=1}^{t+h} (Y_t - \hat{Y}_t)^2}{h}}}{\sqrt{\frac{\sum_{t=1}^{t+h} Y_t^2}{h}}}$$
.

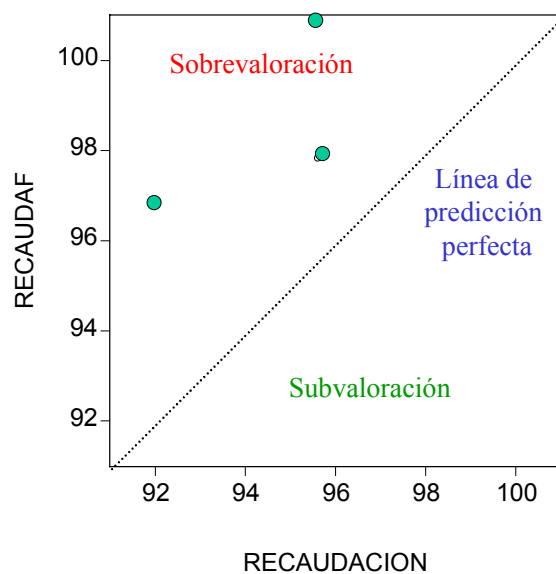
Predicción estática



Predicción dinámica



Otra representación útil para evaluar las predicciones es el *diagrama realización-predicción*, que puede ser construido a partir de los valores reales y previstos o bien de las correspondientes tasas asociadas. En nuestro caso, tomando el diagrama de puntos para valores de recaudación se aprecia que todas las predicciones se sitúan en la zona de sobrevaloración, alejándose más de la diagonal las cifras correspondientes a la predicción dinámica.



SUPUESTO 3

- Sobre una muestra de 40 familias se ha recopilado información relativa a sus gastos de transporte (miles de ptas/año), su renta familiar disponible (millones de pesetas/año) y su tipo de residencia, clasificada en las categorías rural y urbana, a las que se asocia valores 0 y 1 respectivamente. Dicha información se encuentra almacenada en el archivo **GASTO.WF1**.

obs	GASTO	RENTA	AMBITO	obs	GASTO	RENTA	AMBITO
1	10.4	4.3	1	21	7.3	3.2	0
2	7.7	2.9	0	22	7.2	6.6	0
3	10.6	5.1	1	23	11.3	7.8	1
4	11.4	8	1	24	7.3	4.3	0
5	7.3	5.6	0	25	7.1	3.4	0
6	6.9	2.3	0	26	11.5	7.8	1
7	6.5	2	0	27	7.2	4.7	0
8	10.2	3.8	1	28	10.7	8.2	1
9	7.2	4.5	0	29	10.6	6.3	1
10	7.3	6.7	0	30	10.1	3.6	1
11	11.2	9	1	31	7.3	5.5	0
12	7.4	6.8	0	32	8.1	6	0
13	10.2	4	1	33	10.9	7	1
14	11.1	9	1	34	6.9	4	0
15	6.8	2.3	0	35	7.0	3.8	0
16	7.2	3	0	36	11.2	9	1
17	7.3	4	0	37	11.6	8.2	1
18	10.6	5	1	38	7.3	4.8	0
19	10.6	5.7	1	39	7.0	5.1	0
20	7.3	5.4	0	40	7.2	4.6	0

- a) Proponer un modelo econométrico adecuado para explicar el gasto familiar en transporte.
 b) Estimar un modelo *logit* para el ámbito de residencia familiar a partir de la renta y evaluar su capacidad predictiva.

- a) La estimación de un modelo lineal para el gasto familiar en transporte a partir de la renta proporcionaría el resultado:

Dependent Variable: GASTO
 Method: Least Squares
 Sample: 1 40
 Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.415479	0.647942	8.357972	0.0000
RENTA	0.626778	0.114098	5.493352	0.0000
R-squared	0.442627	Mean dependent var		8.757772
Adjusted R-squared	0.427959	S.D. dependent var		1.863226
S.E. of regression	1.409221	Akaike info criterion		3.572658
Sum squared resid	75.46434	Schwarz criterion		3.657102
Log likelihood	-69.45316	F-statistic		30.17691
Durbin-Watson stat	2.085992	Prob(F-statistic)		0.000003

y teniendo en cuenta el bajo coeficiente de determinación y el análisis gráfico de la nube de puntos parece aconsejable tener en cuenta la variable *dummy* asociada al ámbito de residencia:

Dependent Variable: GASTO

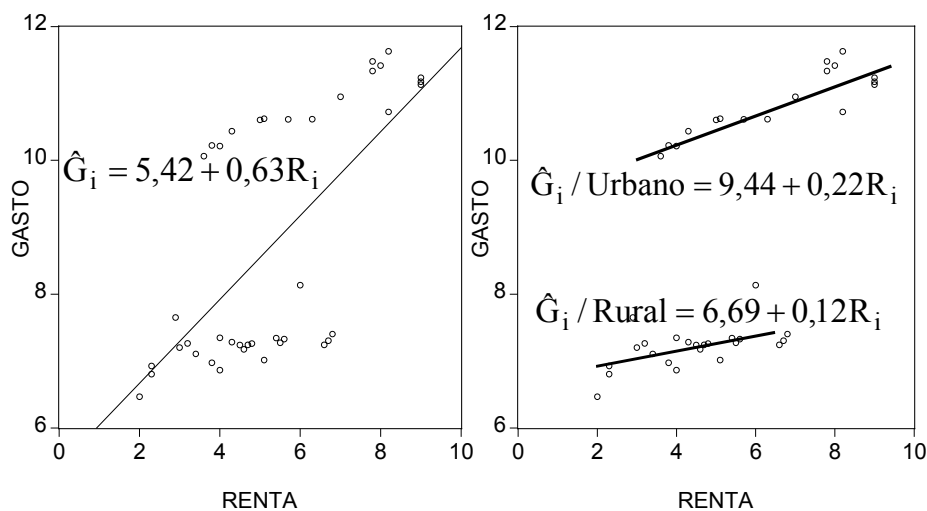
Method: Least Squares

Sample: 1 40

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.688337	0.179134	37.33714	0.0000
RENTA	0.118436	0.038702	3.060173	0.0042
AMBITO	2.746779	0.287856	9.542213	0.0000
AMBITO*RENTA	0.096779	0.050799	1.905140	0.0648
R-squared	0.982155	Mean dependent var		8.757772
Adjusted R-squared	0.980668	S.D. dependent var		1.863226
S.E. of regression	0.259063	Akaike info criterion		0.231147
Sum squared resid	2.416086	Schwarz criterion		0.400035
Log likelihood	-0.622933	F-statistic		660.4567
Durbin-Watson stat	1.735442	Prob(F-statistic)		0.000000

A la vista de estos resultados parece claro que el ámbito de residencia familiar afecta al gasto fijo en transporte incrementándolo en unas 2747 ptas. y también al impacto marginal que la renta familiar tiene sobre el gasto. Así, por cada millón adicional de renta familiar el gasto de transporte se incrementaría en 118 ptas. para familias de ámbito rural y en 215 ptas. para las de ámbito urbano.



b) La estimación de un modelo *logit* para la variable cualitativa *Ambito* proporciona el siguiente resultado con la información disponible:

Dependent Variable: AMBITO

Method: ML - Binary Logit

Sample: 1 40

Included observations: 40

Convergence achieved after 4 iterations

Covariance matrix computed using second derivatives

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-4.227000	1.357714	-3.113322	0.0018
RENTA	0.722692	0.239472	3.017856	0.0025
Mean dependent var	0.425000	S.D. dependent var		0.500641
S.E. of regression	0.428127	Akaike info criterion		1.129606
Sum squared resid	6.965117	Schwarz criterion		1.214050
Log likelihood	-20.59213	Hannan-Quinn criter.		1.160139
Restr. log likelihood	-27.27418	Avg. log likelihood		-0.514803
LR statistic (1 df)	13.36412	McFadden R-squared		0.244996
Probability(LR stat)	0.000256			
Obs with Dep=0	23	Total obs		40
Obs with Dep=1	17			

Como puede apreciarse hemos llegado a la expresión:

$$\hat{L}_i = \text{Log} \frac{P_i}{1 - P_i} = -4,227 + 0,722 \text{ Renta}$$

a partir de la cual se obtienen los logit estimados, que a su vez permiten despejar las probabilidades estimadas P_i con los resultados de la tabla¹³:

obs	Ambito	Renta	Estimación L_i	Estimación P_i	Predicción	Error
1	1	4.3	-1.12	0.25	0	-1
2	0	2.9	-2.13	0.11	0	0
3	1	5.1	-0.54	0.37	0	-1
4	1	8	1.55	0.83	1	0
5	0	5.6	-0.18	0.46	0	0
6	0	2.3	-2.56	0.07	0	0
7	0	2	-2.78	0.06	0	0
8	1	3.8	-1.48	0.19	0	-1
9	0	4.5	-0.97	0.27	0	0
10	0	6.7	0.62	0.65	1	1
11	1	9	2.28	0.91	1	0
12	0	6.8	0.69	0.67	1	1
13	1	4	-1.34	0.21	0	-1
14	1	9	2.28	0.91	1	0
15	0	2.3	-2.56	0.07	0	0
16	0	3	-2.06	0.11	0	0
17	0	4	-1.34	0.21	0	0
18	1	5	-0.61	0.35	0	-1

¹³ Obsérvese que si denotamos los *logit* por L_i se tendría:

$$\text{Log} \frac{P_i}{1 - P_i} = L_i \Rightarrow \frac{P_i}{1 - P_i} = e^{L_i} \Rightarrow P_i = \frac{e^{L_i}}{1 + e^{L_i}}$$

obs	Ambito	Renta	Estimación L_i	Estimación P_i	Predicción	Error
19	1	5.7	-0.11	0.47	0	-1
20	0	5.4	-0.32	0.42	0	0
21	0	3.2	-1.91	0.13	0	0
22	0	6.6	0.54	0.63	1	1
23	1	7.8	1.41	0.80	1	0
24	0	4.3	-1.12	0.25	0	0
25	0	3.4	-1.77	0.15	0	0
26	1	7.8	1.41	0.80	1	0
27	0	4.7	-0.83	0.30	0	0
28	1	8.2	1.70	0.85	1	0
29	1	6.3	0.33	0.58	1	0
30	1	3.6	-1.63	0.16	0	-1
31	0	5.5	-0.25	0.44	0	0
32	0	6	0.11	0.53	1	1
33	1	7	0.83	0.70	1	0
34	0	4	-1.34	0.21	0	0
35	0	3.8	-1.48	0.19	0	0
36	1	9	2.28	0.91	1	0
37	1	8.2	1.70	0.85	1	0
38	0	4.8	-0.76	0.32	0	0
39	0	5.1	-0.54	0.37	0	0
40	0	4.6	-0.90	0.29	0	0

En el caso de utilizar este modelo con fines predictivos, pronosticaríamos ámbito de residencia urbano para probabilidades superiores al 50%. Así, se aprecia que del total muestral (n=40) habría 29 clasificaciones correctas (72,5%) y 11 erróneas (27,5%), de las cuales 7 son “falsos negativos” y los 4 restantes se corresponden con “falsos positivos”.

SUPUESTO 4

⇒ Las empresas *General Electric* y *Westinghouse* han facilitado información relativa a las magnitudes Inversión bruta (I), Valor de mercado (VM) y Valor del stock de maquinaria y establecimientos (VS), todas ellas expresadas en millones de dólares constantes. A partir de dicha información, recopilada en el archivo **EMPRESAS.WF1**, diversos autores¹⁴ han propuesto modelos SUR para la inversión.

Año	General Electric			Westinghouse		
	INVERSIÓN	VALOR ACCIONES	CAPITAL	INVERSIÓN	VALOR ACCIONES	CAPITAL
1935	33.1	1170.6	97.8	12.9	191.5	1.8
1936	45.0	2015.8	104.4	25.9	516.0	0.8
1937	77.2	2803.3	118.0	35.1	729.0	7.4
1938	44.6	2039.7	156.2	22.9	560.4	18.1
1939	48.1	2256.2	172.6	18.8	519.9	23.5
1940	74.4	2132.2	186.6	28.6	628.5	26.5
1941	113.0	1834.1	220.9	48.5	537.1	36.2
1942	91.9	1588.0	287.8	43.3	561.2	60.8
1943	61.3	1749.4	319.9	37.0	617.2	84.4
1944	56.8	1687.2	321.3	37.8	626.7	91.2
1945	93.6	2007.7	319.6	37.3	737.2	92.4
1946	159.9	2208.3	346.0	53.5	760.5	86.0
1947	147.2	1656.7	456.4	55.6	581.4	111.1
1948	146.3	1604.4	543.4	49.6	662.3	130.6
1949	98.3	1431.8	618.3	32.0	583.8	141.8
1950	93.5	1610.5	647.4	32.2	635.2	136.7
1951	135.2	1819.4	671.3	54.4	723.8	129.7
1952	157.3	2079.7	726.1	71.8	864.1	145.5
1953	179.5	2371.6	800.3	90.1	1193.5	174.8
1954	189.6	2759.9	888.9	68.6	1188.9	213.5

- a) Llevar a cabo la estimación de modelos lineales de inversión para Westinghouse y General Electric.
- b) Plantear un modelo de inversión para las empresas, analizando la existencia de correlación contemporánea entre las perturbaciones.
- c) Estimar un modelo SUR, comparando su resultado con los obtenidos anteriormente.
- d) Contrastar la igualdad de coeficientes, estudiando la agregabilidad del sistema de inversión analizado.

¹⁴ Los trabajos originales –que incluyen además de *General Electric* y *Westinghouse* otras empresas como *General Motors*, *Chrysler* y *U.S. Steel*- corresponden a Grunfeld (1958), Grunfeld y Griliches (1960) y J.C.G. Boot y G.M. De Witt (1960).

- a) Comenzamos llevando a cabo la estimación mínimo cuadrática para las dos empresas separadamente:

Dependent Variable: IGE

Method: Least Squares

Sample: 1935 1954

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.956306	31.37425	-0.317340	0.7548
VMGE	0.026551	0.015566	1.705705	0.1063
VSGE	0.151694	0.025704	5.901548	0.0000
R-squared	0.705307	Mean dependent var		102.2900
Adjusted R-squared	0.670637	S.D. dependent var		48.58450
S.E. of regression	27.88272	Akaike info criterion		9.631373
Sum squared resid	13216.59	Schwarz criterion		9.780733
Log likelihood	-93.31373	F-statistic		20.34355
Durbin-Watson stat	1.072099	Prob(F-statistic)		0.000031

Dependent Variable: IW

Method: Least Squares

Sample: 1935 1954

Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.492047	8.093525	-0.060795	0.9522
VMW	0.052665	0.015860	3.320675	0.0040
VSW	0.092829	0.056647	1.638744	0.1196
R-squared	0.740118	Mean dependent var		42.79150
Adjusted R-squared	0.709543	S.D. dependent var		19.13535
S.E. of regression	10.31281	Akaike info criterion		7.642132
Sum squared resid	1808.019	Schwarz criterion		7.791492
Log likelihood	-73.42132	F-statistic		24.20713
Durbin-Watson stat	1.430075	Prob(F-statistic)		0.000011

Puede observarse que los coeficientes estimados presentan el signo esperado, si bien se obtienen niveles críticos superiores al 10% para algunos contrastes de significación individual. Además, teniendo en cuenta que estamos considerando modelos de inversión del tipo:

$$I_t^{GE} = \beta_{10} + \beta_{11} VM_t^{GE} + \beta_{12} VS_t^{GE} + u_t^{GE}$$

$$I_t^W = \beta_{20} + \beta_{21} VM_t^W + \beta_{22} VS_t^W + u_t^W$$

cabe sospechar que existirán correlaciones contemporáneas entre las perturbaciones de ambas ecuaciones, como consecuencia de la presencia de pautas comunes de comportamiento en la inversión de las empresas consideradas.

- b) Para contrastar estadísticamente la presencia de correlación contemporánea entre las perturbaciones de los modelos de inversión de *General Electric* y *Westinghouse* resulta conveniente considerar un sistema de dos ecuaciones que vendría especificado como:

$IGE=C(10)+C(11)*VMGE+C(12)*VSGE$ $IW=C(20)+C(21)*VMW+C(22)*VSW$
--

y cuya estimación podría ser llevada a cabo mediante diferentes procedimientos. En concreto, la estimación mínimo cuadrática equivaldría a considerar las dos ecuaciones como *efectivamente no relacionadas*, pudiendo comprobarse que proporciona resultados coincidentes con los anteriormente obtenidos para los modelos individuales:

System: SISTEMA

Estimation Method: Least Squares

Sample: 1935 1954

Included observations: 20

Total system (balanced) observations 40

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-9.956306	31.37425	-0.317340	0.7529
C(2)	0.026551	0.015566	1.705705	0.0972
C(3)	0.151694	0.025704	5.901548	0.0000
C(4)	-0.492047	8.093525	-0.060795	0.9519
C(5)	0.052665	0.015860	3.320675	0.0022
C(6)	0.092829	0.056647	1.638744	0.1105

Determinant residual covariance 28643.67

Equation: IGE=C(1)+C(2)*VMGE+C(3)*VSGE

Observations: 20

R-squared	0.705307	Mean dependent var	102.2900
Adjusted R-squared	0.670637	S.D. dependent var	48.58450
S.E. of regression	27.88272	Sum squared resid	13216.59
Durbin-Watson stat	1.072099		

Equation: IW=C(4)+C(5)*VMW+C(6)*VSW

Observations: 20

R-squared	0.740118	Mean dependent var	42.79150
Adjusted R-squared	0.709543	S.D. dependent var	19.13535
S.E. of regression	10.31281	Sum squared resid	1808.019
Durbin-Watson stat	1.430075		

A partir de esta estimación es posible además obtener información relativa a los residuos de ambas regresiones¹⁵, resumidos en la matriz de varianzas-covarianzas (*Residual Covariance Matrix*).

	IGE	IW
IGE	660.8294	176.3404
IW	176.3404	90.40097

cuyo determinante aparece recogido en la salida anterior y proporciona un indicador del nivel de dispersión del modelo estimado.

Mediante normalización de la matriz anterior se obtienen también las correlaciones entre residuos (*Residual Correlation Matrix*):

	IGE	IW
IGE	1.000000	0.721474
IW	0.721474	1.000000

¹⁵ Las series completas de residuos de las dos ecuaciones pueden ser recuperadas mediante la opción *Make Residuals* desde el output de la estimación.

La hipótesis de incorrelación de perturbaciones ($\rho_{12}=0$) puede ser contrastada entonces mediante

la expresión $\lambda = T \sum_{i=2}^m \sum_{j=1}^{i-1} r_{ij}^2 \approx \chi^2_{\frac{m(m-1)}{2}}$ que en nuestro caso, con $m=2$, abarca un único

sumando proporcionando el resultado $Tr_{12}^2=20(0.7214)^2=10.41$ que lleva asociado un nivel crítico en torno al 1%.

- c) Una vez que disponemos de la información anterior, planteamos la estimación del sistema mediante metodología SUR (regresiones aparentemente no relacionadas) obteniendo el resultado:

System: SISTEMA

Estimation Method: Iterative Seemingly Unrelated Regression

Sample: 1935 1954

Included observations: 20

Total system (balanced) observations 40

Convergence achieved after: 6 weight matrices, 7 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-29.31763	27.40684	-1.069719	0.2923
C(2)	0.039788	0.013446	2.959022	0.0056
C(3)	0.135863	0.023545	5.770437	0.0000
C(4)	-1.438570	7.016668	-0.205022	0.8388
C(5)	0.058606	0.013478	4.348301	0.0001
C(6)	0.057341	0.049366	1.161548	0.2535

Determinant residual covariance 26934.36

Equation: IGE=C(1)+C(2)*VMGE+C(3)*VSGE

Observations: 20

R-squared	0.688103	Mean dependent var	102.2900
Adjusted R-squared	0.651409	S.D. dependent var	48.58450
S.E. of regression	28.68507	Sum squared resid	13988.16
Durbin-Watson stat	0.969269		

Equation: IW=C(4)+C(5)*VMW+C(6)*VSW

Observations: 20

R-squared	0.733842	Mean dependent var	42.79150
Adjusted R-squared	0.702530	S.D. dependent var	19.13535
S.E. of regression	10.43658	Sum squared resid	1851.678
Durbin-Watson stat	1.366385		

donde se aprecian ligeros cambios en la estimación de los parámetros y los contrastes de significación asociados, sin que se produzcan disminuciones en las varianzas estimadas de los residuos.

Cabe señalar que el método empleado por el programa Eviews para llegar a esta solución parte de una estimación mínimo cuadrática inicial, cuyos residuos proporcionan estimaciones consistentes de la matriz de ponderaciones, que incorporada al modelo proporciona una nueva estimación. Este procedimiento puede ser aplicado de forma iterativa, y permite también la incorporación de autocorrelaciones mediante términos AR.

Concretamente, en el caso de nuestro modelo de inversión se consiguen mejoras sustanciales al incorporar autocorrelaciones en ambas ecuaciones:

Estimation Method: Iterative Seemingly Unrelated Regression

Included observations: 20

Total system (balanced) observations 36

Sequential weighting matrix & coefficient iteration

Convergence achieved after: 6 weight matrices, 25 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-23.95940	24.39713	-0.982058	0.3351
C(2)	0.037186	0.011818	3.146432	0.0041
C(3)	0.137480	0.024995	5.500333	0.0000
C(11)	0.832384	0.169615	4.907497	0.0000
C(12)	-0.594384	0.164549	-3.612193	0.0013
C(4)	8.461310	6.732489	1.256788	0.2200
C(5)	0.034196	0.012523	2.730674	0.0112
C(6)	0.125251	0.037726	3.320053	0.0027
C(21)	0.750787	0.167291	4.487916	0.0001
C(22)	-0.813108	0.173384	-4.689636	0.0001

Determinant residual covariance 8979.023

Equation: IGE=C(1)+C(2)*VMGE+C(3)*VSGE+ [AR(1)=C(11) AR(2)=C(12)]

Observations: 18

R-squared	0.840838	Mean dependent var	109.3167
Adjusted R-squared	0.791864	S.D. dependent var	45.94785
S.E. of regression	20.96227	Sum squared resid	5712.418
Durbin-Watson stat	1.968077		

Equation: IW=C(4)+C(5)*VMW+C(6)*VSW+[AR(1)=C(21) AR(2)=C(22)]

Observations: 18

R-squared	0.855671	Mean dependent var	45.38889
Adjusted R-squared	0.811262	S.D. dependent var	18.24443
S.E. of regression	7.926101	Sum squared resid	816.6999
Durbin-Watson stat	2.697418		

apreciándose además una considerable reducción en la diagonal de la matriz de varianzas covarianzas de los residuos:

	IGE	IW
IGE	317.3565	73.62164
IW	73.62164	45.37222

c) La igualdad de coeficientes puede ser estudiada mediante el test de Wald, que aplicado al último modelo SUR estimado proporciona el resultado:

Wald Test:

Null Hypothesis: C(1)=C(4)
C(2)=C(5)
C(3)=C(6)

Chi-square	3.930577	Probability	0.269059
------------	----------	-------------	----------

obtenido a partir de la comparación de los residuos del modelo restringido con los correspondientes al modelo libre¹⁶.

¹⁶ En este tipo de modelos, la discrepancia asociada al test de Wald viene dada por la expresión:

$$(y - X\hat{\beta}^*) (\hat{\Sigma}^{-1} \otimes I) (y - X\hat{\beta}^*) - (y - X\hat{\beta}) (\hat{\Sigma}^{-1} \otimes I) (y - X\hat{\beta}) \approx \chi_r^2$$

donde $\hat{\Sigma}$ es la matriz de varianzas-covarianzas estimada y r el número de restricciones contrastadas (r=2 en este caso).

Con nuestra información disponible, el test de Wald proporciona un resultado no significativo para rechazar el supuesto de igualdad de parámetros para los dos modelos de inversión. Por tanto, podríamos plantear la agregación de ambas ecuaciones en un único modelo de inversión:

Dependent Variable: INVERSION

Method: Least Squares

Sample(adjusted): 3 20

Included observations: 18 after adjusting endpoints

Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VM	0.028171	0.004928	5.716328	0.0001
VS	0.147490	0.023151	6.370667	0.0000
AR(1)	0.723372	0.193384	3.740610	0.0022
AR(2)	-0.733734	0.193109	-3.799584	0.0020
R-squared	0.863815	Mean dependent var	154.7056	
Adjusted R-squared	0.834632	S.D. dependent var	62.82726	
S.E. of regression	25.54900	Akaike info criterion	9.512203	
Sum squared resid	9138.519	Schwarz criterion	9.710064	
Log likelihood	-81.60983	Durbin-Watson stat	2.263884	
Inverted AR Roots	.36 -.78i	.36+.78i		

SUPUESTO 5

- ⇒ Con el objetivo de analizar la evolución del sector industrial español se ha recopilado la siguiente información temporal referida a los salarios (W) los precios (P), la producción (Q) y el Valor Añadido Bruto (V), que ha sido almacenada en el archivo **INDUSTRIA.WF1**.

Año	W	P	Q	V
1981	65547.3	58.2	82.2	8686.3
1982	75030.5	65.5	81.2	8540.3
1983	85134.8	74.6	83.4	8762.5
1984	93023.3	83.8	84.1	8716.4
1985	101941.8	90.4	85.8	8995.4
1986	113520.8	91.3	88.5	9427.0
1987	121579.3	92.0	92.6	9874.3
1988	128843.0	94.8	95.4	10322.1
1989	136245.8	99	100	10698.5
1990	147886.8	100.9	99.7	10917.9
1991	159067.8	102.3	99.0	11069.2
1992	171040.0	103.8	96.2	11065.2
1993	181918.8	106.2	91.6	10765.0
1994	190550.5	110.7	98.3	11142.7
1995	199145.0	117.8	102.9	11585.6
1996	208047.8	119.8	102.2	11744.2
1997	215188.5	121.0	109.2	12383.3
1998	213250.3	120.3	114.7	13062.9

W: Ganancia salarial/mes, pagos totales, todas las secciones y categorías, total nacional (Fuente: INE, Encuesta de Salarios)

P: Índice de Precios Industriales, base 1990 (Fuente: INE. IPRI, Media anual, índice general)

Q: Índice de Producción Industrial, base 1990 (Fuente: INE. IPI Media anual, índice general)

V: Valor Añadido Bruto del sector industrial, miles de pesetas (Fuente: INE. Contabilidad Nacional de España)

- a) Se ha propuesto el siguiente modelo simplificado de salario-precio [M1]:

$$W_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Q_t + u_{1t}$$

$$P_t = \beta_0 + \beta_1 W_t + u_{2t}$$

Llevar a cabo su estimación mínimo cuadrática comentando los resultados.

- b) Si la segunda ecuación incluye como explicativa el VAB industrial [modelo M2], estudiar su identificabilidad y llevar a cabo su estimación.
- c) Se considera ahora el modelo M3 en el que la primera ecuación incluye la variable adicional precio retardado P_{t-1} . Estudiar los procedimientos adecuados para su estimación.
- d) Contrastar si la producción industrial es causa de los salarios en el sentido de Granger. ¿Podríamos afirmar que las ganancias salariales causan los precios industriales?

a) Estudiando las condiciones de orden y rango sobre el modelo M1 se observa que la ecuación es no identificada y la segunda perfectamente identificada.

SISTEMA		m	k
M1		2	1

ECUACIONES		m'	k'	k-k'	m'-1	Rango(A)	
W=C(10)+C(11)*P+C(12)*Q	EC1	2	1	0	1	1	No identificada
P=C(20)+C(21)*W	EC2	2	0	1	1	1	Identificada

La estimación del sistema por Mínimos Cuadrados Ordinarios proporciona los resultados:

System: M1

Estimation Method: Least Squares

Sample: 1981 1998

Included observations: 18

Total system (balanced) observations 36

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	-154659.0	40659.74	-3.803737	0.0006
C(11)	2213.118	383.5255	5.770459	0.0000
C(12)	886.8702	742.2216	1.194886	0.2412
C(20)	45.73147	3.597217	12.71301	0.0000
C(21)	0.000356	2.36E-05	15.12228	0.0000

Determinant residual covariance 4.24E+08

Equation: W=C(10)+C(11)*P+C(12)*Q

Observations: 18

R-squared	0.940293	Mean dependent var	144831.2
Adjusted R-squared	0.932332	S.D. dependent var	49727.46
S.E. of regression	12935.69	Sum squared resid	2.51E+09
Durbin-Watson stat	0.423725		

Equation: P=C(20)+C(21)*W

Observations: 18

R-squared	0.934609	Mean dependent var	97.33278
Adjusted R-squared	0.930522	S.D. dependent var	18.32651
S.E. of regression	4.830612	Sum squared resid	373.3569
Durbin-Watson stat	0.419091		

pero somos conscientes de que, al encontrarse las variables endógenas explicativas W y P correlacionadas con las perturbaciones, esta estimación sería sesgada e inconsistente.

Si expresamos ahora el modelo en forma reducida [MR1]:

$$W_t = \pi_{10} + \pi_{11}Q_t + v_{1t} \quad \pi_{10} = \frac{\alpha_0 + \alpha_1\beta_0}{1 - \alpha_1\beta_1} \quad \pi_{11} = \frac{\alpha_2}{1 - \alpha_1\beta_1}$$

$$P_t = \pi_{20} + \pi_{21}Q_t + v_{2t} \quad \pi_{20} = \beta_0 + \beta_1\pi_{10} \quad \pi_{21} = \beta_1\pi_{11}$$

podríamos llevar a cabo su estimación mínimo cuadrática, aplicando posteriormente Mínimos Cuadrados Indirectos (MCI) a la ecuación del precio mediante las expresiones:

$$\hat{\beta}_1 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{11}} \quad \hat{\beta}_0 = \hat{\pi}_{20} - \hat{\beta}_1\hat{\pi}_{10}$$

que conducen¹⁷ a la estimación $\hat{P}_t = 44.188 + 0.00036Q_t$, no coincidente con el resultado mínimo cuadrático.

b) El modelo M2 vendría especificado en los siguientes términos:

$$W_t = \alpha_0 + \alpha_1 P_t + \alpha_2 Q_t + u_{1t}$$

$$P_t = \beta_0 + \beta_1 W_t + \beta_2 V_t + u_{2t}$$

y es posible observar que ambas ecuaciones están perfectamente identificadas

SISTEMA		m	k
M2		2	2

ECUACIONES		m'	k'	k-k'	m'-1	Rango(A)	
W=C(10)+C(11)*P+C(12)*Q	EC1	2	1	1	1	1	Identificada
P=C(20)+C(21)*W+C(22)*V	EC2	2	1	1	1	1	Identificada

En consecuencia, la estimación del sistema podría llevarse a cabo indistintamente por Mínimos cuadrados en dos etapas (MC2E) o por Mínimos Cuadrados Indirectos (MCI)¹⁸.

En el primer caso, utilizando como variables instrumentales¹⁹ las predeterminadas Q e Y, el programa Eviews proporcionaría:

System: M2

Estimation Method: Two-Stage Least Squares

Sample: 1981 1998

Included observations: 18

Total system (balanced) observations 36

Instruments: Q V C

	Coefficient	Std. Error	t-Statistic	Prob.
C(10)	-5681.703	105003.8	-0.054109	0.9572
C(11)	4441.634	1320.137	3.364525	0.0021
C(12)	-2972.403	2362.922	-1.257935	0.2181
C(20)	-14.78429	40.36829	-0.366235	0.7168
C(21)	8.17E-05	0.000173	0.471921	0.6404
C(22)	0.009614	0.006199	1.551000	0.1314

Determinant residual covariance 3.87E+08

Equation: W=C(10)+C(11)*P+C(12)*Q

Observations: 18

R-squared	0.805898	Mean dependent var	144831.2
Adjusted R-squared	0.780018	S.D. dependent var	49727.46
S.E. of regression	23323.29	Sum squared resid	8.16E+09
Durbin-Watson stat	0.505414		

Equation: P=C(20)+C(21)*W+C(22)*V

Observations: 18

R-squared	0.885541	Mean dependent var	97.33278
Adjusted R-squared	0.870280	S.D. dependent var	18.32651
S.E. of regression	6.600608	Sum squared resid	653.5204
Durbin-Watson stat	0.515995		

¹⁷ Una vez almacenada la estimación del modelo reducido en un vector de coeficientes (denominado R en nuestro caso), los cálculos con el programa Eviews pueden ser efectuados mediante los comandos: C(21)=R(21)/R(11) C(20)=R(20)-C(21)*R(10).

¹⁸ Puede comprobarse que también los métodos MC3E o MVIC proporcionarían el mismo resultado.

¹⁹ La opción MC2E del programa Eviews (*Two Stage Least Squares* o TSLS) exige, además de la especificación del modelo, la lista de variables instrumentales que se utilizarán en la estimación.

resultado al que se llegaría estimando en una primera etapa las variables endógenas explicativas (W, P) en función de las instrumentales (Q, V) y sustituyendo en una segunda etapa las variables endógenas explicativas por sus estimaciones²⁰.

Por lo que se refiere a la estimación por MCI, comenzaríamos por aproximar el modelo reducido:

System: MR2

Estimation Method: Least Squares

Sample: 1981 1998

Included observations: 18

Total system (balanced) observations 36

	Coefficient	Std. Error	t-Statistic	Prob.
R(10)	-111962.4	38992.94	-2.871350	0.0074
R(11)	-4664.418	1385.055	-3.367677	0.0021
R(12)	67.01259	9.671668	6.928752	0.0000
R(20)	-23.92828	25.54065	-0.936870	0.3563
R(21)	-0.380944	0.907221	-0.419902	0.6775
R(22)	0.015087	0.006335	2.381586	0.0238
Determinant residual covariance		9.52E+08		

Equation: W=R(10)+R(11)*Q+R(12)*V

Observations: 18

R-squared	0.954232	Mean dependent var	144831.2
Adjusted R-squared	0.948129	S.D. dependent var	49727.46
S.E. of regression	11325.53	Sum squared resid	1.92E+09
Durbin-Watson stat	0.736271		

Equation: P=R(20)+R(21)*Q+R(22)*V

Observations: 18

R-squared	0.855426	Mean dependent var	97.33278
Adjusted R-squared	0.836149	S.D. dependent var	18.32651
S.E. of regression	7.418302	Sum squared resid	825.4682
Durbin-Watson stat	0.528453		

obteniendo estimadores de los parámetros estructurales a partir de los reducidos mediante las expresiones:

$$\hat{\alpha}_0 = \hat{\pi}_{20} \begin{pmatrix} \hat{\pi}_{10} & -\hat{\pi}_{12} \\ \hat{\pi}_{20} & \hat{\pi}_{22} \end{pmatrix} \quad \hat{\alpha}_1 = \frac{\hat{\pi}_{12}}{\hat{\pi}_{22}} \quad \hat{\alpha}_2 = \hat{\pi}_{12} \begin{pmatrix} \hat{\pi}_{11} & -\hat{\pi}_{21} \\ \hat{\pi}_{12} & \hat{\pi}_{22} \end{pmatrix}$$

$$\hat{\beta}_0 = \hat{\pi}_{10} \begin{pmatrix} \hat{\pi}_{20} & -\hat{\pi}_{21} \\ \hat{\pi}_{10} & \hat{\pi}_{11} \end{pmatrix} \quad \hat{\beta}_1 = \frac{\hat{\pi}_{21}}{\hat{\pi}_{11}} \quad \hat{\beta}_2 = \hat{\pi}_{12} \begin{pmatrix} \hat{\pi}_{22} & -\hat{\pi}_{21} \\ \hat{\pi}_{12} & \hat{\pi}_{11} \end{pmatrix}$$

²⁰ Es posible reproducir los cálculos necesarios para esta estimación bietápica. Así, la primera estimación (mediante Variables Instrumentales) conduciría a los resultados:

$$\hat{P}_t = -23.93 - 0.38 Q_t + 0.015 V_t$$

$$\hat{W}_t = -111962.4 - 4664.418 Q_t + 67.012 V_t$$

a partir de los cuales, una vez almacenadas las variables estimadas $\hat{P}_t \hat{W}_t$ se llega al resultado de la segunda etapa:

$$\hat{W}_t = -5681.7 + 4441 \hat{P}_t - 2972.4 Q_t$$

$$\hat{P}_t = -14.78 + 0.00008 \hat{W}_t + 0.009 V_t$$

c) Para el modelo M3 se tiene:

SISTEMA		m	k
M3		2	3

ECUACIONES		m'	k'	k-k'	m'-1	$\rho(A)$	
$W=C(10)+C(11)*P+C(12)*Q+ C(13)*P(-1)$	EC1	2	2	1	1	1	Identificada
$P=C(20)+C(21)*W+C(22)*V$	EC2	2	1	2	1	1	Sobreidentificada

y, dado que la segunda ecuación se encuentra sobreidentificada, no resultaría aplicable en este caso Mínimos Cuadrados Indirectos.

La tabla siguiente recoge una síntesis de las estimaciones el modelo M3 llevadas a cabo con varios procedimientos alternativos:

	MCO	MC2E	MC3E	MVIC	MGM
C(10)	-149950.2 (49103.9)	-1955273 (12192340)	3663950 (10562013)	3650594 (533000000)	1521195 (10109117)
C(11)	2725.88 (1349.5)	112652.6 (739940.2)	-226161.9 (641076)	-346325.1 (47764492)	-102843.4 (588944.4)
C(12)	321.84 (801.4)	3506.12 (28055.3)	-8173.05 (24346.3)	32357.74 (4140103)	-2710.67 (25006.9)
C(13)	-30.16 (1276.6)	-98499.49 (662883)	206285.6 (574269.9)	291351.3 (39986492)	95035.67 (529860.5)
C(20)	49.21 (21.2)	43.58 (19.3)	43.58 (17.5)	114.4 (83.03)	36.32 (14.5)
C(21)	0.00037 (0.00008)	0.0003 (0.00008)	0.0003 (0.00008)	0.0007 (0.0004)	0.000221 (0.00006)
C(22)	-0.0005 (0.003)	0.001 (0.003)	0.001 (0.003)	-0.01 (0.01)	0.002941 (0.002)

Especificación del Modelo: $W=C(10)+C(11)*P+C(12)*Q+C(13)*P(-1)$
 $P=C(20)+C(21)*W+C(22)*V$
 Variables Instrumentales: Q, V, P(-1)
 Métodos de estimación:
 MCO: Mínimos Cuadrados Ordinarios (LS)
 MC2E: Mínimos Cuadrados en 2 Etapas (2SLS)
 MC3E: Mínimos Cuadrados en 3 Etapas (3SLS)
 MVIC: Máxima Verosimilitud con Información completa (FIML)
 MGM: Método Generalizado de Momentos (GMM)

Como puede apreciarse, existen diferencias notables entre los resultados, tanto en lo que se refiere a los signos de los coeficientes estimados como a su significación.

Resulta difícil establecer un análisis comparativo entre los resultados ya que, aunque desde un punto de vista asintótico serían preferibles los métodos de estimación de sistemas (MC3E, MVIC, MGM), como contrapartida éstos también resultarían más sensibles ante cualquier error de especificación.

d) El concepto de *causalidad de Granger* (1969) hace referencia a la capacidad de una variable para ayudar en la predicción de otra. Así, se dice que X causa Y en el sentido de Granger

cuando los valores retardados de X resultan significativos para explicar los valores actuales de Y.

Para responder a los interrogantes sobre *Causalidad de Granger* con el programa Eviews bastaría con abrir el grupo de variables consideradas y seleccionar la opción *Group Statistic-Granger Causality Test* que -considerando dos retardos- proporciona el resultado:

Pairwise Granger Causality Tests

Sample: 1981 1998

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
Q does not Granger Cause P	16	0.06873	0.93398
P does not Granger Cause Q		2.14962	0.16292
W does not Granger Cause P	16	7.96556	0.00727
P does not Granger Cause W		0.81424	0.46798
V does not Granger Cause P	16	1.74251	0.22009
P does not Granger Cause V		1.60222	0.24509
W does not Granger Cause Q	16	3.61036	0.06231
Q does not Granger Cause W		0.53963	0.59764
V does not Granger Cause Q	16	1.11190	0.36325
Q does not Granger Cause V		0.39899	0.68033
V does not Granger Cause W	16	0.64822	0.54184
W does not Granger Cause V		3.49063	0.06701

Se aprecia que el único resultado claramente significativo para rechazar corresponde al test de la hipótesis “*Las ganancias salariales W no causan en el sentido de Granger los precios industriales P*”. Por tanto las respuestas a las preguntas planteadas serían:

- ✓ La producción industrial no son causa de los salarios
- ✓ Las ganancias salariales son causa de los precios industriales

Puede comprobarse que estos resultados son obtenidos efectuando la regresión mínimo cuadrática de la variable efecto sobre sus propios valores retardados y sobre los valores retardados de la posible causa. Concretamente, si para las variables salarios y precios industriales (W y P) consideramos 2 retardos se tendría:

Dependent Variable: P

Sample(adjusted): 1983 1998

Included observations: 16 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	31.07246	7.263624	4.277817	0.0013
P(-1)	0.932463	0.213793	4.361516	0.0011
P(-2)	-0.417723	0.161524	-2.586131	0.0253
W(-1)	-6.99E-05	0.000250	-0.280126	0.7846
W(-2)	0.000210	0.000257	0.818742	0.4303
R-squared	0.991466	Mean dependent var		101.7675
Adjusted R-squared	0.988363	S.D. dependent var		13.78730
S.E. of regression	1.487279	Akaike info criterion		3.882080
Sum squared resid	24.33198	Schwarz criterion		4.123514
Log likelihood	-26.05664	F-statistic		319.5088
Durbin-Watson stat	1.938438	Prob(F-statistic)		0.000000

y contrastando ahora sobre este modelo la nulidad de los coeficientes de los precios retardados se obtiene:

Wald Test:

Null Hypothesis: $C(4)=0$

$C(5)=0$

F-statistic	7.965563	Probability	0.007265
--------------------	-----------------	--------------------	-----------------

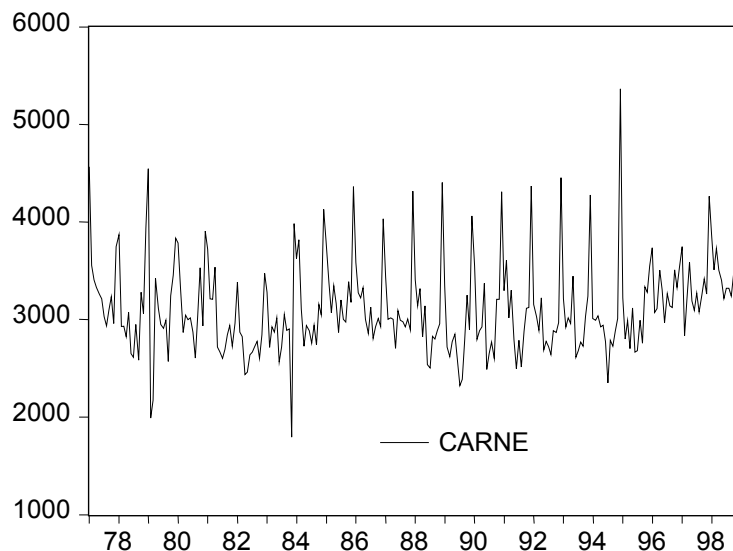
llegando a la conclusión de que los salarios retardados hasta dos períodos presentan una aportación significativa para explicar los precios industriales actuales, esto es, que los salarios causan los precios industriales en el sentido de Granger.

SUPUESTO 6

⇒ Se desea llevar a cabo un análisis cuantitativo de la evolución histórica de la producción de carne sacrificada en Asturias (peso canal en Tm.). La información mensual, elaborada por la consejería de Medio Rural y Pesca del Principado de Asturias desde el año 1977 aparece recogida en el archivo **CARNE.WF1**.

- a) Estudiar el esquema de composición de la serie.
- b) Analizar el impacto de la estacionalidad sobre la producción cárnica regional
- c) Estimar un modelo de alisado adecuado para la serie CARNE en el período 1977.1-1997.12, interpretando los parámetros.
- d) Asumiendo el modelo anterior ¿Cuáles serían las predicciones para la producción de carne hasta el cierre del año 1998? Justificar cómo han sido obtenidas.
- e) Contrastar la estacionariedad de la serie CARNE, proponiendo un modelo ARIMA para la misma en el período muestral 1977.01-1997.12.
- f) A partir del modelo anterior ¿Cuáles serían las predicciones de producción mensual de carne para 1998?.

a) El análisis gráfico de la serie permite apreciar que ésta presenta un patrón estacional:



Para analizar el esquema de composición asociado a esta serie podemos llevar a cabo un estudio de la evolución de su dispersión en relación a las medias anuales. Si dicha dispersión (medida a través de la desviación típica o del rango) se mantiene estable concluiremos que el esquema es

aditivo mientras que patrones de dispersión variables (crecientes o decrecientes) detectarían una hipótesis de tipo multiplicativo.

Así pues, si planteamos un modelo lineal para las desviaciones típicas respecto a las medias anuales: $S_{Y_t} = \beta_0 + \beta_1 \bar{Y}_t + u_t$ el esquema aditivo se corresponde con la hipótesis $H_0 : \beta_1 = 0$ cuyo contraste ha sido implementado en el programa *Hipótesis.prg* descrito en el Anexo A3.

La aplicación de dicho programa con nuestra información muestral proporciona un nivel crítico de 0.0118 que resulta significativo al 5% para rechazar el supuesto de hipótesis aditiva. Así pues, asumiremos como esquema de composición el multiplicativo.

- b) El impacto de la estacionalidad puede ser analizado mediante la opción *Seasonal Adjustment* que permite estimar los componentes estacionales (hipótesis aditiva) y los índices de variación estacional (hipótesis multiplicativa).

En nuestro caso, una vez asumida la hipótesis de composición multiplicativa, se llegaría a los resultados que siguen:

Scaling Factors:	
1	1.154134
2	0.988512
3	0.975946
4	0.998947
5	0.979516
6	0.901363
7	0.907456
8	0.922633
9	0.927655
10	1.017559
11	0.975921
12	1.320284

observándose que la estacionalidad actúa especialmente al alza en el mes de diciembre (mes para el que se estima un 32% de aumento medio sobre el nivel de la serie) y a la baja en el mes de junio, con una caída de producción cárnica en torno al 9.9%²¹.

Otra opción para analizar la estacionalidad consistiría en plantear un modelo explicativo de la producción cárnica en función de una tendencia determinista y de variables dummy asociadas a los componentes estacionales (en principio y dado que las categorías investigadas son los 12 meses, introduciríamos 11 variables cualitativas estacionales, una menos que las categorías, para evitar así la denominada *trampa de la multicolinealidad*).

Las variables cualitativas son generadas mediante el comando @SEAS, designando cada variable con el nombre del correspondiente mes y excluyendo la variable asociada a Enero, ya que este mes se adopta como referencia. El modelo estimado sería en este caso:

²¹ De modo similar, puede comprobarse que si asumiésemos la hipótesis aditiva, el componente estacional estima aumentos de 963.66 Tm para el mes de diciembre y caídas en torno a 322.25 Tm para junio.

Method: Least Squares
 Sample: 1977:01 1998:12
 Included observations: 264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3519.606	74.18623	47.44285	0.0000
T	0.606419	0.253443	2.392718	0.0175
FEBRERO	-530.5352	94.52624	-5.612570	0.0000
MARZO	-572.5838	94.52726	-6.057340	0.0000
ABRIL	-504.7639	94.52896	-5.339781	0.0000
MAYO	-568.5618	94.53134	-6.014532	0.0000
JUNIO	-799.1739	94.53440	-8.453790	0.0000
JULIO	-781.4703	94.53814	-8.266191	0.0000
AGOSTO	-742.7482	94.54255	-7.856231	0.0000
SETIEMBRE	-732.5163	94.54765	-7.747589	0.0000
OCTUBRE	-454.8653	94.55342	-4.810670	0.0000
NOVIEMBRE	-567.5607	94.55988	-6.002130	0.0000
DICIEMBRE	456.4364	94.56701	4.826592	0.0000

R-squared	0.570354	Mean dependent var	3116.761
Adjusted R-squared	0.549813	S.D. dependent var	467.2516
S.E. of regression	313.5070	Akaike info criterion	14.38151
Sum squared resid	24669940	Schwarz criterion	14.55760
Log likelihood	-1885.359	F-statistic	27.76683
Durbin-Watson stat	1.519600	Prob(F-statistic)	0.000000

Estos resultados son coherentes con los anteriores, ya que se observa un impacto positivo para el mes de diciembre (que se cifra en un aumento de 456.4 Tm con respecto al mes de enero) mientras las principales caídas de producción tienen lugar hacia la mitad de año (meses de junio y julio).

- c) A la vista de los resultados anteriores, parece indicada la utilización del procedimiento de alisado Holt-Winters bajo el supuesto de estacionalidad multiplicativa, opción que aplicada al período muestral considerado proporciona los resultados:

Sample: 1977:01 1997:12
 Method: Holt-Winters Multiplicative Seasonal
 Original Series: CARNE
 Parameters: Alpha 0.1300
 Beta 0.0000
 Gamma 0.0000
 Sum of Squared Residuals 21085379
 Root Mean Squared Error 289.2614
 Included observations: 252
 Forecast Series: CARNEHWM

End of Period Levels:	Mean	3349.031
	Trend	-0.042546
	Seasonals:	
	1997:01	1.156538
	1997:02	0.983253
	1997:03	0.966381
	1997:04	0.991510
	1997:05	0.972853
	1997:06	0.897599
	1997:07	0.901329
	1997:08	0.916409
	1997:09	0.921155
	1997:10	1.011112
	1997:11	0.969418
	1997:12	1.312442

Si denotamos por Y_t el valor observado en un período t , siendo T_t , b_t , IS_t los componentes de nivel, pendiente y estacionalidad respectivamente, el método de alisado de Holt-Winters proporciona – bajo un esquema de composición multiplicativo- estimaciones del tipo: $\hat{Y}_t = [\hat{T}_t + \hat{b}_t]I\hat{S}_t$.

La actualización de los distintos componentes de la serie se lleva a cabo según las expresiones.

$$\hat{T}_t = \alpha \frac{Y_t}{I\hat{S}_{t-12}} + (1 - \alpha)(\hat{T}_{t-1} + \hat{b}_{t-1})$$

$$\hat{b}_t = \beta(\hat{T}_t - \hat{T}_{t-1}) + (1 - \beta)(\hat{b}_{t-1})$$

$$I\hat{S}_t = \gamma \frac{Y_t}{\hat{T}_t} + (1 - \gamma)I\hat{S}_{t-12}$$

Con la información muestral disponible, la estimación del parámetro α (que indica las correcciones necesarias en la tendencia) proporciona un resultado 0.13, situándose el nivel final de este componente en torno a 3349 Tm. Por su parte, el valor estimado para el parámetro β es nulo, indicando que la pendiente se mantiene prácticamente estable, situándose este componente en valores de -0.042 al final del período muestral.

También se mantiene estable el patrón de la estacionalidad, como muestra el valor nulo estimado para el parámetro γ , pudiendo apreciarse además que los factores estacionales estimados a final del período coinciden aproximadamente con los índices de variación estacional estimados en el apartado anterior.

- d) Si partiendo del modelo de Holt-Winters llevamos a cabo predicciones para la producción de carne hasta el cierre del año 1999, se obtendrían los resultados siguientes:

Meses	CARNEHWM
1998:01	3873.231
1998:02	3292.861
1998:03	3236.317
1998:04	3320.429
1998:05	3257.909
1998:06	3005.858
1998:07	3018.310
1998:08	3068.769
1998:09	3084.623
1998:10	3385.816
1998:11	3246.158
1998:12	4394.739

a los que se llega mediante las expresiones: $\hat{Y}_t(h) = [\hat{T}_t + \hat{b}_t(h)]I\hat{S}_{t+h-12}$

- e) La estacionariedad de la serie puede ser estudiada a través del test Aumentado de Dickey-Fuller (ADF), cuya hipótesis nula es la presencia de raíces unitarias (no estacionariedad) frente a la alternativa de estacionariedad.

Para llevar a cabo el contraste ADF²² se parte de un modelo AR(1): $Y_t=C+\phi_1Y_{t-1}+u_t$, a partir del cual se obtiene: $\Delta Y_t=C+\beta_1Y_{t-1}+u_t$ con $\beta_1=\phi_1-1$ contrastando la hipótesis de raíz unitaria o no estacionariedad ($H_0: \beta_1=0$) frente a la alternativa de estacionariedad ($H_1: \beta_1<0$).

En términos generales, el test ADF consiste en llevar a cabo una regresión de la primera diferencia de la serie considerada (ΔY_t) sobre la serie con un retardo Y_{t-1} , diferencias de las series con retardos sucesivos ($\Delta Y_{t-1}, \Delta Y_{t-2}, \dots$), una constante y una tendencia temporal. En todos los casos el mecanismo de contraste se basa en el coeficiente de Y_{t-1} que da lugar al

estadístico $t = \frac{\hat{\beta} - \beta}{S_{\hat{\beta}}}$, expresión que – pese a su denominación- no sigue una distribución t de

Student, al no ser la serie estacionaria bajo H_0 .

El procedimiento descrito nos llevará a la conclusión de que la serie es integrada para resultados de t bajos en valor absoluto, concluyendo en otro caso que la serie es estacionaria.

En nuestro caso, la opción ADF del programa Eviews aplicada sobre la serie CARNE proporciona el resultado siguiente:

ADF Test Statistic	-11.82553	1% Critical Value*	-3.4580
		5% Critical Value	-2.8731
		10% Critical Value	-2.5729

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation **Dependent Variable: D(CARNE)**
 Method: Least Squares Sample(adjusted): 1977:02 1997:12
 Included observations: 251 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CARNE(-1)	-0.711761	0.060189	-11.82553	0.0000
C	2199.885	188.1785	11.69042	0.0000
R-squared	0.359639	Mean dependent var		-1.195032
Adjusted R-squared	0.357067	S.D. dependent var		547.1922
S.E. of regression	438.7556	Akaike info criterion		15.01370
Sum squared resid	47934103	Schwarz criterion		15.04179
Log likelihood	-1882.219	F-statistic		139.8432
Durbin-Watson stat	1.973156	Prob(F-statistic)		0.000000

Por tanto la conclusión en este caso es rechazar la hipótesis nula de no estacionariedad, concluyendo que la serie CARNE es estacionaria.

A partir del correlograma muestral podríamos seguir la metodología propuesta por Box y Jenkins (1976) para identificar modelos ARIMA.

²² La versión más sencilla es el contraste propuesto por Dickey y Fuller (1979) que se basa en un modelo AR(1): $Y_t=C+\phi_1Y_{t-1}+u_t$ estableciendo como hipótesis nula la *no estacionariedad* de esta serie:

$H_0: \phi_1=1$ La serie tiene raíces unitarias, es no estacionaria I(1)

$H_1: \phi_1<1$ La serie no tiene raíces unitarias, es estacionaria I(0)

Estas hipótesis permiten expresar la serie en los siguientes términos:

Bajo H_0 la serie es ARI(0,1): $(1-L)Y_t=\mu^* + u_t$

Bajo H_1 la serie es ARI(1,0): $(1-\phi_1L)Y_t=(1-\phi_1L)\mu+u_t$ o bien $(1-\phi_1L)(Y_t-\mu)=u_t$

Entre las posibles especificaciones alternativas para esta serie proponemos un modelo con términos regular y estacional ARIMA(1,0,0)_xARIMA_s(0,0,2) sobre el que posteriormente hemos incorporado variables de intervención asociadas a los *outliers*²³. La estimación proporciona el resultado:

Dependent Variable: CARNE		Method: Least Squares		
Sample(adjusted): 1977:02 1997:12				
Included observations: 251 after adjusting endpoints				
Convergence achieved after 11 iterations				
Backcast: 1975:02 1977:01				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3093.909	57.22403	54.06662	0.0000
D9412	1927.894	221.4657	8.705157	0.0000
D8311	-1344.641	214.7921	-6.260195	0.0000
D7902	-1331.708	270.6417	-4.920559	0.0000
AR(1)	0.298461	0.063025	4.735576	0.0000
MA(12)	0.573268	0.044001	13.02845	0.0000
MA(24)	0.585414	0.041711	14.03494	0.0000
R-squared	0.600753	Mean dependent var	3091.247	
Adjusted R-squared	0.590935	S.D. dependent var	457.5982	
S.E. of regression	292.6714	Akaike info criterion	14.22347	
Sum squared resid	20900201	Schwarz criterion	14.32179	
Log likelihood	-1778.046	F-statistic	61.19170	
Durbin-Watson stat	2.003438	Prob(F-statistic)	0.000000	
Inverted AR Roots	.30			
Inverted MA Roots	.96 -.16i	.96+.16i	.92+.35i	.92 -.35i
	.76 -.62i	.76+.62i	.62 -.76i	.62+.76i
	.35 -.92i	.35+.92i	.16+.96i	.16 -.96i
	-.16+.96i	-.16 -.96i	-.35+.92i	-.35 -.92i
	-.62+.76i	-.62 -.76i	-.76+.62i	-.76 -.62i
	-.92+.35i	-.92 -.35i	-.96 -.16i	-.96+.16i

que puede ser formulado²⁴ como:

$$\hat{Y}_t = \hat{C} + \hat{\phi}_1 Y_{t-1} + \hat{u}_t - \hat{\theta}_{s1} \hat{u}_{t-12} - \hat{\theta}_{s2} \hat{u}_{t-24} + \hat{\gamma}_1 D_t^{9412} + \hat{\gamma}_2 D_t^{8811} + \hat{\gamma}_3 D_t^{7902} =$$

$$= 2170.5 + 0.298 Y_{t-1} + \hat{u}_t + 0.573 \hat{u}_{t-12} + 0.585 \hat{u}_{t-24} + 1927.89 D_t^{9412} - 1344.6 D_t^{8811} - 1331.7 D_t^{7902}$$

Una vez estimado el modelo, el análisis de sus residuos (gráficos, correlograma, test ADF...) sugiere estructura de ruido blanco, validando así nuestra propuesta.

- f) A partir del modelo estimado anteriormente se obtienen las predicciones mensuales para 1998 junto con las correspondientes desviaciones estándar y medidas de bondad de la capacidad predictiva.

Más concretamente las predicciones puntuales proporcionadas por el comando *Forecast* con la opción dinámica (CARNEF) son obtenidas mediante las expresiones:

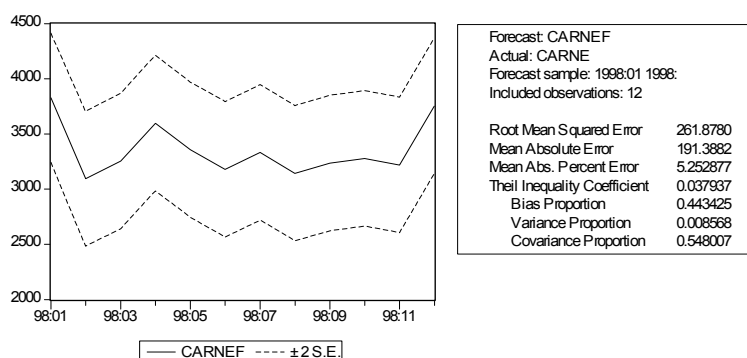
²³ Los valores atípicos detectados corresponden a los períodos 1994.12, 1983.11 y 1979.02 por lo que se han generado las variables *dummy* D9412, D8311 y D7902 asociadas a los mismos.

²⁴ Debemos tener presente que las estimaciones proporcionadas por Eviews corresponden al valor esperado $\hat{\mu} = \frac{\hat{C}}{(1-\hat{\phi}_1)}$ y los valores opuestos de los parámetros MA: $-\hat{\theta}_{s1}, -\hat{\theta}_{s2}$.

$$\begin{aligned} \hat{Y}_{t+1} &= \hat{C} + \hat{\phi}_1 Y_t - \hat{\Theta}_{s1} \hat{u}_{t-11} - \hat{\Theta}_{s2} \hat{u}_{t-23} \\ \hat{Y}_{t+2} &= \hat{C} + \hat{\phi}_1 \hat{Y}_{t+1} - \hat{\Theta}_{s1} \hat{u}_{t-10} - \hat{\Theta}_{s2} \hat{u}_{t-22} \\ &\dots \\ \hat{Y}_{t+12} &= \hat{C} + \hat{\phi}_1 \hat{Y}_{t+11} - \hat{\Theta}_{s1} \hat{u}_t - \hat{\Theta}_{s2} \hat{u}_{t-12} \end{aligned}$$

donde debemos tener en cuenta que en el horizonte de predicción tanto los errores esperados como las variables dummy asociadas a los *outliers* adoptan valores nulos, y los parámetros estimados presentan valores: $\hat{C} = 2170.5, \hat{\phi}_1 = 0.298, \hat{\Theta}_{s1} = -0.573, \hat{\Theta}_{s2} = -0.585$.

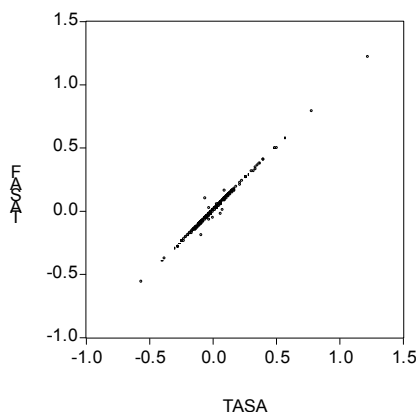
Por lo que se refiere a la capacidad predictiva del modelo, la raíz del error cuadrático



medio asciende en este caso a 261.87, resultado inferior al asociado a las predicciones de Holt-Winters²⁵ (272.37).

El índice de Theil por su parte adopta un valor reducido (0.037), con un sesgo moderado (44%) y destacando el peso del componente de covarianza (54.8%).

Podría ser interesante analizar también el diagrama realización-predicción sobre tasas de variación mensual, generadas mediante las sentencias `GENR TASA=@PCH(CARNE)` y `GENR TASAF=@PCH(CARNEF)` y cuya representación gráfica resulta cercana a la diagonal de predicciones perfectas:



²⁵ La opción de Holt-Winters del programa Eviews (salida de la página 37) proporciona únicamente la raíz del error cuadrático medio asociado al período de estimación (289.26 en este caso). No obstante, podemos cuantificar esta medida mediante la expresión `=@SQR(@SUMSQ(CARNE-CARNEHWM)/12)` que para el horizonte de predicción (recorrido 1998.01 1998.12) proporciona el resultado 272.35.

SUPUESTO 7

⇒ Partiendo de ciertos procesos generadores de datos, se simulan muestras aleatorias por el método de Montecarlo, estudiando a partir de ellas las correspondientes características muestrales y llevando a cabo la identificación de un modelo adecuado²⁶.

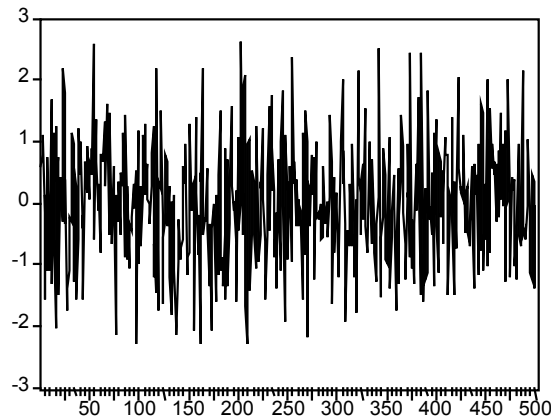
El punto de arranque será la simulación por el método de Montecarlo de 500 observaciones de una serie temporal con estructura de *ruido blanco gaussiano* $e_t \approx N(0,1)$. A partir del mismo se generan las siguientes series:

Serie	Especificación
Y1	$Y_t = 2 + e_t + 0.8 e_{t-1}$
Y2	$Y_t = 80 + u_t + 0.6u_{t-1} - 0.3u_{t-2}$ con $u_t \approx N(0, \sigma^2=16)$
Y3	$Y_t = 60 + 0.5Y_{t-1} + u_t$ con $Y_1=100$
Y4	$Y_t = 60 + Y_{t-1} + u_t$ con $Y_1=100$
Y5	$Y_t = 30 + 0.5t + u_t$
Y6	$Y_t = 75 - 0.2Y_{t-1} + 0.5Y_{t-2} + u_t$, con $Y_1=80, Y_2=60$
Y7	$Y_t = 75 - 0.8Y_{t-1} + 0.5Y_{t-2} + u_t$, con $Y_1=80, Y_2=60$

Para cada una de las series anteriores, efectuar un análisis teórico (estudio del modelo, sus características poblacionales, su estacionariedad e invertibilidad) y llevar a cabo su estimación a partir de la información muestral generada.

²⁶ El archivo SIMULA.WF1 contiene los resultados obtenidos mediante simulación de las series propuestas en este supuesto.

Para llevar a cabo estas simulaciones, comenzamos definiendo un archivo de trabajo en Eviews con rango para 500 observaciones, construyendo a continuación una serie $e_t \approx N(0,1)$ mediante el comando de generación de números aleatorios normales (GENR $e=@NRND$). Las distintas vistas de la serie permiten apreciar su carácter estacionario.



y el correlograma incluye la salida del test de Ljung-Box, que permite asumir la no existencia de correlaciones:

Sample: 1 500		Included observations: 500					
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob		
. .	.	1	-0.044	-0.044	0.9673	0.325	
. .	.	2	-0.029	-0.031	1.3872	0.500	
. .	.	3	-0.027	-0.029	1.7442	0.627	
. .	.	4	0.026	0.023	2.0971	0.718	
. .	.	5	0.019	0.019	2.2710	0.811	
. .	.	6	0.016	0.019	2.4047	0.879	
. .	.	7	0.050	0.054	3.6811	0.816	
. .	.	8	-0.011	-0.005	3.7425	0.880	
. .	.	9	-0.031	-0.029	4.2440	0.895	
. *	*	10	0.097	0.096	9.0985	0.523	
. .	.	11	0.026	0.030	9.4562	0.580	
. .	.	12	-0.011	-0.007	9.5223	0.658	
. .	.	13	0.009	0.016	9.5643	0.729	
. .	.	14	0.024	0.020	9.8526	0.773	
. .	.	15	0.048	0.048	11.056	0.749	
. .	.	16	-0.044	-0.039	12.059	0.740	
. .	.	17	0.011	0.000	12.124	0.793	
. .	.	18	0.056	0.055	13.751	0.745	
. .	.	19	0.041	0.049	14.648	0.745	
. .	.	20	-0.046	-0.050	15.765	0.731	

El contraste de la hipótesis de no autocorrelación $H_0: \rho_k=0 \forall k$ (necesaria para que la serie sea ruido blanco) frente a la alternativa $H_1: \exists k, \rho_k \neq 0$ se lleva a cabo mediante el estadístico Ljung-Box dado por la expresión: $Q = T(T+2) \sum_{s=1}^p \frac{r_s^2}{T-s}$ que bajo la hipótesis nula sigue modelo chi-cuadrado con tantos grados de libertad como autocorrelaciones consideradas (p).

Como puede apreciarse, para la serie considerada se obtienen -tal y como cabía esperar- niveles críticos suficientemente altos.

Por lo que se refiere a las características muestrales, una síntesis de las mismas junto con la representación gráfica de la distribución es proporcionada por la opción *Views-Histogram*, obteniéndose en este caso una media próxima a cero con dispersión cercana a la unidad.

⇒ La **SERIE Y1** propuesta ($Y_t=2+e_t+0.8e_{t-1}$) se corresponde con un modelo MA(1) de parámetro $\theta_1=-0.8$, siendo por tanto sus características teóricas:

$$E(Y1)=\mu=2,$$

$$\text{Var}(Y1)=\sigma^2(1+\theta_1^2)=1,64 \quad \text{Std}(Y1)=1,28$$

$$\text{FAC: } \rho_1 = -\frac{\theta_1}{(1+\theta_1^2)} = 0.49 \text{ siendo las restantes correlaciones nulas}$$

Teniendo en cuenta que el modelo considerado es MA, su estacionariedad queda garantizada siendo necesario analizar únicamente las condiciones de invertibilidad.

Partiendo de la expresión: $Y_t=2+e_t(1+0.8L)$ será necesario estudiar las raíces del polinomio MA, que en este caso proporcionan el resultado $-\frac{1}{0.8} = -1.25 -1/0.8=-1.25$ y, al superar en valor absoluto a la unidad, permiten concluir que el proceso es invertible.

Por lo que se refiere a la simulación, el modelo propuesto se genera en función del ruido blanco anterior e mediante la sentencia: `GENR Y1=2+e+0.8*e(-1)`. La estimación muestral de sus características puede obtenerse mediante las opciones de histograma y correlograma, mientras que la estimación del modelo MA(1) conduce a un resultado coherente con las características poblacionales investigadas (parámetros C y θ_1 significativamente distintos de 0 y residuos de estimación con estructura de ruido blanco).

⇒ La **SERIE Y2** viene dada por la expresión $Y_t=80+u_t+0.6u_{t-1}-0.3u_{t-2}$ con $u_t \approx N(0, \sigma^2=16)$, es decir, un MA(2) con $C=80$, $\theta_1=-0.6$, $\theta_2=0.3$ cuyas características poblacionales serían:

$$E(Y2)=\mu=80$$

$$\text{Var}(Y2)= (1+\theta_1^2+\theta_2^2)\sigma^2=23.2 \quad \text{Std}(Y2)= 4.82$$

$$\rho_1 = \frac{-\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} = 0.29$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} = -0.21$$

$$\rho_j = 0 \quad \forall j \neq 1,2$$

Resolviendo el polinomio MA asociado a esta serie $(1+0.6L-0.3L^2)=0$ se obtienen raíces reales 3.08 y -1.08 ambas fuera del círculo de raíz unitaria por lo cual se garantiza su invertibilidad.

La serie muestral se genera partiendo de una perturbación aleatoria normal homoscedástica con $\sigma^2=16$, creada como $U=4*\text{@NRND}$. El proceso MA(2) considerado viene entonces dado por la expresión: $Y_2=80+U+0.6*U(-1)-0.3*U(-2)$, observándose el correlograma muestral:

Sample: 1 500

Included observations: 498

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
. **	. **	1 0.294	0.294	43.311	0.000
** .	*** .	2 -0.218	-0.333	67.189	0.000
. .	. **	3 -0.023	0.197	67.466	0.000
. .	* .	4 0.013	-0.159	67.550	0.000
. .	. *	5 0.035	0.162	68.167	0.000
. .	* .	6 0.014	-0.119	68.271	0.000
. .	. *	7 -0.015	0.093	68.381	0.000
. .	* .	8 0.007	-0.062	68.408	0.000
. .	. *	9 0.020	0.071	68.614	0.000
. .	* .	10 -0.002	-0.062	68.616	0.000

que permite identificar un modelo MA(2) cuya estimación proporciona los resultados esperados (coeficientes C, $-\theta_1$ y $-\theta_2$ significativos, e inversas de las dos raíces inferiores a la unidad en valor absoluto).

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	80.24252	0.225603	355.6796	0.0000
MA(1)	0.600761	0.042942	13.98998	0.0000
MA(2)	-0.300907	0.043156	-6.972606	0.0000
Inverted MA Roots	.33	-.93		

⇒ La **SERIE Y3** se corresponde con un modelo autorregresivo AR(1) dado por la expresión $Y_t=60+0.5Y_{t-1}+u_t$ que lleva asociadas las características poblacionales:

$$E(Y_3) = \frac{C}{1 - \phi_1} = 120$$

$$\text{Var}(Y_3) = \frac{\sigma^2}{1 - \phi_1^2} = 21.33 \Rightarrow \text{Std}(Y_3) = 4.618$$

$$\rho_1 = \phi_1^1 = 0.5 \quad \rho_2 = \phi_1^2 = 0.25 \quad \rho_3 = \phi_1^3 = 0.125$$

$$\phi_{11} = 0.5 \quad \phi_{jj} = 0 \quad \forall j \neq 1$$

Al tratarse de un modelo AR queda garantizada su invertibilidad, y la raíz del polinomio autorregresivo $(1-0.5L)$ es $2>1$ por lo cual la serie resulta estacionaria.

Desde la óptica muestral, una vez introducido el valor inicial $Y_3=100$, la serie se genera mediante la expresión $Y_3=60+0.5*Y_3(-1)+U$.

Su correlograma proporciona resultados del tipo:

Sample: 2 500

Included observations: 499

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. ****	. ****	1	0.500	0.500	125.32	0.000
. **	. .	2	0.239	-0.014	154.10	0.000
. *	. .	3	0.104	-0.013	159.60	0.000
. *	. .	4	0.079	0.046	162.73	0.000
. *	. .	5	0.067	0.018	164.97	0.000
. .	. .	6	0.048	0.001	166.14	0.000
. .	. .	7	0.019	-0.014	166.33	0.000
. .	. .	8	0.021	0.019	166.55	0.000
. .	. .	9	0.018	0.001	166.71	0.000
. .	. .	10	-0.011	-0.033	166.76	0.000

y la estimación del modelo (regresión LS Y5 C AR(1)) proporciona aproximaciones de la esperanza μ y del parámetro $\phi_1=0.5$ que coincide con la inversa de la raíz.

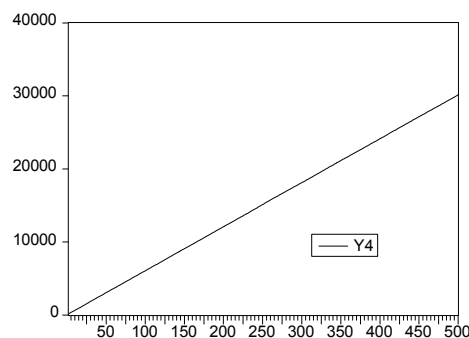
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	120.3777	0.350464	343.4808	0.0000
AR(1)	0.506665	0.037641	13.46059	0.0000
Inverted AR Roots	.51			

Además, el análisis gráfico y el correlograma de los residuos de esta regresión permiten apreciar que éstos presentan estructura de ruidos blancos.

⇒ La **SERIE Y4** viene dada por la expresión $Y_t=60+Y_{t-1} +u_t$, siendo por tanto similar a la anterior en cuanto a su estructura AR(1) con la salvedad de que Y4 presenta una raíz unitaria.

Como consecuencia, la serie Y4 no será estacionaria, siendo imposible obtener un valor esperado para la misma. Nos encontraríamos ante una serie con tendencia estocástica (*paseo aleatorio con deriva*) para la cual resulta conveniente efectuar una diferenciación, obteniendo así $Z_t=Y_t(1-L)=60+u_t$.

La simulación de la serie Y4 proporciona una representación gráfica en la que se aprecia claramente su carácter no estacionario:



observándose también un correlograma no convergente:

Sample: 2 500
Included observations: 499

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *****	. *****	1	0.994	0.994	496.00	0.000
. *****	. .	2	0.988	-0.003	987.00	0.000
. *****	. .	3	0.982	-0.003	1473.0	0.000
. *****	. .	4	0.976	-0.002	1954.1	0.000
. *****	. .	5	0.970	-0.003	2430.3	0.000
. *****	. .	6	0.964	-0.003	2901.5	0.000
. *****	. .	7	0.958	-0.003	3367.8	0.000
. *****	. .	8	0.952	-0.003	3829.3	0.000
. *****	. .	9	0.946	-0.003	4285.8	0.000
. *****	. .	10	0.940	-0.003	4737.5	0.000

y quedando la no estacionariedad confirmada mediante la aplicación del test ADF, que no permite en este caso rechazar la presencia de una raíz unitaria:

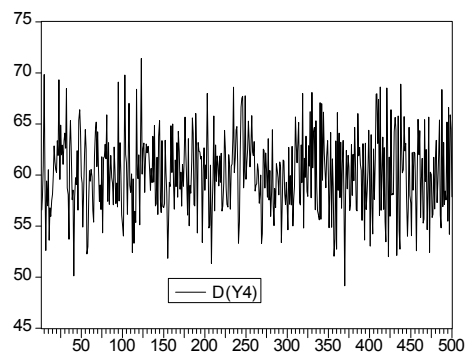
ADF Test Statistic	-2.708087	1% Critical Value*	-3.9807
		5% Critical Value	-3.4208
		10% Critical Value	-3.1328

*MacKinnon critical values for rejection of hypothesis of a unit root.

Augmented Dickey-Fuller Test Equation **Dependent Variable: D(Y4)**
Sample: 2 500 Included observations: 499

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y4(-1)	-0.030992	0.011444	-2.708087	0.0070
C	63.38747	1.186862	53.40762	0.0000
@TREND(2)	1.866225	0.689313	2.707371	0.0070

La aplicación del mismo test sobre la serie con una diferencia²⁷ conduce a un resultado significativo para rechazar la presencia de raíz unitaria. Así pues, generaríamos la serie en primeras diferencias D(Y4) que sí muestra comportamiento estacionario²⁸.



²⁷ Más concretamente, el test ADF proporciona en este caso el resultado:

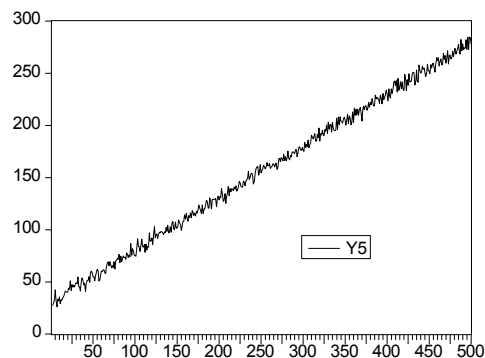
ADF Test Statistic	-22.20708	1% Critical Value*	-3.4458
		5% Critical Value	-2.8676
		10% Critical Value	-2.5700

²⁸ La estimación directa de la serie Y4 como AR(1) no sería adecuada, obteniéndose una raíz unitaria, síntoma de un proceso no estacionario. Para la serie en primeras diferencias D(Y4) resultaría únicamente significativa la constante, para la que se estima un valor cercano a 60.

⇒ La **SERIE Y5** presenta tendencia determinista, al venir dada por la expresión $Y_t=30+0.5t+u_t$. Si bien esta serie tampoco es estacionaria, el tratamiento adecuado sería ahora distinto al caso anterior, ya que al diferenciar se obtendría $Y_t(1-L)=0.5+u_t(1-L)$, es decir, un modelo MA(1) no invertible.

Así pues, en caso de presentar tendencia determinista, la modelización de la serie debería efectuarse incorporando dicha tendencia.

La generación de la serie Y5 se llevaría a cabo mediante la orden GENR $Y5=30+0.5*T+U$ donde T ha sido generada como tendencia lineal:



estimándose para la misma un modelo determinista del tipo:

Dependent Variable: Y5

Method: Least Squares

Sample: 1 500

Included observations: 500

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	30.36190	0.346428	87.64277	0.0000
T	0.499358	0.001198	416.7345	0.0000
R-squared	0.997141	Mean dependent var	155.4511	
Adjusted R-squared	0.997135	S.D. dependent var	72.25155	
S.E. of regression	3.867374	Akaike info criterion	5.547020	
Sum squared resid	7448.376	Schwarz criterion	5.563879	
Log likelihood	-1384.755	F-statistic	173667.7	
Durbin-Watson stat	1.994380	Prob(F-statistic)	0.000000	

⇒ La **SERIE Y6** $Y_t=75-0.2Y_{t-1}+0.5Y_{t-2}+u_t$ se corresponde con un modelo AR(2) de parámetros $\phi_1 = -0.2, \phi_2 = 0.5$ cuya estacionariedad queda garantizada por verificar sus parámetros las

condiciones $\phi_1+\phi_2<1, \phi_2-\phi_1<1$, siendo su valor esperado: $E(Y6) = \frac{C}{1 - \phi_1 - \phi_2} = 107.14$.

El carácter estacionario queda también garantizado mediante el estudio de las raíces del polinomio autorregresivo: $1+0.2L-0.5L^2$ que en este caso conducen a los resultados -1.23 y 1.63 , ambos superiores en valor absoluto a la unidad.

La generación de la serie anterior conduce a características muestrales coherentes con la descripción efectuada anteriormente. Así, la estimación de un modelo AR(2) proporciona el resultado:

Dependent Variable: Y6

Sample: 3 500

Included observations: 498

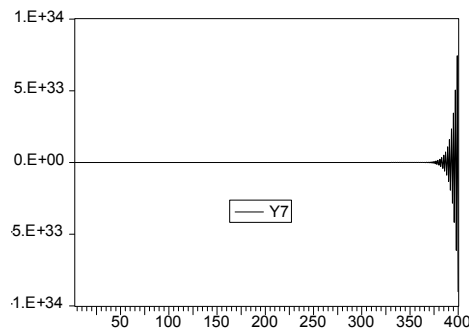
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	107.4224	0.248076	433.0231	0.0000
AR(1)	-0.202295	0.032581	-6.208972	0.0000
AR(2)	0.502693	0.031823	15.79668	0.0000
R-squared	0.413509	Mean dependent var		107.3445
Adjusted R-squared	0.411139	S.D. dependent var		5.037934
S.E. of regression	3.865972	Akaike info criterion		5.548309
Sum squared resid	7398.141	Schwarz criterion		5.573674
Log likelihood	-1378.529	F-statistic		174.5014
Durbin-Watson stat	1.989110	Prob(F-statistic)		0.000000
Inverted AR Roots	.62	-.82		

⇒ Por último, la **SERIE Y7** $Y_t = 75 - 0.8Y_{t-1} + 0.5Y_{t-2} + u_t$ cuya estructura AR(2) es en principio similar al modelo anterior, muestra sin embargo un carácter no estacionario ya que en este caso los parámetros adoptan valores $\phi_1 = -0.8, \phi_2 = 0.5$ para los que se cumple $\phi_1 + \phi_2 < 1$ pero sin embargo se observa $\phi_2 - \phi_1 > 1$.

De modo similar, las raíces del polinomio autorregresivo $1 + 0.8L - 0.5L^2$ proporcionan en este caso los resultados 2.42 y -0.82 , este último inferior en valor absoluto a la unidad.

Simulando una serie de la tipología descrita (GENR $Y7 = 75 - 0.8 * Y7(-1) + 0.5 * Y7(-2) + U$, con valores iniciales 80 y 60) la no estacionariedad se aprecia claramente en su gráfico explosivo:



y no resulta posible estimar un modelo AR(2) debido a la no invertibilidad de la matriz, justificada por la alta correlación existente entre las observaciones en gran parte del recorrido muestral.

SUPUESTO 8

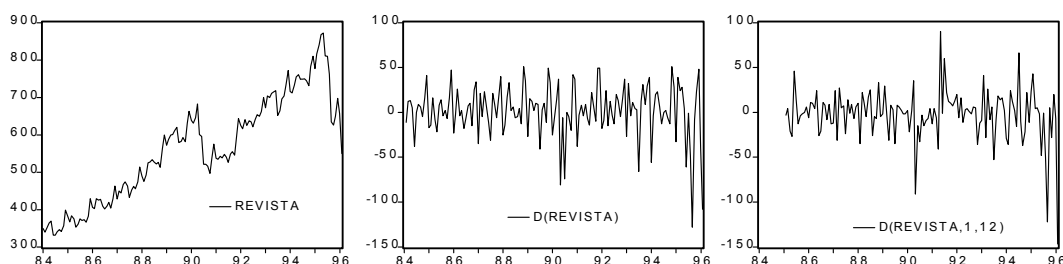
⇒ Una revista juvenil publicada mensualmente se distribuye a través de centros educativos y asociaciones culturales, de forma gratuita hasta Marzo de 1996 y al precio de 50 pesetas a partir de esa fecha. Los datos relativos a los ejemplares distribuidos mensualmente aparecen recogidos en el archivo **REVISTA.WF1**.

Proponer una modelización univariante para la serie en el período comprendido hasta 1997, estimando el efecto que ha tenido el precio sobre la distribución de ejemplares, y efectuar predicciones referidas a los años 1998 y 1999.

Como puede apreciarse, la serie de ejemplares distribuidos mensualmente presenta un “escalón” -en este caso decreciente- en el mes de Marzo del año 1996. Para estimar dicho efecto y elaborar un modelo univariante adecuado para la serie REVISTA analizaremos inicialmente su comportamiento en el período muestral previo al cambio.

El análisis de estacionariedad para dicho recorrido permite apreciar que es necesario diferenciar la serie, ya que la aplicación del test ADF a la serie original conduce a resultados no significativos para rechazar la presencia de raíces unitarias, mientras que el mismo contraste para la serie en primeras diferencias llevaría a concluir que ésta es estacionaria y lo mismo sucedería para la serie con una diferencia regular y otra estacional.

ADF Test Statistic	-1.715790	1% Critical Value*	-3.4764
Serie REVISTA		5% Critical Value	-2.8814
		10% Critical Value	-2.5772
ADF Test Statistic	-10.91623	1% Critical Value*	-3.4767
Serie D(REVISTA)		5% Critical Value	-2.8815
		10% Critical Value	-2.5773
ADF Test Statistic	-6.448865	1% Critical Value*	-2.5812
Serie D(REVISTA,1,12)		5% Critical Value	-1.9423
		10% Critical Value	-1.6170



Entre las diferentes especificaciones posibles, hemos optado por los dos modelos siguientes:

$$(1-\phi_{12}L^{12})(1-L)(1-L^{12})Y_t=u_t \quad \text{ARIMA}(0,1,0) \times \text{ARIMA}_S(1,1,0)$$

$$(1-L)(1-L^{12})Y_t=(1-\theta_{12}L^{12})u_t \quad \text{ARIMA}(0,1,0) \times \text{ARIMA}_S(0,1,1)$$

donde Y_t denota la serie REVISTA.

Las estimaciones de ambos modelos proporcionan los resultados:

Dependent Variable: D(REVISTA,1,12)

Sample(adjusted): 1986:02 1996:02

Included observations: 121 after adjusting endpoints

Convergence achieved after 2 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SAR(12)	-0.484287	0.106387	-4.552122	0.0000
R-squared	0.141456	Mean dependent var	-2.413223	
Adjusted R-squared	0.141456	S.D. dependent var	29.38896	
S.E. of regression	27.23112	Akaike info criterion	9.454828	
Sum squared resid	88984.07	Schwarz criterion	9.477934	
Log likelihood	-571.0171	Durbin-Watson stat	1.831140	
Inverted AR Roots	.91+.24i	.91 -.24i	.67+.67i	.67 -.67i
	.24 -.91i	.24+.91i	-.24 -.91i	-.24+.91i
	-.67 -.67i	-.67 -.67i	-.91 -.24i	-.91+.24i

Dependent Variable: D(REVISTA,1,12)

Sample(adjusted): 1985:02 1996:02

Included observations: 133 after adjusting endpoints

Convergence achieved after 7 iterations

Backcast: 1984:02 1985:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SMA(12)	-0.844036	0.030924	-27.29372	0.0000
R-squared	0.304270	Mean dependent var	-2.240602	
Adjusted R-squared	0.304270	S.D. dependent var	28.53099	
S.E. of regression	23.79782	Akaike info criterion	9.184556	
Sum squared resid	74756.40	Schwarz criterion	9.206288	
Log likelihood	-609.7730	Durbin-Watson stat	1.811017	
Inverted MA Roots	.99	.85+.49i	.85 -.49i	.49+.85i
	.49 -.85i	-.00 -.99i	-.00+.99i	-.49 -.85i
	-.49+.85i	-.85+.49i	-.85 -.49i	-.99

apreciándose que, desde esta óptica, el segundo modelo resultaría preferible.

Por lo que se refiere a la capacidad predictiva, si bien se esperan resultados desfavorables en ambos casos debido al cambio estructural del año 1996, podemos llevar a cabo un análisis comparativo de los modelos, obteniéndose los resultados sintetizados en la tabla que sigue, según los cuales resultaría preferible el modelo SMA(12):

INDICADOR	MODELO SAR(12)	MODELO SMA(12)
AIC	6.616951	6.346681
EAM (10 m)	384.5459	381.4324
Raíz ECM (10 m)	386.1515	382.0218
EAPM (10 m)	225.8366	223.7742
COEF. THEIL	0.530871	0.527875

Una vez que hemos estimado varios modelos de partida y dado que somos conscientes del cambio estructural que se ha producido en la serie, deberíamos llevar a cabo un análisis de intervención sobre los modelos anteriores, contemplando el impacto de los precios sobre la publicación.

Propondríamos así un modelo del tipo: $Y_t = \frac{(1 - \theta_{12}L^{12})}{(1-L)(1-L^{12})}u_t + I_t$, siendo la variable de

intervención tipo impulso (I_t adoptará nula hasta 1996.02 y valor unitario a partir de entonces).

El modelo puede ser expresado de forma alternativa como:

$$(1-L)(1-L^{12})Y_t = (1-\theta_{12}L^{12})u_t + (1-L)(1-L^{12})I_t$$

y su estimación con el programa Eviews se lleva ahora a cabo para el período muestral hasta 1997 obteniéndose el resultado:

Dependent Variable: D(REVISTA,1,12)
 Sample(adjusted): 1985:02 1997:12
 Included observations: 155 after adjusting endpoints
 Convergence achieved after 11 iterations
 Backcast: 1984:02 1985:01

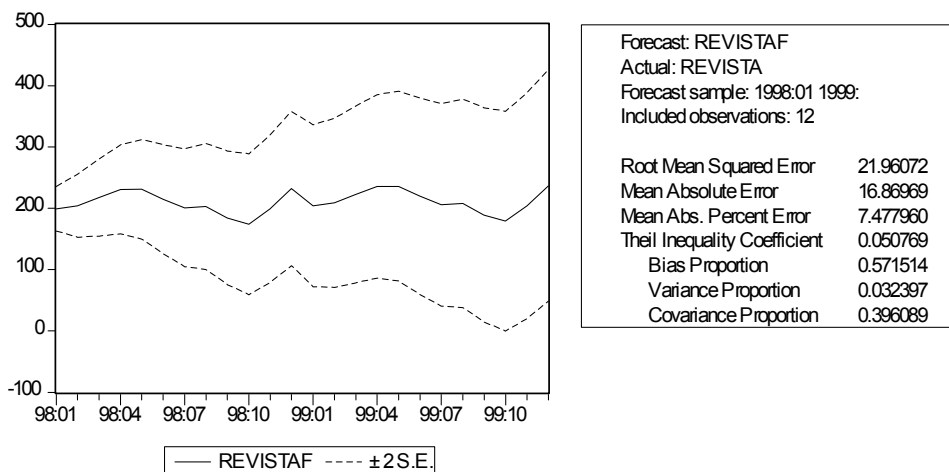
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(IT,1,12)	-400.4512	22.29738	-17.95957	0.0000
SMA(12)	-0.885741	0.000357	-2482.865	0.0000
R-squared	0.837260	Mean dependent var		0.064516
Adjusted R-squared	0.836196	S.D. dependent var		55.44574
S.E. of regression	22.44036	Akaike info criterion		9.072418
Sum squared resid	77046.15	Schwarz criterion		9.111688
Log likelihood	-701.1124	Durbin-Watson stat		1.962403
Inverted MA Roots	.99	.86 -.49i	.86+.49i	.49 -.86i
	.49+.86i	.00+.99i	-.00 -.99i	-.49+.86i
	-.49 -.86i	-.86+.49i	-.86 -.49i	-.99

en el que se aprecia un descenso estimado de 400 ejemplares a partir del marzo del 1996, confirmando así nuestras expectativas.

Por lo que respecta a la etapa de *chequeo* del modelo, el análisis de los residuos asociados revela una estructura estacionaria (conclusión apoyada por el test ADF sobre dichos residuos), no correlación serial (confirmada por el test de Ljung-Box) pero sin embargo se plantean dificultades en cuanto a la normalidad, debido a la presencia de *outliers*.

Examinando el gráfico de valores verdaderos, estimados y residuos, se llegaría a la conclusión de que los valores atípicos más destacados corresponden a 1995.09, 1996.02 y 1990.05, siendo posible mejorar el modelo anterior mediante la introducción de variables cualitativas asociadas a dichas observaciones (de hecho el error estándar, situado en 22.44, se reduciría hasta 18).

Si ahora utilizamos el modelo con intervención y tratamiento de outliers para realizar predicciones sobre la serie REVISTA, asumiendo que no se producen nuevos cambios en la política de precios durante el horizonte de predicción (años 1998 y 1999) se obtendría:



resultados a los que se ha llegado mediante la aplicación del comando *Forecast* con su opción dinámica, que –como puede apreciarse en el gráfico- proporciona no sólo predicciones puntuales²⁹ sino también los correspondientes valores de la desviación estándar del error de predicción: $S(e_T(h)) = \sqrt{(1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2)}\sigma_U$, a partir de los cuales se construyen los correspondientes intervalos de confianza $\hat{Y}_T(h) \pm kS(e_T(h))$, que el programa Eviews representa gráficamente, para un nivel de confianza de aproximadamente el 95%.

Por otra parte, teniendo en cuenta que disponemos de la información correspondiente al año 1998, esta primera parte del horizonte predictivo ha permitido obtener las medidas habituales de la bondad de predicción (raíz del error cuadrático medio, error absoluto medio e índice de Theil).

²⁹ Dado que el modelo ha sido planteado sobre la variable D(REVISTA,1,12) el programa permite obtener predicciones tanto para esta serie (diferenciada regular y estacionalmente) como para la serie original REVISTA, alternativa esta última que hemos seleccionado.

SUPUESTO 9

- Sobre un grupo de 45 empresas se han investigado durante el período 1992-1998 (7 años) los beneficios anuales netos (BENEFICIOS) y las cantidades destinadas anualmente a ayuda al desarrollo (AYUDA), ambas expresadas en dolares corrientes. La información (almacenada inicialmente como series temporales consecutivas de 7 años para cada empresa) ha sido recogida en el archivo **COOPERA.WF1**.
- a) Estimar un modelo único para describir el comportamiento de las ayudas a la cooperación.
 - b) ¿Cambiarían sustancialmente los resultados si se estima un modelo para los Efectos medios?
 - c) Llevar a cabo la estimación de un modelo de panel con Efectos fijos, contrastando la hipótesis de igualdad de los efectos.
 - d) Estimar un modelo de Efectos aleatorios, comparando sus resultados con los del apartado anterior.
 - e) Estudiar la inclusión de términos autorregresivos sobre el modelo de Efectos fijos.

- a) La información que manejamos en este supuesto (series temporales de 7 años para 45 empresas) aparece almacenada inicialmente en dos únicas series (AYUDA, BENEFICIO) que contienen 315 observaciones cada una.

La estimación mínimo cuadrática de un modelo lineal para las ayudas a la cooperación proporcionaría el resultado:

$$\hat{AYUDA} = 14.38 + 0.0084 \text{ BENEFICIO}$$

(1.026) (0.0003)

al que también podría llegarse con un modelo de datos de panel.

Para trabajar con este tipo de modelos es necesario almacenar la información de cada empresa de forma individualizada, llegando así a tener 45 series de las magnitudes investigadas, con 7 observaciones temporales cada una. En el programa Eviews el procedimiento necesario para llegar a esta nueva situación consiste en diseñar un nuevo fichero de trabajo sobre el que se define un *Pool* en el que se almacenan las variables AYUDA, BENEFICIO. Una vez copiada la información para dichas variables, el programa extrae desde dicho *Pool* cada una de las series individuales de las empresas³⁰.

Partiendo del objeto *Pool*, la estimación de un modelo único se lleva a cabo mediante el comando de estimación³¹, seleccionando la opción de intercepto común:

Dependent Variable: AYUDA?

Method: Pooled Least Squares

Sample: 1992 1998

Included observations: 7

Number of cross-sections used: 45

Total panel (balanced) observations: 315

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	14.38115	1.026224	14.01366	0.0000
BENEFICIO?	0.008454	0.000350	24.16696	0.0000
R-squared	0.651075	Mean dependent var		36.66780
Adjusted R-squared	0.649961	S.D. dependent var		13.50565
S.E. of regression	7.990500	Sum squared resid		19984.45
F-statistic	584.0421	Durbin-Watson stat		0.273753
Prob(F-statistic)	0.000000			

obteniéndose resultados idénticos a los del modelo global anteriormente estimado.

³⁰ En concreto, el proceso para este supuesto consiste en definir el archivo PANELCOOP.WF1 con el período temporal de estudio (años 1992 1998) sobre el cual se define un objeto *Pool* en el que es necesario identificar las unidades transversales (en nuestro caso las empresas han sido numeradas de 1 hasta 45). Posteriormente, en la opción *View/Spreadsheet* se introduce el nombre de las variables seguido de un signo de interrogación y se copia la información desde el archivo inicial.

Como resultado de este procedimiento, aparecen almacenadas 45 series de cada una de las variables, que el programa designa con los nombres de dichas variables seguidos de su número (AYUDA01, AYUDA02,...)

³¹ Obsérvese que añadimos una interrogación al final del nombre de las variables, indicando así que se consideran sus distintas observaciones para cada unidad transversal.

- b) Por lo que se refiere a la estimación de los efectos medios, si bien el programa Eviews no lleva incorporada esta opción de forma específica, cabría la posibilidad de efectuar la estimación sobre los efectos medios (regresión *between*) mediante la incorporación de la función @MEAN sobre las variables AYUDA y BENEFICIO:

Dependent Variable: @MEAN(AYUDA?)

Method: Pooled Least Squares

Number of cross-sections used: 45

Total panel (balanced) observations: 45

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	15.18439	2.656811	5.715270	0.0000
@MEAN(BENEFICIO?)	0.009104	0.000975	9.336788	0.0000
R-squared	0.669677	Mean dependent var		36.98719
Adjusted R-squared	0.661995	S.D. dependent var		14.62107
S.E. of regression	8.500427	Sum squared resid		3107.062
F-statistic	87.17560	Prob(F-statistic)		0.000000

Este resultado proporciona la estimación para una muestra de 45 datos, correspondientes a las medias de beneficios y ayudas a la cooperación de cada empresa en el período temporal considerado. Con respecto al modelo anterior, se aprecian valores estimados más elevados tanto para la “ayuda fija” (que pasaría de 14.38 a 15.18 millones) como para el efecto marginal de los beneficios sobre las ayudas (de 0.0084 pasa a 0.009).

- c) La consideración de un modelo de Efectos Fijos sería equivalente a contemplar el efecto diferenciado de cada una de las empresas mediante variables *dummy*.

Dependent Variable: AYUDA?

Method: Pooled Least Squares

Sample: 1992 1998

Included observations: 7

Number of cross-sections used: 45

Total panel (balanced) observations: 315

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BENEFICIO?	0.001955	0.000819	2.388560	0.0176
Fixed Effects				
01—C	32.81013			
02—C	31.18592			
03—C	30.24262			
04—C	28.03342			
.....	
.....	
42--C	40.85035			
43--C	35.82476			
44--C	37.03107			
45--C	32.83065			
R-squared	0.936886	Mean dependent var		36.66780
Adjusted R-squared	0.926328	S.D. dependent var		13.50565
S.E. of regression	3.665774	Sum squared resid		3614.795
Durbin-Watson stat	1.426859			

Debemos ahora contrastar la hipótesis de igualdad de efectos sobre el modelo anterior. A partir de la salida de la estimación (que hemos reproducido sólo parcialmente) se aprecian ya algunas diferencias sustanciales en los coeficientes, que permiten sospechar que la hipótesis no podrá ser admitida.

En efecto, planteando un contraste de restricciones sobre coeficientes llegaríamos a la conclusión esperada. El método consistiría en comparar las sumas de residuos cuadráticos en los modelos libre y restringido mediante una F de Snedecor y, si bien no aparece implementado de forma directa en el programa Eviews su construcción resulta muy sencilla (las sumas de residuos cuadráticos aparecen en los outputs de las regresiones y son también recuperables mediante el comando @SSR aplicado inmediatamente después de la estimación de un modelo).

Los resultados de este procedimiento serían los siguientes:

SSR Modelo libre (Efectos fijos)	3614.79
SSR Modelo restringido (Efectos comunes)	19984.45
Discrepancia F	27.68
Nivel crítico p	0.00

y conducen por tanto al rechazo de la hipótesis de efectos iguales.

- d) Planteamos por último la estimación de un modelo de Efectos aleatorios³², que proporciona el siguiente resultado (reproducido también parcialmente):

Dependent Variable: AYUDA?		Method: GLS (Variance Components)		
Sample: 1992 1998		Included observations: 7		
Number of cross-sections used: 45		Total panel (balanced) observations: 315		
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.91323	1.978421	11.58157	0.0000
BENEFICIO?	0.005218	0.000618	8.442889	0.0000
Random Effects				
01--C	6.930237			
02--C	-2.036055			
03--C	-1.337437			
.....	
44--C	3.357975			
45--C	-0.770996			
GLS Transformed Regression				
R-squared	0.919075	Mean dependent var	36.66780	
Adjusted R-squared	0.918817	S.D. dependent var	13.50565	
S.E. of regression	3.848122	Sum squared resid	4634.917	
Durbin-Watson stat	1.127092			
Unweighted Statistics including Random Effects				
R-squared	0.932633	Mean dependent var	36.66780	
Adjusted R-squared	0.932418	S.D. dependent var	13.50565	
S.E. of regression	3.510997	Sum squared resid	3858.383	
Durbin-Watson stat	1.353929			

³² El modelo de efectos aleatorios asume que la perturbación es suma de la actuación conjunta de componentes independientes $\mu_i \approx \text{IID}(0, \sigma_\mu^2)$ $v_{it} \approx \text{IID}(0, \sigma_v^2)$. Su estimación se lleva a cabo partiendo del modelo de efectos fijos, cuyos residuos proporcionan una primera estimación de la varianza *within*. Dicha varianza, junto con la asociada a la regresión *between*, conducen a las ponderaciones necesarias para la aplicación de Mínimos Cuadrados Generalizados (GLS).

Como consecuencia del procedimiento de estimación empleado en este caso (basado en la transformación del modelo inicial) se recogen en este caso las medidas de bondad referidas tanto a los residuos del modelo ponderado (GLS) como a los del inicial.

Cabe señalar además la posibilidad de contrastar la presencia de efectos aleatorios, mediante la construcción del test de Hausman, que si bien no aparece como opción directa del programa Eviews puede ser programado con relativa facilidad. (Aunque no incluimos aquí los resultados de este contraste, en nuestro caso concreto, la conclusión apoyaría un modelo de efectos fijos).

- e) Hasta ahora habíamos supuesto que las perturbaciones se encontraban incorreladas a lo largo del tiempo. Planteando ahora la inclusión de términos AR en el modelo de efectos fijos se obtendría:

Dependent Variable: AYUDA?

Method: Pooled Least Squares

Sample: 1992 1998

Included observations: 7

Number of cross-sections used: 45

Total panel (balanced) observations: 270

Convergence achieved after 5 iteration(s)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
BENEFICIO?	0.000986	0.001159	0.851195	0.3956
AR(1)	0.259804	0.063852	4.068837	0.0001
Fixed Effects				
01--C	32.79046			
02--C	35.93567			
03--C	32.02868			
04--C	31.45536			
05--C	45.88883			
.....			
40--C	37.33277			
41--C	30.60358			
42--C	43.11038			
43--C	37.90074			
44--C	39.71296			
45--C	35.22310			
R-squared	0.947021	Mean dependent var		36.61456
Adjusted R-squared	0.936092	S.D. dependent var		13.33901
S.E. of regression	3.372098	Sum squared resid		2535.742
F-statistic	3986.196	Durbin-Watson stat		2.007194
Prob(F-statistic)	0.000000			

Es decir, los términos AR, aunque presentan coeficientes estimados de pequeño valor, resultan estadísticamente significativos, mejorando la bondad global del modelo.

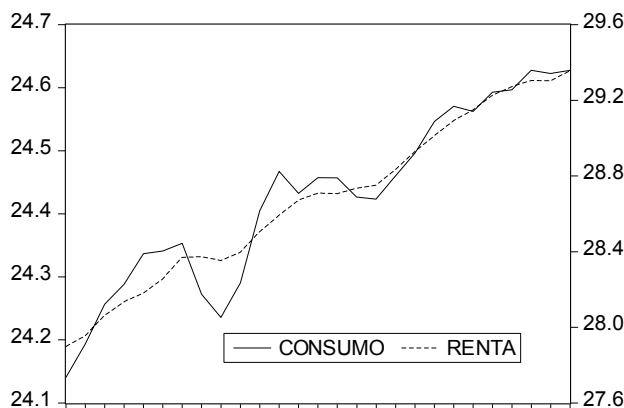
SUPUESTO 10

- ⇒ El fichero **CONSUMO.WF1** contiene información referida a las variables Renta Familiar Disponible (RENTA) y Consumo de los hogares (CONSUMO) para cierta región, ambas expresadas en miles de euros:

Años	CONSUMO	RENTA
1972	24.14057	27.89822
1973	24.19394	27.95565
1974	24.25661	28.06311
1975	24.28814	28.13483
1976	24.33689	28.1804
1977	24.34095	28.25646
1978	24.35333	28.36915
1979	24.27251	28.37246
1980	24.2357	28.3529
1981	24.2906	28.39738
1982	24.40474	28.50586
1983	24.46723	28.59263
1984	24.43256	28.67266
1985	24.45744	28.70956
1986	24.45706	28.70595
1987	24.42658	28.73516
1988	24.42344	28.75095
1989	24.45989	28.83446
1990	24.49638	28.92998
1991	24.54653	29.01396
1992	24.57053	29.09268
1993	24.56264	29.14931
1994	24.59273	29.22652
1995	24.59664	29.2719
1996	24.62734	29.30458
1997	24.62265	29.30238
1998	24.62734	29.3577

A partir de la información anterior estudiar la posible existencia de cointegración entre las variables, estimando si es posible un Modelo de Corrección de Error.

En una primera aproximación, el análisis conjunto del consumo y la renta permite apreciar la existencia de tendencias comunes crecientes a lo largo del período investigado



Para llevar a cabo un análisis de cointegración sería posible aplicar dos metodologías alternativas: el procedimiento bietápico de Engle y Granger (1987) y el método máximo-verosímil de Johansen (1988), encontrándose la segunda directamente implementada en el programa Eviews.

La aplicación de la opción *Cointegration Test* sobre las variables CONSUMO y RENTA proporciona el resultado siguiente:

Included observations: 25
Series: CONSUMO RENTA
Lags interval: 1 to 1

Data Trend:	None	None	Linear	Linear	Quadratic
Rank or	No Intercept	Intercept	Intercept	Intercept	Intercept
No. of CEs	No Trend	No Trend	No Trend	Trend	Trend
Log Likelihood by Model and Rank					
0	98.58701	98.58701	104.4943	104.4943	104.6944
1	104.7849	107.4625	112.5962	112.9809	113.0884
2	104.7953	112.9066	112.9066	118.2445	118.2445
Akaike Information Criteria by Model and Rank					
0	-7.566961	-7.566961	-7.879543	-7.879543	-7.735553
1	-7.742790	-7.876999	-8.207694	-8.158474	-8.087074
2	-7.423621	-7.912532	-7.912532	-8.179557	-8.179557
Schwarz Criteria by Model and Rank					
0	-7.371941	-7.371941	-7.587013	-7.587013	-7.345513
1	-7.352750	-7.438204	-7.720143	-7.622169	-7.502013
2	-6.838561	-7.229961	-7.229961	-7.399477	-7.399477
L.R. Test:	Rank = 0	Rank = 2	Rank = 1	Rank = 1	Rank = 2

Como se puede apreciar, en la salida anterior aparecen resumidas las conclusiones del test de Johansen para todos los casos posibles respecto a la existencia de tendencia determinista en los datos y de constante y tendencia en la ecuación de cointegración³³. De hecho, en la primera situación (rango de cointegración 0) se concluiría que las variables CONSUMO y RENTA no están cointegradas, mientras en los casos 2 y 5 la conclusión sería que el rango de cointegración es 2, equivalente a la estacionariedad de las series³⁴. Únicamente las situaciones 3 y 4 apoyan la existencia de cointegración entre el consumo y la renta familiar.

A la vista de estos resultados podríamos centrarnos en el caso 3 que, por ser muy habitual en la práctica, es el incorporado en el programa por defecto, y resulta adecuado a nuestra información muestral³⁵:

Included observations: 25

Test assumption: Linear deterministic trend in the data

Series: CONSUMO RENTA

Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.476988	16.82472	15.41	20.04	None *
0.024532	0.620947	3.76	6.65	At most 1

(**) denotes rejection of the hypothesis at 5%(1%) significance level
L.R. test indicates 1 cointegrating equation(s) at 5% significance level

Unnormalized Cointegrating Coefficients:

CONSUMO	RENTA
-7.033321	2.107143
-0.948868	0.789976

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

CONSUMO	RENTA	C
1.000000	-0.299594 (0.01506)	-15.83758

Log likelihood 112.5962

La aplicación del test de Johansen bajo esta estructura concreta (tendencia determinística en los datos y ecuación con término independiente) proporciona los resultados de la tabla anterior, que incluyen los contrastes sobre número de relaciones de cointegración y una estimación de la relación de cointegración, una vez normalizados los coeficientes.

³³ El test de cointegración de Johansen se basa en la estimación máximo-verosímil de un modelo VAR entre las variables investigadas, obteniéndose el número de relaciones de cointegración que coincide con el rango de la matriz Π asociada a dicho modelo.

³⁴ En el caso de que el rango de cointegración coincidiera con el número de variables investigadas la conclusión sería que ninguna de las series presenta raíz unitaria, por lo cual el VAR puede especificarse en términos de las series en nivel. Este resultado puede ser en ocasiones contradictorio con los asociados a los test de raíces unitarias tipo ADF.

³⁵ De hecho, puede comprobarse fácilmente que la tendencia determinista (generada como T=@TREND) resultaría significativa para modelizar individualmente las variables CONSUMO y RENTA.

Por lo que se refiere a los tests sobre rango de cointegración, éstos se llevan a cabo de modo secuencial contrastando en primer lugar la hipótesis $r=0$ que, con nuestros datos, debe ser rechazada al 5%. A continuación se pasaría a contrastar el supuesto $r \geq 1$ que en nuestro caso no resulta rechazable.

Una vez asumida la existencia de una relación de cointegración, ésta vendría estimada por la expresión:

$$\hat{C}_t = 115.837 + 0.2996R_t$$

(0.015)

y el teorema de Granger garantizaría la posibilidad de describir la relación entre las variables mediante un Modelo de Corrección de Error (MCE). Esta opción, accesible en Eviews desde las salidas del test de Johansen, proporciona el resultado:

Included observations: 25 after adjusting endpoints
Standard errors & t-statistics in parentheses

Cointegrating Eq:	CointEq1	
CONSUMO(-1)	1.000000	
RENTA(-1)	-0.299594	
	(0.01506)	
	(-19.8961)	
C	-15.83758	
Error Correction:	D(CONSUMO)	D(RENTA)
CointEq1	-0.831971	-0.246605
	(0.20500)	(0.23045)
	(-4.05835)	(-1.07013)
D(CONSUMO(-1))	0.751564	0.567370
	(0.19578)	(0.22008)
	(3.83883)	(2.57804)
D(RENTA(-1))	-0.068513	0.091101
	(0.21020)	(0.23629)
	(-0.32594)	(0.38555)
C	0.006692	0.040025
	(0.01205)	(0.01354)
	(0.55536)	(2.95501)
R-squared	0.533809	0.334614
Adj. R-squared	0.467210	0.239559
Sum sq. resids	0.017841	0.022544
S.E. equation	0.029147	0.032765
F-statistic	8.015309	3.520211
Log likelihood	55.09080	52.16601
Akaike AIC	-4.087264	-3.853280
Schwarz SC	-3.892244	-3.658260
Mean dependent	0.017336	0.056082
S.D. dependent	0.039932	0.037573
Determinant Residual Covariance		4.20E-07
Log Likelihood		112.5962
Akaike Information Criteria		-8.207694
Schwarz Criteria		-7.720143

a partir del cual se describen los MCE estimados para las variables. En concreto, para el consumo se tendría:

$$(1-L)\hat{C}_t = -0.83(C_{t-1} - 0.2995R_{t-1} - 15.837) + 0.7515(1-L)C_{t-1} - 0.068(1-L)R_{t-1} + 0.00669$$

expresión en la que se aprecian la relación de equilibrio a largo entre las variables y los desequilibrios a corto plazo o errores.

Una vez estimado el MCE, la conexión entre consumo y renta podría ser validada mediante el test de causalidad de Granger, cuyo resultado en este caso permite concluir que, con un nivel de significación del 5%, debemos rechazar el supuesto de que la renta no causa el consumo (sin que fuese posible, al mismo nivel, rechazar la causalidad en sentido contrario).

Pairwise Granger Causality Tests

Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
CONSUMO does not Granger Cause RENTA	25	3.03904	0.07040
RENTA does not Granger Cause CONSUMO		7.71484	0.00329

REFERENCIAS BIBLIOGRÁFICAS

- ALCAIDE, A.; ALVAREZ, N. (1992): *Econometría. Métodos deterministas y estocásticos*. (Teoría y Aplicaciones), Ed. Centro de Estudios Ramón Areces, S.A.
- AZNAR, A.; TRÍVEZ, F.J. (1992): *Métodos de predicción en Economía*, Tomo I, Ed. Ariel.
- AZNAR, A.; TRÍVEZ, F.J. (1993): *Métodos de predicción en economía*, Tomo II: Análisis de series temporales, Ed. Ariel, Barcelona.
- BALTAGI, B.H. (1998): *Econometrics*, Springer.
- BOOT y DE WITT (1960): “Investment demand: an empirical contribution to the aggregation problem”, *International Economic Review*, 1, 3-30.
- BOX, G.E.P.; JENKINS, G.M. (1976): *Time Series Analysis: Forecasting and Control*, (2ª ed.), Ed. Holden-Day, San Francisco.
- CHAREMZA, W.W.; DEADMAN, D.F. (1997): *New Directions in Econometric Practice*, Edward Elgar.
- CHOW, G.C. (1960): “Tests of Equality Between Subsets of Coefficients in Two Lineal Regression Models”, *Econometrica*, 28, 591-605.
- DICKEY, D.A.; FULLER, W.A. (1979): “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”, *Journal of the American Statistical Association*, 74, 427-431.
- DOTI, J.L.; ADIBI, E. (1998): *The Practice of Econometrics with Eviews*, Quantitative Micro Software.
- DURBIN, J.; WATSON, G.S. (1934): “Testing for Serial Correlation in Least Squares Regression”, *Biometrika*, 37, 409-428 y 38, 159-178.
- ENGLE, R.F. (1984): “Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics”, en Griliches, Z.; Intrilligator, M.D. (eds): *Handbook of Econometrics*, 2.
- ENGLE, R.F.; GRANGER, C. (1987): “Co-integration and Error Correction: Representation, Estimation and Testing”, *Econometrica*, 35, 251-276.
- GOLDBERGER, A.S. (1991): *A Course in Econometrics*, Harvard University Press.
- GOLDFELD, S.M.; QUANDT, R.E. (1965): “Some Tests for Homoscedasticity”, *Journal of the American Statistical Association*, 60, 539-547.
- GRANGER, C. W. J. (1969): “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods”, *Econometrica*, 37, 424-438.
- GREENE, W.H. (1998): *Análisis Económico*, Ed. Prentice Hall.

- GRIFFITHS, W.E.; CARTER, R.; JUDGE, G.G. (1993): *Learning and Practicing Econometrics*, John Wiley & Sons, Inc.
- GRUNFELD, Y.; GRILICHES, Z. (1960): "Is Aggregation Necessarily Bad?" *Review of Economics and Statistics*, 42, 1-13.
- GUJARATI, D. (1995): *Econometría*, Ed. McGraw-Hill.
- JARQUE, C.M.; BERA, A.K. (1980): "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals", *Economics Letters*, 6, 255-259.
- JOHANSEN, J. (1988): "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.
- KAWAKATSU, H. (1998): *A Computer Handbook using Eviews*, Ed Irwin & McGraw-Hill.
- KMENTA, J. (1986): *Elementos de Econometría*, Vicens Universidad.
- LJUNG, G.; BOX, G. (1979) "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, 66, 265–270.
- MADDALA, G.S. (1992): *Introduction to Econometrics*. Ed. Prentice Hall.
- NEWAY, W.; WEST, K. (1987): "A Simple Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
- OTERO, J.M. (1994): *Modelos econométricos y predicción de series temporales*, Ed. AC, Madrid.
- PÉREZ, R. LÓPEZ, A.J. (1997): *Análisis de datos económicos II. Métodos inferenciales*, Ed. Pirámide.
- PINDYCK, R.S.; RUBINFELD, D.L.(1998): *Econometric Models and Economic Forecasts*, Ed. McGraw-Hill.
- PULIDO, A. (1987): *Modelos econométricos*, Ed. Pirámide, Madrid.
- QUANTITATIVE MICRO SOFTWARE (1998): *EViews User's Guide, Command and Programming Reference*.
- THEIL, H. (1958): *Economic Forecasts and Policy*, North Holland Publishing.
- THEIL, H. (1966): *Applied Economic Forecasting*, North Holland Publishing.
- THOMAS, R. L. (1993): *Introductory Econometrics. Theory and Applications*, Addison Wesley.
- THOMAS, R. L. (1996): *Modern Econometrics*, Addison Wesley.
- URIEL, E.; GEA, I. (1997): *Econometría Aplicada*, Ed AC.
- WHITE, H. (1980) "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

ANEXOS

📄 A1: OBTENCIÓN DE VECTORES DE ESTIMADORES MÍNIMO CUADRÁTICOS

La construcción de los vectores de estimadores mínimo cuadráticos puede ser llevada a cabo partiendo de objetos definidos como vectores y matrices. Así, para el archivo VENTAS.WF1 comenzamos por almacenar en un vector las observaciones de las ventas mediante la orden:

$$\text{VECTOR Y}=@\text{CONVERT}(\text{VENTAS})$$

De modo similar, podemos crear la matriz X de datos, para lo cual generamos un grupo (que denominamos DATOS) en el que se incluyen el término independiente y las variables explicativas PRECIO y PUBLICIDAD³⁶. La conversión de este grupo a una matriz se llevaría a cabo entonces mediante el comando:

$$\text{MATRIX X}=@\text{CONVERT}(\text{DATOS})$$

Para obtener el vector de estimadores mínimo cuadráticos (que denominamos B) bastaría operar sobre los objetos anteriores con los correspondientes comandos de *Eviews*³⁷.

En una sola operación se tendría:

$$\text{VECTOR B}=@\text{INVERSE}(@\text{INNER}(\text{X})) * @\text{TRANSPOSE}(\text{X}) * \text{Y}$$

si bien la definición de cálculos parciales puede ser útil para la obtención de medidas de bondad. Así, el vector de residuos (que las opciones automáticas del programa generan como RESID) podría ser obtenido como:

$$\text{VECTOR U}=\text{Y}-\text{X} * \text{B}$$

llegándose a la matriz de varianzas-covarianzas de los estimadores como:

$$\text{MATRIX VARB}=(@ \text{INNER}(\text{U}) / (@ \text{REGOBS} - @ \text{NCOEF})) * @ \text{INVERSE}(@ \text{INNER}(\text{X}))$$

A partir de dicha matriz podemos obtener los valores t y niveles críticos asociados a los distintos contrastes de significación individual. Así, para el coeficiente de los precios se tendría:

$$\text{SCALAR TPRECIO}=\text{B}(2) / @ \text{SQR}(\text{VARB}(2,2))$$

$$\text{SCALAR PPRECIO}=@ \text{TDIST}(\text{TPRECIO}, @ \text{REGOBS} - @ \text{NCOEF})$$

comandos que proporcionan respectivamente los resultados $t=-2.19$ y $p=0.0332$, coincidentes con los recogidos en la salida del modelo de regresión.

³⁶ Para generar el grupo basta mostrar los regresores (mediante la orden SHOW C PRECIO PUBLICIDAD) nombrando este grupo como DATOS y almacenando a continuación su información en una matriz X.

Otra alternativa para definir vectores y matrices consiste en generar objetos con las dimensiones requeridas (50 filas y 1 columna para y, 50 filas y 3 columnas para X) que en principio estarían “vacíos”, y a los que posteriormente podemos asignar los correspondientes valores mediante el comando @CONVERT.

³⁷ Los principales comandos utilizados son el de transposición de matrices (@TRANSPOSE), cálculo de inversas (@INVERSE) y premultiplicación por la transpuesta (@INNER).

A2: PROGRAMA PARA LA APLICACIÓN DEL TEST DE GOLDFELD Y QUANDT (GQ.Prg)

Este programa contrasta el supuesto de homoscedasticidad mediante el test propuesto por Goldfeld y Quandt. Los autores parten de los enunciados de las hipótesis:

$$H_0 : \sigma_i^2 = \sigma^2 \quad \forall i = 1, \dots, n$$

$$H_1 : \sigma_i^2 = \sigma^2 X_i^2$$

siendo por tanto necesario partir de un modelo de regresión y de una variable asociada a la presencia de heteroscedasticidad.

Ambos parámetros deben ser proporcionados al programa *GQ.prg* para su ejecución. Con esta información (en el actual ejemplo escribiríamos MODELO PUBLICIDAD) el programa ordena las observaciones según la variable asociada a la heteroscedasticidad (PUBLICIDAD en este caso) y plantea regresiones parciales en dos submuestras, una vez eliminado un número de observaciones centrales suficiente para garantizar la independencia.

Con la información sobre sumas de cuadrados de residuos en ambas regresiones y los correspondientes grados de libertad, el programa obtiene el estadístico F asociado al contraste y el nivel crítico asociado, devolviendo una tabla final del tipo:

G_Q Estadístico=		2.225
Dist. F, Gr.l.=	26	25
Nivel Crítico		0.0246
=====		
Observaciones en la Regresión		
n=	62	
p=	6	
Primera muestra		
Obs. Regresión:	28	
Sample:	1	66
Sum. u^2=	583659.9475	
Segunda muestra		
Obs.	29	
Sample:	72	Final
Sum. u^2=	1298767.382	

A3: PROGRAMA PARA EL ANÁLISIS DEL ESQUEMA DE COMPOSICIÓN DE UNA SERIE (Hipótesis.Prg)

Este programa contrasta la presencia de un esquema de composición aditivo frente a la alternativa de un esquema multiplicativo.

Una vez introducidos como parámetros el nombre de la variable de referencia y su fecha de inicio, el programa obtiene los valores de las medias anuales y las desviaciones típicas anuales estimando a partir de ellas el modelo $S_{Y_t} = \beta_0 + \beta_1 \bar{Y}_t + u_t$ sobre el que se contrasta el supuesto de hipótesis aditiva $H_0 : \beta_1 = 0$

El programa devuelve una tabla con los resultados del valor de la discrepancia t de Student, el nivel crítico asociado y la correspondiente conclusión (*hipótesis multiplicativa* si el resultado es significativo, al 5% (*) o el 1% (**) e *hipótesis aditiva* en otro caso).

Asimismo, el programa construye el gráfico desviación típica-media que permite ilustrar la conclusión a la que se llega con la información muestral disponible.

En el caso del supuesto 6, con datos mensuales de producción de carne en Asturias, los resultados obtenidos al ejecutar el programa aparecen recogidos a continuación:

Contraste para el tipo de hipótesis	
Valor t	Nivel crítico
2.4484110	0.0118485
<i>Hipótesis multiplicativa (*)</i>	

