

EL USO DE LA ECONOMETRÍA B + Á + S + I + C + A

ISBN: 978-958-5123-34-2

$$V_t = \beta_0 + \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 LGP_t + \beta_5 V_{t-1} + \mu_{ti}$$

Diagram illustrating the components of the regression equation:

- Variable endógena:** V_t
- Intercepto:** β_0
- Coeficientes:** $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$
- Variabes explicativas:** $PE_t, PC_t, GP_t, LGP_t, V_{t-1}$
- Componente sistemático:** $\beta_0 + \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 LGP_t + \beta_5 V_{t-1}$
- Error aleatorio:** μ_{ti}
- Componente aleatorio:** μ_{ti}

**Mario Eduardo
Hidalgo Villota**



Editorial
Universidad de Nariño

**Universidad de Nariño
Facultad de Ciencias Económicas y Administrativas
Departamento de Economía**

Del Autor.

Mario Eduardo Hidalgo Villota. Profesor Asistente adscrito al Departamento de Economía de la Universidad de Nariño, Colombia. Actualmente se desempeña como Director del mismo Departamento. Economista de la Universidad de Nariño, Especialista en Finanzas Públicas de la Escuela Superior de Administración Pública (ESAP), Especialista en Gerencia de Proyectos de la Universidad del Cauca en convenio con la Universidad de Nariño y Magister en Políticas Públicas de la Universidad del Valle. Ex-Director del Centro de Estudios de Desarrollo Regional (CEDRE) y Ex-Director de la Oficina de Planeación y Desarrollo de la Universidad de Nariño. Investigador asociado al Grupo de Investigación en Economía, Gobierno y Políticas Públicas de la Universidad de Nariño.

EL USO DE LA ECONOMETRÍA BÁSICA

Un análisis de caso realizado paso a paso

Mario Eduardo Hidalgo Villota



Universidad de Nariño
Facultad de Ciencias Económicas y Administrativas
Departamento de Economía

Hidalgo Villota, Mario Eduardo

El uso de la Econometría Básica: Un análisis de caso realizado paso a paso/ Mario Eduardo Hidalgo.
-1ª.-ed.- Pasto: Editorial Universidad de Nariño, 2020.
305p.: tablas.

Incluye bibliografía
ISBN 978-958-5123-34-2

1. Econometría-enseñanza aprendizaje 2. Econometría básica 3. Econometría--ejercicios prácticos
4. Análisis económico aplicado.

330.015195 H632 -SCDD-Ed 22

Biblioteca Alberto Quijano Guerrero

EL USO DE LA ECONOMETRÍA BÁSICA

Un análisis de caso realizado paso a paso

© Mario Eduardo Hidalgo Villota
mariohidalgo@udenar.edu.co

© Editorial Universidad de Nariño

ISBN: 978-958-5123-34-2

Primera Edición

Corrección de Estilo: Mg. Manuel E. Martínez R.

Diseño y Diagramación: María Elena Mesías P.

Centro de Publicaciones Universidad de Nariño

Fecha de Publicación: Octubre de 2020

San Juan de Pasto - Nariño - Colombia

Prohibida la reproducción total o parcial, por cualquier medio o con cualquier propósito, sin la autorización escrita de su Autor o de la Editorial Universidad de Nariño.

Las bases de datos en archivos excel y wf1 (EViews) que acompañan este texto, se pueden descargar en el siguiente link:

<http://sired.udenar.edu.co/6846/>

AGRADECIMIENTOS

El autor expresa su profundo agradecimiento académico a Alonso Antón, A.; Fernández Macho, J. & Gallastegui Zulaica, I., catedráticos de Econometría de la Universidad del País Vasco/Euskal Herriko Unibertsitatea (España), autores del libro Econometría publicado por la editorial Pearson Prentice Hall en 2005, por su generosidad, al haber permitido utilizar el formato de la historieta que se desarrolla en este documento, así como las bases de datos incluidas en su libro de texto. A ellos, mi más sincera gratitud.

[El concepto de la Historia de la Editorial, las bases de datos y la Figura 6.1 han sido tomadas de: Alonso, Fernández y Gallastegui. Econometría. España. Pearson Prentice Hall, 2005. Reproducidas con autorización]

DEDICATORIA

A la memoria de mis padres y hermana. Donde quiera que estén, sé que se sentirán orgullosos de mí.

A mi esposa e hijos, por su paciencia y su inmenso amor. Gracias a ellos por haber cedido de manera generosa y desinteresada, parte del tiempo de la familia. Ustedes son mi inspiración y el motor de mi vida.

A los estudiantes de Economía y de otras disciplinas, quienes inspiran día a día el deseo de engrandecer la profesión del maestro universitario.

CONTENIDO

INTRODUCCIÓN.	12
1. VENCIENDO EL MIEDO: UNA APROXIMACIÓN CONCEPTUAL A LA ECONOMETRÍA.	16
1.1 Así comienza la historia de la editorial.	17
1.2 Componentes constitutivos de la econometría.	21
1.3 Diferencias entre un modelo matemático y un modelo econométrico.	26
1.4 Relaciones estadísticas y relaciones deterministas	33
1.5 Regresión y causalidad.	34
1.6 La causalidad y la noción de <i>ceteris paribus</i> en el análisis econométrico.	35
1.7 Regresión y correlación.	35
1.8 Denominación de las variables en un modelo econométrico.	37
1.9 Etapas en la elaboración de un modelo econométrico.	39
1.9.1 Especificación.	39
1.9.2 Estimación.	39
1.9.3 Validación.	40
1.9.4 Uso del modelo econométrico.	40
2. ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: ESPECIFICACIÓN Y ESTIMACIÓN.	41
2.1 Formalización de la discusión.	41
2.1.1 Relación entre dos variables.	43
2.1.2 Interpretación de los coeficientes.	57
2.1.3 Coeficiente de determinación (R^2).	71

2.1.4	El coeficiente de determinación ajustado (\bar{R}^2).	76
2.1.5	Coeficiente de correlación.	79
3.	ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: ESPECIFICACIÓN Y ESTIMACIÓN.	85
3.1	Especificación del modelo.	92
3.2	Estimación del modelo.	101
3.3	Estadísticos de la estimación.	106
4.	ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: RESTRICCIONES Y CONTRASTES.	113
4.1	Contraste de significancia individual Modelo 1: Regresión simple.	115
4.2	Contraste de significancia individual Modelo 2: Regresión múltiple.	120
4.3	Contraste conjunto del modelo 2.	130
4.4	Contraste de significación para un subconjunto de coeficientes.	135
4.5	Contraste de hipótesis informativas.	136
5.	ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: DIAGNÓSTICO ESTRUCTURAL.	140
5.1	Selección de variables explicativas: error de especificación.	142
5.2	Errores de especificación en las variables explicativas.	143
5.2.1	Omisión de variables relevantes.	144
5.2.2	Inclusión de variables irrelevantes.	147
5.3	Análisis de la estabilidad estructural.	150
5.3.1	Contraste de Chow.	150
5.3.2	Contraste de predicción de Chow.	153
5.3.3	Estimación de residuos recursivos.	155
5.3.4	Estimación de coeficientes recursivos.	156
5.3.5	El contraste de CUSUM de mala especificación.	159
5.3.6	Error de especificación de la forma funcional.	161
5.3.7	Normalidad de las perturbaciones.	164

6.	ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: PREDICCIÓN.	168
6.1	Intervalo de confianza de la predicción.	176
6.2	Evaluación de la capacidad predictiva del modelo.	176
6.3	Simulación de la capacidad predictiva del modelo.	179
7.	FACTORES EXPLICATIVOS CUALITATIVOS Y VARIABLES FICTICIAS.	215
8.	MULTICOLINEALIDAD.	221
9.	HETEROSCEDASTICIDAD.	225
9.1	Especificación y causas de la heteroscedasticidad.	226
9.2	Detección de la heteroscedasticidad mediante métodos gráficos.	227
9.3	Detección de la heteroscedasticidad mediante contrastes.	236
9.3.1	Contraste de White.	236
9.3.2	Contraste de Breusch-Pagan-Godfrey.	242
10.	AUTOCORRELACIÓN.	248
10.1	Detección de la autocorrelación mediante métodos gráficos.	248
10.2	Detección de la autocorrelación mediante contrastes.	253
10.2.1	El estadístico de Durbin-Watson.	253
10.2.2	Contraste de Breusch-Godfrey.	255
10.2.3	Contraste de Ljung-Box.	257
11.	REGRESORES ESTOCÁSTICOS Y MODELOS DINÁMICOS.	261
11.1	Contraste de exogeneidad de Hausman.	273
12.	USO DEL MODELO ECONÓMICO PARA FINES DE CONTROL O DE POLÍTICA.	278
13.	UNA SÍNTESIS DE LA HISTORIA DE LA EDITORIAL.	283
13.1	Planteamiento de la teoría o hipótesis relacionada con las ventas de la editorial.	283

13.2	Especificación del modelo matemático de ventas de la editorial.	284
13.3	Especificación del modelo econométrico de ventas de la editorial.	284
13.4	Estimación de los parámetros del modelo econométrico de las ventas de la editorial.	285
13.5	Prueba de hipótesis del modelo econométrico de ventas de la editorial.	286
13.6	Predicción.	287
13.7	Uso del modelo para fines de control o de política.	290
14.	LIMITACIONES DE LA ECONOMETRÍA.	298
	BIBLIOGRAFÍA.	300

LISTA DE TABLAS

Tabla 1.1	Elementos que conforman la econometría.	21
Tabla 1.2	Ventas trimestrales de libros 2009-2018 (pesos constantes de 2009).	31
Tabla 1.3	Denominación de las variables.	38
Tabla 2.1	Relación de ventas trimestrales de la editorial 2009-2018.	52
Tabla 2.2	Resultados MCO de las ventas contra el precio.	56
Tabla 2.3	Interpretaciones de los coeficientes.	58
Tabla 2.4	Resultados MCO de ventas contra el logaritmo natural del precio	65
Tabla 2.5	Resultados MCO del logaritmo natural de ventas contra el precio.	67
Tabla 2.6	Resultados MCO del logaritmo natural de las ventas contra el logaritmo natural del precio.	68
Tabla 2.7	Interpretaciones de los coeficientes o parámetros de los modelos de ventas de la editorial.	69
Tabla 2.8	Información para la estimación del coeficiente de determinación.	76
Tabla 2.9	Información para la estimación del coeficiente de correlación lineal.	80
Tabla 2.10	Coefficiente de correlación lineal.	83
Tabla 3.1	Ventas trimestrales de libros 2009-2018 (pesos constantes de 2009).	89
Tabla 3.2	Tasa de crecimiento anual media de las variables del modelo.	92
Tabla 3.3	Matriz de correlación lineal.	98
Tabla 3.4	Estadísticos descriptivos.	99
Tabla 3.5	Variables del modelo de la editorial.	100
Tabla 3.6	Estimación del modelo por MCO.	102
Tabla 3.7	Intervalos de confianza.	104

Tabla 3.8	Matriz de coeficientes de Covarianza.	112
Tabla 4.1	Modelo 1. Ventas de la editorial.	117
Tabla 4.2	Prueba de Wald.	120
Tabla 4.3	Modelo 2. Ventas de la editorial.	121
Tabla 4.4	Prueba de Wald.	125
Tabla 4.5	Prueba de Wald.	127
Tabla 4.6	Prueba de Wald.	128
Tabla 4.7	Prueba de Wald.	130
Tabla 4.8	Prueba de Wald.	135
Tabla 4.9	Prueba de Wald.	136
Tabla 5.1	Prueba de variables omitidas.	146
Tabla 5.2	Prueba de variables redundantes.	149
Tabla 5.3	Prueba de Chow.	152
Tabla 5.4	Prueba de pronóstico de Chow.	154
Tabla 5.5	Contraste de RESET de Ramsey.	163
Tabla 6.1	Evaluación del pronóstico.	178
Tabla 6.2	Estimación del modelo de ventas de la editorial.	180
Tabla 7.1	Horas de lectura semanales.	187
Tabla 7.2	Estimación del modelo de ventas contra el efecto estacional.	188
Tabla 7.3	Estimación del modelo con variables cualitativas.	191
Tabla 7.4	Estimación del modelo de ventas con variables cuantitativas y cualitativas.	193
Tabla 7.5	Variables cuantitativas y cualitativas (efecto estacional).	195
Tabla 7.6	Variables cuantitativas y cualitativas (efecto estacional y gasto en publicidad en periodo vacacional).	197
Tabla 7.7	Estimación modelo de ventas con efecto estacional y gasto en publicidad en periodo vacacional.	198
Tabla 7.8	Estimación modelo de ventas con gasto en publicidad en periodo vacacional.	201
Tabla 7.9	Estimación modelo utilizando distintas variables cualitativas.	203
Tabla 7.10	Estimación modelo lectura contra edad de los clientes y aparatos en el hogar.	204

Tabla 7.11	Estimación modelo utilizando distintas variables cualitativas.	207
Tabla 7.12	Estimación modelo lectura contra variables cualitativas relevantes.	208
Tabla 7.13	Estimación modelo lectura contra educación de las clientes mujeres.	210
Tabla 7.14	Estimación modelo lectura contra mujeres con educación universitaria.	211
Tabla 7.15	Estimación modelo final de lectura.	212
Tabla 8.1	Matriz de correlación lineal.	218
Tabla 8.2	Regresiones auxiliares: Precio medio de la editorial.	220
Tabla 8.3	Regresiones auxiliares: Precio medio de la competencia.	220
Tabla 8.4	Regresiones auxiliares: Gasto en publicidad.	221
Tabla 8.5	Regresiones auxiliares: Renta personal.	221
Tabla 9.1	Contraste de Glesjer.	236
Tabla 9.2	Detección de heteroscedasticidad (contraste de White).	237
Tabla 9.3	Prueba de heteroscedasticidad de White.	238
Tabla 9.4	Prueba de heteroscedasticidad de White.	239
Tabla 9.5	Prueba de heteroscedasticidad de White.	240
Tabla 9.6	Prueba de heteroscedasticidad de White.	241
Tabla 9.7	Prueba de heteroscedasticidad de Breusch-Pagan-Godfrey.	244
Tabla 9.8	Estimador robusto de White	246
Tabla 9.9	Estimador robusto de White-Hinkley.	247
Tabla 10.1	Breusch-Godfrey Serial Correlation LM Test.	256
Tabla 11.1	Estimación modelo final de ventas de la editorial.	269
Tabla 11.2	Contraste Breusch-Godfrey Serial Correlation LM Test.	271
Tabla 11.3	Estimación del modelo por el método de mínimos cuadrados en dos etapas.	275

INTRODUCCIÓN

El libro que el lector tiene en sus manos, no es precisamente un manual de econometría, ni tampoco pretende ser considerado como tal; más bien, es un complemento a los excelentes textos de econometría y de economía aplicada disponibles en las librerías en el mundo que sirven de soporte para la enseñanza-aprendizaje de este campo disciplinar que forma parte de los planes de estudio tanto de pregrado como de posgrado en economía y en otras ciencias sociales. El propósito de esta obra es que el estudiante logre vencer el miedo cuando se enfrenta al estudio de esta disciplina y de servir de guía para el aprendizaje de algunos aspectos fundamentales de la econometría.

Este objetivo académico se pretende lograr mediante una historieta divertida que narra un hecho económico cotidiano en la economía de mercado como la teoría de la demanda, cuyos interrogantes son resueltos por los personajes que participan en ella, de manera secuencial y didáctica mediante el análisis econométrico.

La pretensión del autor es que los estudiantes y otras personas interesadas en el estudio de la econometría, pueda comprender de manera fácil y amena, el instrumental básico de que dispone esta herramienta para el análisis económico aplicado. Si bien, el estudio de la econometría es un tema un tanto complejo y poco atractivo para los estudiantes y profesionales universitarios, el interés académico se enfoca en simplificar el entendimiento de la econometría básica a partir del desarrollo de un caso práctico, considerando dos enfoques: por un lado, mediante procedimientos manuales –que permitan apreciar de manera fácil y clara–, cómo se estiman los coeficientes o parámetros de un modelo

econométrico, y segundo, mediante la automatización con la ayuda del software econométrico *EViews 10*, realizar paso a paso las rutinas necesarias para construir un modelo econométrico capaz de explicar y predecir un fenómeno económico cotidiano, en este caso, las ventas de una editorial. Hoy, el uso de la econometría es más simple, precisamente por la disponibilidad de software científico tanto de tipo comercial como libre.

Este libro además de los beneficios descritos, le permite comprender al lector cómo hoy, los paquetes econométricos procesan cuantiosa información de manera precisa, veloz y a un bajo costo, dentro de la lógica de un proyecto de investigación econométrica, asunto que hace un par de décadas hubiese sido prácticamente imposible. Si bien, el uso de la tecnología vuelve más simple la práctica econométrica, en ningún modo, puede convertirla en algo mecánico.

El contenido de este trabajo está inspirado en el modo innovador y pedagógico como *Alonso Antón, A., Fernández Macho, J. & Gallastegui Zulaica, I.*, catedráticos de econometría de la Universidad del País Vasco (España) escribieron su libro de texto: *Econometría* publicado en 2005 por la prestigiosa editorial Pearson Prentice Hall, quienes pretenden enseñar de modo didáctico esta compleja, y a la vez apasionante, técnica del análisis económico. Asumiendo este reto académico, el autor intentará interpretar y complementar el brillante trabajo realizado por los docentes mencionados, para lo cual se ha mantenido los nombres de los personajes de la historieta (*Don Pablo* y su hijo *Emilio*), parte de sus diálogos y discusiones teóricas y empíricas en torno al mejor modelo que explica las ventas de la editorial, así como una parcela significativa de la información contenida en las bases de datos que acompañan el mencionado libro de texto.

En síntesis, mediante un trabajo detallado y minucioso que se va enriqueciendo de capítulo en capítulo se ha intentado adentrarse en la comprensión y uso de la econometría básica conservando la esencia del mensaje y la doctrina académica de Alonso, Fernández & Gallastegui (2005).

No es intención del presente libro, hacer una exposición teórica detallada y exhaustiva de cada uno de los temas que se abordan en un curso de econometría

básica, sino más bien, exponer un par de conceptos claves haciendo uso del mínimo instrumental matemático y estadístico, privilegiando ante todo, las aplicaciones econométricas partiendo de un modelo específico que recoge la teoría del modelo lineal de regresión simple hasta transformarse en un modelo multivariado que incorpora variables ficticias e instrumentales. Estoy seguro que un estudio de caso como el aquí propuesto, seduce al estudiante y al profesional universitario en economía y en otras disciplinas, a interesarse por el aprendizaje de la econometría, la que en palabras de Gujarati (2010), *la cual le ha dado contenido empírico a gran parte de la teoría económica*.

A lo largo de sus catorce capítulos, con el estudio de caso desarrollado paso a paso, este libro aborda de manera general pero comprensible la temática fundamental de un curso de introducción a la econometría. Por su alcance, no pretende profundizar en la utilización de fórmulas estadísticas y comprobaciones matemáticas, y menos, en el excesivo acervo teórico que caracteriza a los manuales de econometría. Como estrategia didáctica, el estudio de caso aquí desarrollado atraviesa los aspectos esenciales de la econometría mediante aplicaciones que el estudiante puede replicar a partir de las bases de datos utilizadas en este libro.

En el apéndice de este documento no se incluyen las tablas estadísticas como suele ocurrir con los manuales de estadística y econometría. En su lugar, se propone que el estudiante curioso utilice el buscador de valores en las tablas estadísticas disponible en el software econométrico GRETL de uso gratuito, como se propone en el capítulo 4. Esto ahorrará esfuerzo y tiempo

El capítulo 1 hace una aproximación conceptual a la econometría desde la perspectiva del estudio de caso en cuestión; la especificación y estimación del modelo de regresión lineal simple se encara en el capítulo 2; el capítulo 3 desarrolla el modelo de regresión lineal general en lo que a especificación y estimación se refiere; el capítulo 4 se ocupa de las restricciones y contrastes del modelo de regresión lineal general; el capítulo 5 desarrolla las pruebas de diagnóstico y especificación del modelo lineal general; el capítulo 6 se dedica la predicción y evaluación de la capacidad predictiva del modelo;

el capítulo 7 se ocupa de los factores explicativos cualitativos y variables ficticias; el capítulo 8 se ocupa de la multicolinealidad; los capítulos 9 y 10 de la heteroscedasticidad y autocorrelación, respectivamente; en el capítulo 11 se estudia los regresores estocásticos y los modelos dinámicos; en el capítulo 12 aborda el uso del modelo econométrico para fines de control y de política; en el capítulo 13 se hace una síntesis de la historia de la editorial donde se resalta la evolución del modelo econométrico desde un modelo simple hasta uno más sofisticado que responde a las expectativas de los personajes de la historieta; y finalmente, considerando que la econometría no es perfecta, en el capítulo 14 se exponen algunas de sus limitaciones expuestas por sus opositores, sobre las cuales los expertos vienen desarrollando investigación que permita superar sus puntos débiles.

**VENCIENDO EL MIEDO:
UNA APROXIMACIÓN
CONCEPTUAL A LA
ECONOMETRÍA**

Para Portillo (2006), la concepción moderna de la ciencia económica, que surgió con la revolución marginal, incorporó el lenguaje matemático en la construcción de esta ciencia social. Jevons (1886: 321), uno de los pioneros de este enfoque, afirma: “que la economía, científicamente hablando, (...) es una especie de matemática que calcula las causas y los efectos de la actividad humana”. Ya en los inicios de la econometría como disciplina científica, Schumpeter (1933) califica a la economía como la más cuantitativa, no sólo de las ciencias sociales, sino de todas las ciencias, argumentando que los componentes fundamentales de los fenómenos económicos son cuantificables. Para este autor, la econometría es el reconocimiento explícito de este hecho y el intento de hacer frente a sus consecuencias (p. 3).

La aproximación al conocimiento y uso de la econometría se hará a partir del desarrollo de una divertida historieta protagonizada por dos personajes peculiares que discuten apasionadamente diferentes posiciones (hipótesis) sobre los factores determinantes de las ventas realizadas por la empresa editorial “*El mejor libro*”, cuyo fundador y líder es Don Pablo, un hombre de negocios de ideas conservadoras, y a la vez, extremadamente curioso. Don Pablo es asistido en la gestión de la editorial por su hijo Emilio, una persona joven, de ideas liberales, con formación universitaria y acervo teórico, quien considera que los negocios, hoy, se administran a partir de decisiones fundamentadas en la teoría, en la experiencia y con el uso de un buen modelo que se aproxime a la realidad.

Desde su experiencia y conocimiento del mundo editorial, cada uno de estos divertidos personajes exponen sus tesis o hipótesis sobre las variables más influyentes en las ventas de la editorial; preliminarmente, desde una posición teórica basada en el razonamiento lógico y en las percepciones individuales, y luego, apoyados en evidencia empírica.

1.1 Así comienza la historia de la editorial ...

La Editorial “*El mejor libro*” es una empresa familiar muy tradicional con más de medio siglo de existencia dedicada a la comercialización de diferentes tipos de libros nacionales y extranjeros, desde obras de literatura básica hasta volúmenes especializados en distintas áreas del conocimiento. En ocasiones excepcionales, la editorial se ha responsabilizado del diseño gráfico y del tiraje de algunas obras de interés cultural regional y nacional, en particular, cuando éstas requieren un toque diferenciador.

La editorial se encuentra administrada por Don Pablo, su fundador, quien trabaja en este sector desde muy joven. En los últimos años su hijo Emilio ha comenzado a trabajar en la editorial con la intención de hacerse cargo del negocio familiar cuando su padre se retire. Las dos generaciones, a pesar de llevarse muy bien, tienen distinta manera de entender el mundo editorial. Por esta razón, discuten con frecuencia sobre la política comercial que debe orientar mejor la selección de libros que comercializa la editorial.

Don Pablo con su vasta experiencia como librero considera que el elemento clave para incrementar las ventas de la empresa, es esencialmente, el precio de los libros. Según su razonamiento y experiencia acumulada, deduce que a precios bajos las ventas aumentan, pero ante aumentos del precio, las ventas se estancan o disminuyen (Don Pablo, no siendo economista, ha descrito de manera muy clara y sabia, la ley de la demanda¹). Don Pablo apoyado en este argumento, proporciona a su hijo la información que registra las ventas de la editorial por cada trimestre en los últimos diez años (2009-2018) extractados

¹ La ley de la demanda refleja la relación existente entre la cantidad demandada de un bien o un servicio.

de la contabilidad junto con el precio promedio² de los libros. En resumidas cuentas, la hipótesis de Don Pablo se sustenta en la intuición y experiencia personal de que *el precio es la variable explicativa de las ventas*, lo demás es poco significativo.

Emilio, en cambio, no considera que el precio sea determinante de las ventas. Por su parte, él sostiene que el acceso a los medios de comunicación electrónicos (páginas web, tiendas virtuales, redes sociales, entre otros) ha hecho cambiar muy rápidamente la forma de considerar a los libros por parte de sus potenciales compradores, por lo que estaría dispuesto a afirmar que, en algunas circunstancias, incluso podría darse el caso de que las ventas de libros aumenten ante incrementos en el precio. Además, agrega que el precio de la competencia y la renta personal de los clientes, también son factores explicativos de dichas ventas.

De este modo, la historieta parte de dos hipótesis fundamentales que el instrumental econométrico debe resolver:

- *Hipótesis 1:* Según Don Pablo, el precio es la variable determinante de las ventas de la editorial, no haya otra igual.
- *Hipótesis 2:* De acuerdo a Emilio, aunque el precio explica en cierto modo las ventas de la editorial, el precio de la competencia, el gasto en publicidad y la renta personal de los clientes son también factores significativos, los cuales deben ser incorporados al modelo explicativo.

Debido a que ninguno de los dos está dispuesto a ceder en sus opiniones, deciden buscar la manera de dirimir la cuestión basándose en la información que tienen sobre las ventas de la editorial. Emilio recuerda que en su época de estudiante de economía, recibió un curso de econometría; sus profesores *Alfonso, Fernández & Gallastegui (2005)* en aquel entonces exponían la necesidad de relacionar variables económicas para determinar los efectos

² Como la librería vende muchos volúmenes a diferentes precios según las características específicas de cada libro, para efectos de simplificación del análisis se estima un precio medio como resultado de sumar las ventas totales anuales y dividir las sobre el número de ejemplares vendidos en cada año.

causales de una o más variables explicativas sobre una variable dependiente y de este modo demostrar o no, los efectos o cambios provocados por una variable sobre otra. Este tipo de recuerdos, un tanto vagos, despierta en Emilio, el interés de desempolvar su biblioteca de economía; pero antes de empezar con el estudio de los temas de su interés y decidido a derrumbar la hipótesis de su padre, considera importante comprender el concepto de econometría para luego explicar a su padre, cómo el uso de esta herramienta puede ayudar a contrastar dos opiniones hasta el momento divergentes y explicar cuál de los dos tiene la razón.

Aunque Emilio encuentra en su colección de libros, varios manuales de econometría (unos muy simples y abstractos y otros sumamente rigurosos y avanzados), le llama la atención, uno en especial. Se trata del libro titulado: *ECONOMETRÍA* de sus maestros *Alonso, Fernández & Gallastegui*, catedráticos de econometría de la Universidad del País Vasco (España). Para Emilio, este libro toca casi todos los aspectos de la econometría básica e intermedia y está escrito en un lenguaje accesible y ameno; además, de ser la fuente de inspiración de esta historieta que se desarrolla a lo largo del presente texto, con la que se pretende que el lector apropie con mayor facilidad los aspectos fundamentales de la econometría básica.

Emilio empieza su exposición diciendo que no hay un concepto o definición de econometría única. Como lo señalan jocosamente Stock y Watson (2012:1):

“Pregunte a media docena de económetras qué es la econometría y obtendrá media docena de respuestas diferentes”.

También le parece interesante la definición de Maddala (1996:1) quien afirma que:

“la econometría es la aplicación de métodos estadísticos y matemáticos al análisis de datos económicos con el propósito de dar contenido empírico a las teorías económicas y verificarlas o refutarlas”.

Removiendo el baúl de los recuerdos, Emilio encuentra una definición suficientemente completa de la econometría que ofrece Judge et al, (1988), apuntando que:

La econometría, utilizando teoría económica, economía matemática e inferencia estadística como fundamentos analíticos y los datos como fuente de información, proporciona a la ciencia económica una base para: 1) modificar, refinar o posiblemente refutar las conclusiones contenidas en el cuerpo de conocimientos, conocido como teoría económica; y 2) conseguir signos, magnitudes y proposiciones fiables acerca de los coeficientes de las variables en las relaciones económicas, de modo que esta información pueda servir de base para la toma de decisiones y la elección (p. 1).

Como puede observarse, Emilio encuentra en los libros de texto diversas definiciones de econometría, desde la definición etimológica (econometría = medición económica) hasta otras más sofisticadas, pero se inclina por la definición incluida en el libro de sus profesores *Alonso, Fernández & Gallastegui (2005:2)*, la cual se considera incluyente, sencilla, concisa, de fácil memorización y que de manera simple integra los aspectos fundamentales de la econometría:

“La econometría se ocupa de formular relaciones entre variables económicas, cuantificarlas y valorar los resultados obtenidos”.

Por su parte, Gujarati (2010) concluye que “la teoría económica hace afirmaciones o formula hipótesis de naturaleza sobre todo cualitativa. Por ejemplo, la teoría microeconómica establece que, si no intervienen otros factores, se espera que la reducción del precio de un bien aumente la cantidad demandada de ese bien. Así, la teoría económica establece una relación negativa o inversa entre el precio y la cantidad demandada de un bien. Pero la teoría por sí sola no proporciona medida numérica alguna de la relación entre los dos; *no dice cuánto aumentará o se reducirá la cantidad como resultado de un cambio determinado en el precio del bien*. El trabajo del econométrico es proporcionar tales estimaciones numéricas. En otras palabras, la econometría da contenido empírico a gran parte de la teoría económica” (p. 2). Esta última frase despierta el interés de Emilio, donde a la econometría se le otorga un gran papel.

Adentrándonos un poco más en el tema de discusión, en la hipótesis de Don Pablo, existe una relación estrecha entre las ventas de la editorial y el precio de los libros, y en la opinión de su hijo Emilio, la relación entre las ventas y el precio de la competencia, el gasto en publicidad y el ingreso personal es

mucho más estrecha que la primera; se advierte que Emilio no desconoce que el precio ejerce algún efecto sobre las ventas de libros. Aunque los dos perciben la existencia de relaciones entre las variables económicas en cuestión, hasta el momento no han logrado cuantificar y valorar dichas relaciones. Ante esto, Emilio le comenta a su padre que la econometría precisamente proporcionará la respuesta a favor o en contra de las afirmaciones de cada uno. Las palabras de Emilio, en realidad, no logran convencer aún a Don Pablo, quien se sostiene firmemente en su opinión inicial.

Si hay algo que caracteriza a Emilio, es su carácter y su espíritu de búsqueda de la verdad. Él tiene un reto muy grande: demostrar que su padre está equivocado en su apreciación. Para ello decida armar una argumentación sólida en torno a la bondad de la investigación econométrica.

1.2 Componentes constitutivos de la econometría.

La econometría en su función de formular relaciones entre variables económicas, cuantificarlas y valorar los resultados obtenidos como se verá enseguida, fusiona tres elementos esenciales para darle consistencia al análisis económico, tales como: 1) Teoría económica, 2) datos y 3) estadística inferencial, como se muestra en la Tabla 1.1.

Tabla 1.1.
Elementos que conforman la econometría.

Elementos	Descripción
Teoría económica	Se ocupa del análisis de la economía, utilizando el método científico, con el fin de establecer leyes generales en el comportamiento económico que se traducen en la existencia de relaciones estables entre variables relevantes de la economía. En el caso que nos ocupa, la relación negativa o inversa entre las ventas de libros (V_t) y el precio promedio (P_t) de los ejemplares, según la opinión de Don Pablo. Este relacionamiento de variables se halla soportado en la ley de la demanda antes explicada.

Datos	El análisis que se realiza se refiere a una realidad concreta, representada por los valores que se hayan observado para las diferentes variables involucradas en el análisis. En este caso, se cuenta con información estadística de ventas y precios promedio con frecuencia trimestral a lo largo del tiempo (series temporales) para el periodo 2009-2018.
Estadística	Proporciona una estructura básica de métodos de tratamiento de datos que permiten tanto cuantificar las relaciones entre las variables de manera adecuada como valorar los resultados de acuerdo con estándares establecidos. En este caso, inicialmente se considera el uso de la técnica de mínimos cuadrados ordinarios como la más adecuada.

Fuente: Elaboración propia.

Emilio uniendo la esencia de los tres elementos arriba enunciados, llega a la misma definición de sus maestros Alonso et al, (2005:2):

“La econometría se ocupa de formular relaciones entre variables económicas, cuantificarlas y valorar los resultados obtenidos”.

Emilio necesita convencer a su padre de que la técnica propuesta, es en verdad confiable y contundente. Para lograr este cometido, de manera ingeniosa, a la definición anterior, le da un matiz mucho más aplicado y convincente. Él sabe que su padre como hombre pragmático es poco amante de los procedimientos estadísticos y de la rigurosidad científica, por ello, decide simplificar el concepto de econometría del siguiente modo:

¡Padre!, en el caso que nos ocupa y ante la necesidad de conocer quién de los dos tiene la razón, te propongo que se resuelva la validez de cada una de nuestras hipótesis, a partir del relacionamiento estadístico de las ventas de los libros con el precio promedio de los mismos, a fin de valorar el efecto que ejerce el precio promedio sobre las ventas totales. Por favor, proporcióname la información (base de datos) de las variables en cuestión, y yo haré el análisis pertinente. Posteriormente, se discutirá los resultados; espero que al final de este ejercicio, puedas convencerte de que tu tesis es equivocada.

Como puede verse, Emilio cuenta con los elementos fundamentales para contrastar la hipótesis de su padre. Primero, tiene a la mano una teoría económica ampliamente aceptada (la ley de la demanda); segundo, la editorial cuenta con una base de datos histórica con información de ventas trimestrales y de precios promedio por ejemplar; y tercero, dispone de un *modelo econométrico* y de un método estadístico robusto como mínimos cuadrados ordinarios (MCO) para cuantificar el efecto del precio sobre las ventas de los libros de la editorial.

Don Pablo escuchando la exposición de su hijo Emilio, siente una confusión con respecto a los términos modelo económico, modelo matemático y modelo econométrico. Para él estos conceptos comparten algo en común: son una perdedera de tiempo y demasiado esfuerzo intelectual; sin embargo, a regañadientes acepta que su hijo le explique de manera muy didáctica en qué consiste la modelación econométrica.

Recordando a sus mencionados profesores, Emilio empieza su exposición, hilando tres elementos:

Elemento 1: Teoría económica

Emilio le explica a su padre que la lógica más sencilla indica que existe una relación inversa o negativa entre las ventas y el precio, que se puede representar a través de la siguiente función:

$$\text{Ventas} = f(\text{Precio})$$

(-)

Donde $f(\cdot)$ es una función genérica y el signo (-) indica que, de acuerdo con la teoría, las ventas crecen si el precio se reduce y viceversa.

Emilio con el desarrollo de este paso, ha formalizado un modelo económico que resume el comportamiento de las ventas de la editorial, pero con poca información aportante. Indudablemente, Don Pablo requiere un modelo específico que establezca cómo reaccionan las ventas de la editorial ante cambios en el precio promedio de los libros.

Emilio recuerda una vez más las enseñanzas de sus maestros *Alonso, Fernández & Gallastegui (2005)* quienes afirman que creer que la variable explicativa Precio promedio de la editorial (PE_i) se relaciona de manera estable con las ventas (V_i) significa pensar en una relación en la que los coeficientes permanezcan constantes durante el periodo de análisis. De otro lado, introducir datos numéricos y buscar igualdades supone obtener valores cambiantes para esos coeficientes. La solución se encuentra añadir al modelo un nuevo elemento, al que se denotará μ_i llamado término de error estocástico o perturbación. Emilio comprende que ante las relaciones inexactas entre las ventas (V_i) y el precio promedio (PE_i), el modelo económico anterior se ha transformado en un modelo más realista que el anterior; se está precisamente ante el denominado *modelo econométrico*.

Anatomía de un modelo econométrico general

$$y_i = \underbrace{\beta_0 + \beta_1 X_i}_{\text{(Parte sistemática)}} + \underbrace{\mu_i}_{\text{(Perturbación estocástica)}}$$

Dónde:

y_i : Es la variable que se quiere explicar, también denominada variable dependiente, endógena, explicada, regresada o respuesta.

X_i : Es la variable utilizada para explicar las variaciones de la variable y_i . También se le denomina variable explicativa, independiente, exógena, regresora o de control.

β_0 y β_1 : Representan los coeficientes (o parámetros) que determinan la relación entre las variables (ventas y precio promedio). Son constantes y desconocidos, por lo tanto, se tendrá que estimarlos mediante métodos econométricos.

μ_i : Es la variable que recoge el resto de efectos presentes en los datos muestrales no recogidos por las variables explicativas del modelo, también se le llama perturbación aleatoria o término de error.

i : Representa una observación cualquiera del conjunto de observaciones presentes en la muestra: $i = 1, 2, 3, \dots, n$.

La parte sistemática o determinista ($\beta_0 - \beta_1 X_i$) expresa el valor esperado (o media) de la variable dependiente Y condicionada al valor que tome la variable explicativa X .

La perturbación o término de error (μ_i) es una variable aleatoria que incorpora toda aquella información que no recoge la variable explicativa (X), es decir, aquello que provoca que una observación o un dato se alejen del valor esperado de Y para un valor concreto de X .

¿Por qué es importante incluir en un modelo el término de error? La perturbación³ (μ_i) puede contener: a) otras variables explicativas, que aun siendo relevantes no se encuentran especificadas en el modelo de regresión simple, por ejemplo, el precio de la competencia, el gasto en publicidad, la renta de los clientes y otras más; b) no se dispone de observaciones muestrales de otras variables X ; y c) refleja errores de medida en la variable dependiente Y .

Ante la argumentación anterior, Emilio decide incluir en su modelo la parte aleatoria o término de perturbación (μ_i). Él sabe que la relación entre las variables económicas (por ejemplo, relación entre las ventas y el precio promedio) es inexacta (no determinista), por lo tanto, es consciente de que las ventas de la editorial no dependen exclusiva y absolutamente del precio de los libros, es decir, hay otras variables como las enumeradas arriba que también explican las variaciones en las ventas. Precisamente, en la perturbación se recogerán los efectos de dichas variables no contempladas, por ahora, en el modelo.

³ Wooldridge (2015: 55) resalta que existe una diferencia entre los errores (perturbaciones) y los residuales, aspecto crucial en la construcción de un estimador de σ^2 . La ecuación $y_i = \beta_0 + \beta_1 X_i + u_i$; $i = 1, 2, \dots, n$, muestra cómo escribir el modelo poblacional en términos de una observación muestral aleatoria como $y_i = \beta_0 + \beta_1 X_i + u_i$ donde u_i es el error en la observación i . También se puede expresar y_i en términos de su valor ajustado y su residual como en la ecuación $y_i = \hat{y}_i + u_i$, así: $y_i = \beta_0 + \hat{\beta}_1 X_i + u_i$. Comparando estas dos ecuaciones, se observa que el error aparece en la ecuación que contiene los parámetros poblacionales β_0 y β_1 . Por otro lado, los residuales aparecen en la ecuación estimada con $\hat{\beta}_0$ y $\hat{\beta}_1$. Los errores jamás son observables, mientras que los residuales se calculan a través de los datos.

A la memoria de Emilio, regresa nuevamente un concepto explicado por sus influyentes profesores Alonso et al., (2005):

(...) en la perturbación se incluye todo lo que no es estable, ni sistemático, ni predecible, sino errático” En la perturbación se puede considerar: i) los efectos que influyen ligeramente en las ventas de cada periodo pero que no están explícitamente recogidos por el modelo, ii) pequeñas fluctuaciones de los datos y iii) efectos no sistemáticos que no pueden anticiparse cuándo se producirán y ni en qué cuantía. (p. 6).

1.3 Diferencias entre un modelo matemático y un modelo econométrico.

Don Pablo siente mucha confusión cuando Emilio habla de modelos, término recurrente que parece complicarlo todo. Don Pablo decide poner a prueba a su hijo, pidiéndole una explicación sencilla y breve. Mirando fijamente a su hijo, le pregunta: Hijo, tú que tanto hablas de modelación económica, dime: ¿cuáles son las diferencias entre un modelo matemático y un modelo econométrico? Por favor, evita los tecnicismos en tu explicación.

Emilio en atención a la petición de su padre, empieza por un concepto general de modelo diciendo lo siguiente: mira papá, cuando una pareja de recién casados decide construir una vivienda, la cual albergará en un futuro a la familia, inicia contratando los servicios de un arquitecto, cuya función principal será el diseño de la casa contenido en un conjunto de bocetos, planos y en una maqueta (si se requiere). Como la casa que se pretende construir tiene un volumen apreciable, el arquitecto elige una escala disminutiva para la elaboración de los planos y la maqueta, es decir, no es relevante para él, ni para los futuros propietarios, observar el diseño de la casa en tamaño natural (escala 1:1). El uso de los planos y la maqueta a una escala apreciable, ayuda a los dueños de la vivienda y al constructor que se elija, a comprender los aspectos más relevantes de la futura edificación, sin entrar en detalles poco irrelevantes como la textura de los acabados y enchapes, los colores de las paredes, las marcas de los accesorios, la decoración, la localización de los diferentes artefactos dentro la vivienda, etcétera.

El modelo económico es una simplificación vaga de una realidad muy compleja cargada de multiplicidad de elementos, detalles e interrelaciones. Como bien lo define Varian (2016), “el poder de un modelo se deriva de la supresión de los detalles irrelevantes, que permite al economista fijarse en los rasgos esenciales de la realidad económica que intenta comprender” (p. 1). El modelo, como el caso de los planos y de la maqueta de la casa, ayuda a comprender una realidad, tomando únicamente los aspectos más sobresalientes de dicha realidad y descartando a aquellos que se consideren redundantes. En este caso, el arquitecto sólo se ocupa de las dimensiones y el número de espacios y su ubicación, tratando de optimizar la disponibilidad de metros cuadrados, los aislamientos para la generación de iluminación natural, los elementos de ventilación, el diseño de la fachada; por citar algunos.

Don Pablo percibe que Emilio no ha sido lo suficientemente claro en su explicación, por tal razón, en tono enérgico, le pide ir al grano de una vez por todas, dejando a un lado lo intrascendente.

Emilio responde: Papá, solo quiero que entiendas de manera didáctica la importancia y aplicación de un modelo económico. Emilio, prosigue con la explicación argumentando lo siguiente: Un modelo matemático establece una relación exacta (determinista) entre las variables utilizadas; esto equivaldría a pensar y a aceptar que las ventas de la editorial dependen única y exclusivamente de los precios de los libros, por lo tanto, el modelo matemático de la editorial, sería:

$$V_t = \beta_0 + \beta_1 PE_t \quad (1.1)$$

Dónde: V_t = Ventas, PE_t = Precio promedio de la editorial, β_0 y β_1 coeficientes o parámetros del modelo (β_0 es el intercepto y β_1 el coeficiente de la pendiente de la variable Precio). Nótese que en el modelo (1.1) no aparece ningún término de error, es decir, no existe una parte sistemática que recoja los efectos no sistemáticos.

Dada la exactitud de un modelo matemático, esto implicaría que la predicción de las ventas utilizando el precio como variable explicativa tendría una absoluta precisión, lo cual en la vida real, es totalmente absurdo. Por lo

tanto, este tipo de modelos son de uso exclusivo de las ciencias exactas como la física y la astronomía, las cuales utilizan datos experimentales.

En las ciencias sociales como la economía⁴, el uso de modelos matemáticos para relacionar variables es prácticamente imposible, de ahí la necesidad de utilizar para el caso de la economía, los modelos econométricos, los cuales son variados y de diversa complejidad.

Los modelos econométricos aceptan la inexactitud existente en las relaciones entre variables económicas, por ende, existe una probabilidad de ocurrencia de los pronósticos. En el caso de la editorial, las ventas dependen en alguna medida del precio de los libros, pero se acepta que existen también otras variables determinantes. El éxito en la especificación de dicho modelo estaría en la identificación de las variables más relevantes o con mayor poder explicativo.

Emilio muy satisfecho por la contextualización realizada, en tono sonriente, le responde a Don Pablo: Las ventas de la editorial no podrían ser explicadas a través de ningún modelo matemático; por defecto, debería utilizarse un modelo econométrico univariado (con dos variables) o multivariado (con tres o más variables), como los siguientes:

a. Modelo bivariado (regresión simple):

$$V_t = \underbrace{\beta_0 + \beta_1 PE_t}_{\text{Parte sistemática}} + \underbrace{\mu_t}_{\text{Parte aleatoria}} \quad t = 1, 2, 3, \dots, T \quad (1.2)$$

⁴ La economía estudia el comportamiento o la conducta humana en su dimensión económica. Ante la complejidad en la comprensión de la conducta económica de los agentes y la imposibilidad de realizar experimentos controlados en laboratorio (de uso cotidiano de la física, química y biología), la economía recoge, clasifica, organiza, procesa y analiza datos observacionales extraídos del comportamiento humano, los cuales le permiten hacer aproximaciones en la comprensión de los fenómenos económicos. Desde esta perspectiva, para muchos estudiosos, la economía no es una ciencia en el sentido estricto; se asemeja más a una disciplina en constante desarrollo, actualmente influenciada por el excesivo uso de las matemáticas; las cuales como lo afirman Ekelund y Hébert (2005), distraen inevitablemente la atención de las verdades básicas del proceso económico desarrollado por Adam Smith, considerado el precursor de la economía moderna. Como en economía es imposible utilizar datos experimentales, para la modelación se recurre al uso de datos observacionales.

Dónde: V_t son las Ventas (variable dependiente), PE_t es el Precio promedio de la editorial (variable explicativa), β_0 es el intercepto, β_1 es el coeficiente de la variable Precio de la editorial (PE_t) y μ_t es el término de perturbación. Este es un ejemplo de modelo de regresión lineal, donde la función econométrica de las ventas de la editorial plantea como hipótesis que la variable dependiente Ventas (V_t) está relacionada linealmente con la variable explicativa Precio de la editorial (PE_t), pero que la relación entre las dos variables (*ventas y precio*) no es exacta, por consiguiente, dicha relación está sujeta a variaciones individuales, las cuales son recogidas en la parte aleatoria μ_t .

b. Modelo multivariado (regresión múltiple):

$$V_t = \underbrace{\beta_0 + \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t}_{\text{Parte sistemática}} + \underbrace{\mu_t}_{\text{Parte aleatoria}} \quad t = 1, 2, 3, \dots, T \quad (1.3)$$

Donde: Corresponde a las ventas de la editorial; β_0 es el intercepto; β_1 , β_2 , β_3 y β_4 son los parámetros del modelo; PE_t , PC_t , GP_t y R_t son las variables explicativas y μ_t es el término de error. Aunque las ventas de la editorial están relacionadas linealmente con el precio promedio de la editorial (PE_t), el precio de la competencia (PC_t), el gasto en publicidad (GP_t) y la renta personal de los clientes (R_t); esta relación no es exacta, por consiguiente, dicha relación está sujeta a variaciones individuales, las cuales son recogidas en la parte aleatoria μ_t . De hecho, existen otras variables que explican las ventas de la editorial como el sexo, la edad, el nivel de estudios de los clientes, etcétera que podrían ser incluidas en este modelo.

Don Pablo replica: Ahora, si comprendo la diferencia entre los dos modelos, así como también la validez del modelo econométrico y de su poder explicativo. Papá, por favor, observa el diagrama de dispersión (véase Figura 1.1); la recta con pendiente negativa se abre camino en medio de los diversos puntos que corresponden a las parejas ($X = \text{Ventas}$ y $Y = \text{Precio promedio}$). Hay puntos localizados exactamente sobre la recta, pero también hay otros más cercanos y más alejados, tanto por encima de la recta de regresión (residuos positivos) como por debajo (residuos negativos) de la misma. La distancia entre cada punto y la recta corresponde a una perturbación (μ_t) o

variación individual de cada pareja de puntos (X, Y) , entre más larga sea la distancia, mayor será el error. Si todos los puntos se posaran exactamente sobre la recta, la relación entre el precio y las ventas sería totalmente exacta (esto es poco probable en el análisis económico) con un ajuste perfecto entre las dos variables y un poder de predicción perfecto.

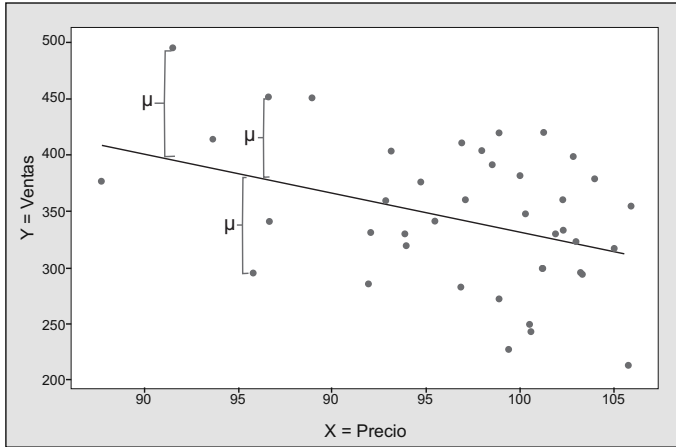


Figura 1.1. Relación entre ventas y precio promedio

Elemento 2: Datos

Emilio sabe que el modelo específico que debe explicar a su padre requiere de información concreta expresada en las ventas de la editorial en un periodo de tiempo determinado y del precio promedio de los ejemplares vendidos. En síntesis, Emilio requiere un modelo específico que se adapte a los datos disponibles.

$$\text{Ventas} = f(\text{Precio}) \quad t = 2009:1, \dots, 2018:4$$

El subíndice t indica el momento en que se considera la relación (en el caso que nos ocupa, cada trimestre).

Don Pablo proporciona a Emilio inicialmente la siguiente información contenida en la Tabla 1.2 para que pueda construir su famoso modelo econométrico, y de una vez por todas, conocer quién tiene la razón. Esta información corresponde a una serie temporal de ventas y precio promedio del periodo 2009...2018 organizada por trimestres.

Tabla 1.2***Ventas trimestrales de libros 2009-2018 (pesos constantes de 2009).***

Trimestre	Ventas (millones de \$)	Precio promedio por unidad* (miles de pesos)	Trimestre	Ventas (millones de \$)	Precio promedio por unidad* (miles de pesos)
2009:1	275,5	98,6	2014:1	229,9	99,2
2009:2	285,6	96,6	2014:2	342,9	86,4
2009:3	336,8	102,0	2014:3	361,9	92,7
2009:4	333,4	91,9	2014:4	301,3	100,9
2010:1	357,0	105,6	2015:1	332,5	101,7
2010:2	325,2	102,8	2015:2	343,2	95,3
2010:3	362,2	102,1	2015:3	421,9	98,6
2010:4	232,0	99,1	2015:4	401,9	102,5
2011:1	252,2	100,3	2016:1	421,8	101,0
2011:2	322,1	104,8	2016:2	361,7	96,9
2011:3	297,3	85,6	2016:3	393,6	98,3
2011:4	298,9	103,0	2016:4	287,6	91,8
2012:1	246,8	100,4	2017:1	380,0	77,8
2012:2	322,4	93,8	2017:2	495,6	81,6
2012:3	383,5	99,7	2017:3	452,6	86,6
2012:4	321,8	104,6	2017:4	404,9	93,0
2013:1	351,5	100,0	2018:1	421,3	83,7
2013:2	381,1	103,7	2018:2	333,7	93,7
2013:3	412,5	96,7	2018:3	379,1	94,5
2013:4	217,5	105,5	2018:4	407,0	97,7

* Como la librería vende muchos volúmenes a diferentes precios según las características específicas de cada libro, para efectos de simplificación del análisis se estima un precio medio como resultado de sumar las ventas anuales total y dividir las sobre el número de ejemplares vendidos en cada año.

Fuente: El autor con base en la información proporcionada por Alonso et al., (2005:14).

Con la información disponible, Emilio procede a graficar el comportamiento de las variables en el tiempo. Las Figuras 1.2 y 1.3 sugieren a simple vista, que en los periodos en los que las ventas de la editorial se incrementaron, el precio promedio de los libros experimentó una disminución, lo que deja entrever la existencia de una relación inversa entre estas dos variables. Por su parte, la Figura 1.4 confirma la dirección en que se mueven las dos variables

en cuestión, los aumentos o disminuciones de las ventas de la editorial giran en torno a los aumentos o disminuciones de los precios de los libros. Emilio se pregunta: ¿Será que mi papá tiene razón? En esencia, ¿el precio determina las ventas de la editorial? Partiendo de esta intuición, si la editorial quiere incrementar sus ventas, necesariamente debería bajar los precios de los libros. ¿Esta discusión será así de sencilla?

Figura 1.2. Ventas de la editorial
2009-2018

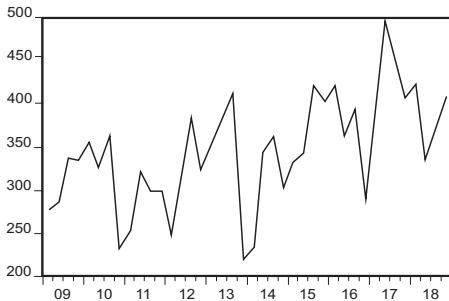


Figura 1.3. Precio promedio
2009-2018

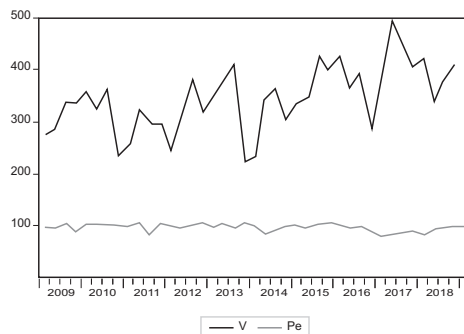
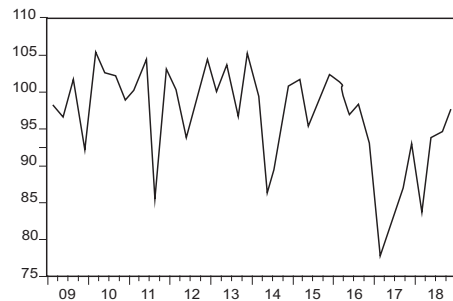


Figura 1.4. Ventas y precio promedio
2009-2018

Elemento 3: Estadística

La cuantificación de los valores de los coeficientes del modelo se hace mediante la inferencia estadística como lo señala Alonso et. al., (2005), además, de permitir obtener un valor numérico para cada uno de los coeficientes, proporcionará, información en términos probabilísticos, sobre la proximidad o lejanía que se puede tener respecto de los valores que se quiere conocer,

por ejemplo, el coeficiente β_1 que acompaña la variable precio promedio estimado a partir de la información muestral disponible (2009:1...2018:4), en términos probabilísticos nos permitirá conocer si se está próximos o lejanos al verdadero valor. Además, el instrumental estadístico facilita elegir los procedimientos más adecuados para la obtención de los mejores estimadores, es decir, a aquellos que expliquen y predigan mejor la variable dependiente (por ejemplo, las ventas de la editorial) en su media.

1.4 Relaciones estadísticas y relaciones deterministas.

También quedó claro en el capítulo anterior, la diferencia entre un modelo matemático y un modelo econométrico. Para Gujarati (2010:19), “en el análisis de regresión interesa lo que se conoce como dependencia *estadística* entre variables, no así la *funcional* o *determinista*, propia de la física clásica (por ejemplo, la ley de la gravedad de Newton, la ley de Ohm, ley de los gases de Boyle, la ley de la electricidad de Kirchhoff y la ley del movimiento de Newton). En las relaciones estadísticas entre variables se analizan, en esencia, variables aleatorias o estocásticas⁶, es decir, variables con distribuciones de probabilidad. Por otra parte, en la dependencia funcional o determinista también se manejan variables, pero no son aleatorias o estocásticas”.

En nuestro caso, las ventas de la editorial dependen del precio al que vende la editorial, el precio de la competencia, el gasto en publicidad, entre otras. Parafraseando a Gujarati (2010), dicha dependencia es de naturaleza estadística porque las variables explicativas, si bien son importantes, no permiten al analista de mercados predecir en forma exacta las ventas de libros en un periodo determinado debido a los errores propios de la medición de estas variables y a otros factores que en conjunto afectan las ventas pero que son difíciles de identificar individualmente. De este modo, habrá alguna variabilidad “intrínseca” o aleatoria en la variable dependiente (ventas) que no

⁶ La palabra *estocástico* viene de la voz griega *stokhos*, que significa “centro del blanco”. El resultado de lanzar dardos sobre un tablero es un proceso estocástico, es decir, un proceso lleno de tiros fallidos. (Gujarati, 2010: 19).

puede explicarse en su totalidad sin importar cuántas variables explicativas se consideren en el modelo; es decir, a pesar de todas las variables explicativas que se tomen siempre existirá errores de medición (residuos). En conclusión, en economía como en otras ciencias sociales, las relaciones que se establecen entre variables serán siempre de tipo estadístico (probabilísticas).

1.5 Regresión y causalidad.

Gujarati (2010), señala que "a pesar de que el análisis de regresión tiene que ver con la dependencia de una variable respecto de otras, esto no implica causalidad necesariamente. En palabras de Kendall y Stuart: "Una relación estadística, por más fuerte y sugerente que sea, nunca podrá establecer una conexión causal: nuestras ideas de causalidad deben provenir de estadísticas externas y, en último término, de una u otra teoría" (p. 19). "Para aducir causalidad se debe acudir a consideraciones *a priori* o teóricas" (p. 20). En el caso de las ventas de la editorial, es posible acudir a la teoría económica (ley de la demanda) para afirmar que las ventas de la editorial dependen del precio de los libros. El análisis clásico de regresión se basa en el supuesto de que el modelo del análisis es el correcto, por consiguiente, la dirección de la causalidad puede estar implícita en el modelo postulado. La relación de causalidad entre la variable endógena y las variables explicativas es unidireccional, es decir, las variables explicativas pueden influir en la variable endógena, pero no a la inversa.

Según Wooldridge (2015:12), en la mayoría de las pruebas de teorías económicas, así como en la evaluación de políticas públicas, el objetivo de los economistas es inferir que una variable (por ejemplo, el precio de los libros) tiene un efecto causal sobre otra variable (por ejemplo, las ventas de la editorial). Encontrar una relación entre dos o más variables puede ser sugestivo, pero no concluyente, a menos de que pueda establecerse causalidad. "En las ciencias sociales, dado el carácter no experimental de la mayor parte de los datos que suelen recolectarse, hallar relaciones causales no es una tarea fácil" (p. 16).

1.6 La causalidad y la noción de *ceteris paribus* en el análisis econométrico.

Como bien lo precisa Wooldridge (2015), la noción *ceteris paribus* – “si todos los demás factores relevantes permanecen constantes” desempeña un papel importante en el análisis causal, puesto que la mayor parte del análisis económico tiene un carácter *ceteris paribus*. Por ejemplo, cuando se analiza las ventas de la editorial interesa conocer el efecto que tiene una modificación en el precio de los libros sobre las ventas totales, mientras todos los demás factores – tales como: el precio de la competencia, el gasto en publicidad y el ingreso personal – se mantienen constantes. Si no permanecen constantes los demás factores, entonces no se puede saber cuál es el efecto de una modificación en el precio sobre la cantidad demanda. Si se logran mantener constantes todos los demás factores relevantes y se encuentra una relación entre el precio promedio de los libros y las ventas de la editorial, puede concluirse que el precio tiene un efecto causal sobre las ventas.

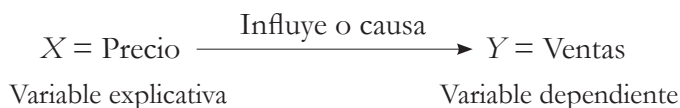
1.7 Regresión y correlación.

Existe una relación estrecha entre el análisis de correlación y el análisis de regresión, aunque estos dos conceptos son diferentes entre sí. El análisis de correlación a través del coeficiente de correlación mide la fuerza o el grado de asociación lineal entre dos variables, mientras que el análisis de regresión trata de estimar o predecir el valor promedio de una variable dependiente con base en los valores fijos de otras variables explicativas.

Aplicando estos dos conceptos al caso de la editorial, se diría que el análisis de correlación se interesa por medir a través del coeficiente de correlación, el grado de asociación lineal existente entre las ventas de la editorial (Y) y el precio promedio de los libros (X). Como la correlación no mide ninguna relación causal, el orden de las variables se puede invertir (no importa su orden), por tanto, la correlación también podría medir el grado de asociación lineal existente entre el precio promedio de los libros (X) y las ventas de la editorial (Y). En los dos casos, el resultado es exactamente igual.

$$Y \rightarrow X \text{ o también } X \rightarrow Y$$

Por su parte, el análisis de regresión centra su interés en medir la relación causal entre las ventas y el precio promedio (qué tanto el precio promedio de los libros explica las ventas de la editorial) y mediante una estimación de la recta de regresión, se predecirá para ciertos periodos de tiempo, el valor promedio de la ventas con base en valores fijos del precio promedio, por ejemplo, si el precio promedio de los libros aumenta/disminuye en un 10%, las ventas disminuirán/aumentarán en un 14%. El cálculo del coeficiente de correlación y la estimación de la recta de regresión lineal le permitirán a la editorial fijar una política de precios en aras de optimizar sus beneficios. Se debe considerar, que en el análisis de regresión el orden de las variables es fundamental; en nuestro caso, el precio promedio de los libros (X) explica en forma parcial el volumen de ventas de la editorial (Y), si se invierten las variables, el resultado también se modifica.



Nótese que al invertir el orden de las variables, la dirección de la relación causal también se altera, y por ende, los valores de los coeficientes y la interpretación de estos. Aquí radica la gran diferencia entre la correlación y la regresión. A este respecto, también se afirma:

(...) en el análisis de regresión hay una asimetría en el tratamiento a las variables dependientes y explicativas. Se supone que la variable dependiente es estadística, aleatoria o estocástica, es decir, que tiene una distribución de probabilidad. Por otra parte, se asume que las variables explicativas tienen valores fijos (en muestras repetidas⁷). En el análisis de correlación, por otra parte, se tratan dos variables cualesquiera en forma simétrica; no hay distinción entre las variables dependiente y explicativa. (...). Además, las dos variables se consideran aleatorias. Como se puede ver, la mayor parte de la teoría de correlación parte del supuesto de aleatoriedad de las variables, mientras que la mayor parte de la teoría de regresión que se expondrá en este texto está condicionada al supuesto de que la variable dependiente es estocástica y que las variables explicativas son fijas o no estocásticas". (Gujarati, 2010:20)

⁷ Es de crucial importancia notar que las variables explicativas pueden ser intrínsecamente estocásticas, pero, para fines del análisis de regresión, suponemos que sus valores son fijos en el muestreo repetido (es decir, que X toma los mismos valores en diversas muestras), de modo que, en efecto, no resultan aleatorias ni estocásticas). Citado en Gujarati (2010: 20).

Para el caso de la editorial, la variable dependiente ($Y =$ Ventas de la editorial) es estadística, aleatoria o estocástica, mientras que la variable explicativa ($X =$ Precio promedio) tiene un valor fijo (en muestras repetidas). Por cuestiones de causalidad cada variable ocupa una posición específica, por ejemplo, el precio promedio (variable explicativa = X) es una variable determinante de las ventas de la editorial (variable dependiente = Y), es decir, existe una relación de X hacia Y ($X \rightarrow Y$), por ende, no es posible invertir la relación entre estas variables sin que se modifique la ecuación de regresión estimada (al invertir la relación, la ecuación de regresión se transforma, al igual que su interpretación). En cambio, en el análisis de correlación, las ventas como el precio promedio se tratan como variables aleatorias, por consiguiente, es indiferente medir el grado de asociación lineal entre las ventas y el precio promedio o viceversa; el coeficiente de correlación siempre dará el mismo resultado.

Precio promedio (P) \rightarrow Ventas de la editorial (V) = Ventas de la editorial (V) \rightarrow Precio promedio (P)
Precio promedio (P) \leftrightarrow Ventas de la editorial (V)

1.8 Denominación de las variables en un modelo econométrico.

Don Pablo ha escuchado de boca de Emilio, usar en repetidas veces las palabras variable *dependiente* y variable *independiente o explicativa*, pero necesita se le aclare por qué las llama de diversas maneras. Emilio utilizando el modelo econométrico de las ventas de la editorial le responde a su padre: “los términos *variable dependiente* y *variable explicativa* se definen de varias maneras”, siendo todas sinónimos. A continuación en la Tabla 1.3 se presenta una lista de tales sinónimos.

Tabla 1.3
Denominación de las variables.

Ventas editorial = Y	Precio = X
Variable dependiente	Variable explicativa
Variable explicada	Variable independiente
Predicha	Predictora
Regresada	Regresora
Respuesta	Estímulo
Endógena	Exógena
Resultado	Covariante
Variable controlada	Variable de control

Fuente: Gujarati (2010:21).

En el presente documento y con el deseo de no confundir al lector, se utiliza la terminología de variable dependiente/variable explicativa y, en algunas ocasiones, a la variable explicativa también se la denominará regresor(a). Si se estudia la dependencia de una variable respecto de una única variable explicativa, como por ejemplo, *las ventas que dependen del precio*, dicho estudio se conoce como análisis de regresión *simple*, o con dos variables (las *ventas* como variable dependiente y el *precio promedio* como variable explicativa). Sin embargo, si se estudia la dependencia de una variable respecto de más de una variable explicativa, como las ventas en función del precio de la editorial, el precio de la competencia, el gasto en publicidad y la renta personal, y posiblemente otras; se trata de un análisis de regresión múltiple. En otras palabras, en una regresión simple sólo hay una variable explicativa, mientras que en la regresión múltiple se utiliza más de una variable explicativa. El análisis de regresión múltiple es el más utilizado en investigación, precisamente por la multiplicidad de factores que influyen o explican un determinado fenómeno económico.

1.9 Etapas en la elaboración de un modelo econométrico.

Habitualmente se distinguen cuatro etapas en la elaboración de un modelo econométrico, las cuales son coherentes con los objetivos de la econometría. Como lo señala Alonso et al, (2005), el primer paso en la construcción de un modelo econométrico, corresponde a la selección del problema que se quiere analizar, en función de la identificación de la variable que se quiere explicar y los factores que la explican, la teoría económica existente y la información o datos disponibles.

1.9.1 Especificación.

Corresponde al planteamiento del modelo econométrico en coherencia con la teoría económica existente. La una correcta especificación de un modelo supone que las variables explicativas X son variables relevantes que explican a la endógena Y , es decir, no existe ninguna variable X que no explique nada de la Y .

Procedimiento: 1) Elección de las variables explicativas: X_1, X_2, \dots, X_k ; 2) Elegir la forma funcional: $f(\cdot)$ en $Y = f(X)$. Por lo general, se supondrá la forma funcional lineal: $f(\cdot) \equiv \text{Lineal}$; y 3) elegir el comportamiento probabilístico (distribución) de la perturbación aleatoria μ , para que la se supondrá: $\mu_t \sim \text{iid}(0, \sigma^2)$; y 4) selección de la muestra de datos u observaciones relacionadas con las variables dependiente y explicativas.

Nuestro modelo bivariado de ventas de la editorial quedará especificado de la siguiente manera:

$$V_t = \beta_0 + \beta_1 PE_t + \mu_t \quad t = 1, 2, 3, \dots, T \quad (1.4)$$

1.9.2 Estimación.

En esta etapa se cuantificarán los parámetros desconocidos a partir de la información disponible. Para esto, se elige el método estadístico adecuado y se obtienen los valores numéricos relativos a los parámetros desconocidos del modelo. La estimación se representa sobreponiendo un sombrero, gorro o circunflejo ($\hat{\cdot}$) al parámetro de interés.

$$V_t = \beta_0 + \beta_1 PE_t + \mu_t \quad (1.5)$$

1.9.3 Validación.

En esta etapa se evalúa si el modelo representa correctamente el problema que dio origen a su especificación inicial. Si el modelo supera esta prueba, se constituirá en una herramienta adecuada para el análisis económico; en caso contrario, el modelo deberá someterse a re-especificación, re-estimación y nuevamente a evaluación o contrastación.

1.9.4 Uso del modelo econométrico.

Un modelo que supera la etapa de validación o contrastación es utilizado con fines explicativos y predictivos de la variable dependiente Y ; en nuestro caso, el modelo de ventas de la editorial servirá para explicar los factores que influyen en mayor medida en las ventas. También, servirá para pronosticar las ventas de la editorial trimestre a trimestre a partir de valores específicos que toman las variables explicativas; por ejemplo, ¿cuál será el valor medio de las ventas cuando el precio promedio de los libros disminuya en un 10% o viceversa?.

**ENCARANDO EL MODELO DE
REGRESIÓN LINEAL SIMPLE:
ESPECIFICACIÓN Y ESTIMACIÓN**

Supuesto 1: *El modelo de regresión es lineal en los parámetros, aunque puede o no ser lineal en las variables. Simbólicamente, se tiene que: $Y_i = \beta_0 + \beta_1 X_i + u_i$.*

Emilio para salvar su discusión, debe utilizar el análisis de regresión, como lo afirma Gujarati (2010), “el cual trata del estudio de la dependencia de una variable (*variable dependiente*) respecto de una o más variables (*variables explicativas*) con el objetivo de estimar o predecir la media o valor promedio poblacional de la primera en términos de los valores conocidos o fijos (en muestras repetidas) de las segundas” (p.15).

En el caso particular de la editorial, el análisis de regresión estudiará la dependencia de las ventas de libros (*variable dependiente*) respecto del precio promedio de los ejemplares (*variable explicativa*) con el objetivo de estimar o predecir el valor promedio poblacional de las ventas en términos de un valor conocido o fijo del precio.

2.1 Formalización de la discusión.

Además, tras una discusión, Don Pablo afirma: Estoy dispuesto a formalizar mi teoría a partir del modelo econométrico propuesto por Emilio para que pueda comprobarse si los datos lo avalan de manera que mi hijo no pueda poner en tela de juicio el resultado; de esta manera se saldrá de dudas. En

mi opinión, es el precio de los libros la variable fundamental que influye en las ventas de los mismos, que desincentiva a los compradores cuando este aumenta o estimula su consumo, en caso de que el precio se reduzca.

Para zanjar su discusión, padre e hijo acuerdan desarrollar el siguiente modelo:

$$V_{ti} = V_{ti} = \beta_0 + \beta_{1i} PE_{ti} + u_{ti} \quad i = 1, 2, 3, \dots, T \quad (2.1)$$

Dónde:

V_t = Ventas de la editorial

β_0 = Intercepto o constante

β_1 = Parámetro o coeficiente de la pendiente

μ_t = Término de perturbación estocástica

i = i -ésima observación con T como tamaño de la población

El modelo especificado de ventas de la editorial es lineal en sus parámetros, puesto que el coeficiente β_1 está elevado a la potencia 1.

Tomando las ventas de la editorial (V_t) y el precio promedio de los ejemplares (PE_t), Don Pablo y su hijo Emilio, deciden construir un gráfico de dispersión que permita observar la relación existente entre estas dos variables. La disposición de los datos en la nube de puntos (véase Figura 1.1), confirma efectivamente que existe una relación negativa entre el precio de los libros (X) y las ventas de los mismos (Y). Tal como lo expone la teoría de la demanda, las ventas dependen en cierta medida del comportamiento de los precios de los bienes en cuestión (a mayor precio menor demanda y viceversa). Si bien, esta evidencia ayuda en algo a resolver en alguna medida, la discusión entre Don Pablo y Emilio, no es lo suficientemente contundente para dar por terminada tal discusión. A Don Pablo le interesa conocer la magnitud del efecto del precio sobre las ventas para comprobar su teoría y resolver de una vez por todas, esta azarosa discusión.

La Figura 2.1 muestra una serie de detalles relacionados con la recta de regresión lineal del precio promedio con respecto a las ventas de la editorial.

Los puntos que aparecen por encima y por debajo de la recta corresponden a los residuos que miden la distancia vertical entre la recta y cada uno de los puntos. Entre mayor sea la distancia de cada punto a la recta, mayor es la dispersión de los datos y menor será la capacidad explicativa y predictiva del modelo. Un modelo tendría un gran ajuste si todos los puntos estuviesen localizados sobre la recta de regresión, pero esto en la práctica, es poco posible. La verdadera función del método de Mínimos Cuadrados Ordinarios (en adelante MCO) es minimizar la suma de los residuos al cuadrado $[\sum(Y_i - \hat{Y}_i) = 0]$ para que los estimadores de $\hat{\beta}_0$ y $\hat{\beta}_1$ sean lo más aproximados a los verdaderos valores de β_0 y β_1 y \hat{Y}_i como estimador de $E(Y|X_i)$.

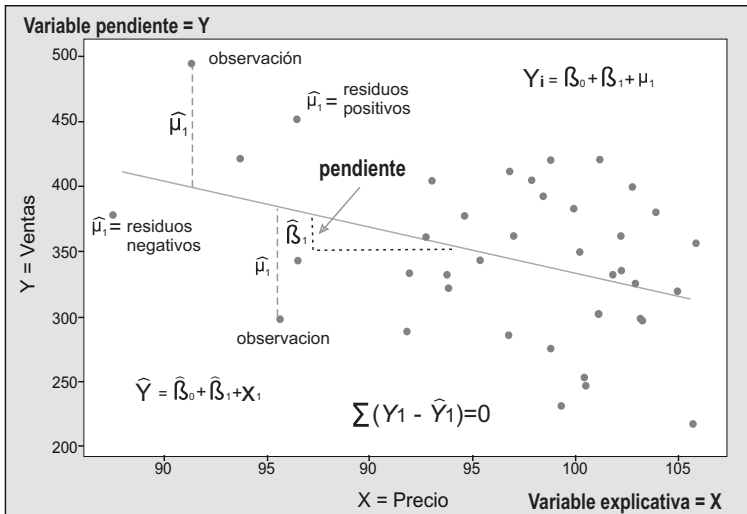


Figura 2.1. Relación entre ventas y precio.

2.1.1. Relación entre dos variables.

Don Pablo y Emilio observan en el gráfico de dispersión (Figura 2.1) la aparente relación existente entre el Precio promedio (P_i) y las Ventas de la editorial (V_i), pero en su discusión, ellos pretenden ir más allá en sus análisis. Para dar respuesta a esta discusión, X representara el precio promedio de los libros vendidos por la editorial y Y las ventas de la editorial en los últimos diez años ($T = 10 \times 4 = 40$ trimestres).

Emilio hace memoria nuevamente, y recuerda que los Profesores *Alonso, Fernández & Gallastegui* (2005:18) hablaron en algún momento en el curso de econometría del método MCO, el cual sirve para cuantificar las relaciones entre variables. Ellos de manera insistente solían decir que el ajuste mínimo cuadrático permite estimar la mejor ecuación de la recta de regresión a través de la minimización de la suma de los residuos al cuadrado. Los residuos mínimo cuadráticos miden para cualquier valor de X , la distancia entre el valor observado de Y y el valor que la recta mínimo cuadrática proporcional. Como este concepto es un tanto confuso para Don Pablo, en particular; Emilio le sugiere a su padre que examine nuevamente el gráfico de dispersión mencionado arriba. En este se observa una línea con pendiente negativa con trazo de izquierda a derecha que viaja de manera aproximada por el centro de la nube de puntos. Esta recta (recta ajustada) es la denominada recta de regresión mínimo cuadrática, la cual posteriormente, será expresada en una ecuación.

Don Pablo siente mucha inquietud por los puntos alrededor de la recta. Con sus gafas puestas, observa que algunos de estos puntos se acercan más a la recta, en cambio, otros se alejan; pero ciertos puntos se localizan precisamente sobre la recta. Ante este asombro, Don Pablo le pide a Emilio una explicación concreta. Emilio de manera rápida le explica a su padre, que a la distancia entre los puntos y la recta se le denomina los residuos, entre más amplia sea esta distancia, los residuos serán mayores y el modelo econométrico menor capacidad explicativa tendrá. En palabras de sus recordados maestros: “Por lo tanto, resulta natural pensar en elegir una recta siguiendo un procedimiento que trate de minimizar las distancias entre la recta y todos los puntos observados” (Alonso et al., 2005, p.16). Ante esto Don Pablo opina: Ahora comprendo el por qué a este método se le conoce como Ajuste Mínimo Cuadrático y por qué la inclusión de la perturbación (parte aleatoria) en el modelo econométrico inicial. *Eureka* dice Don Pablo: Al fin he podido comprender que las relaciones entre las variables económicas no son exactas, de ahí la existencia de los residuos, y por ende, la necesidad de incluir en el modelo una perturbación o término de error.

Don Pablo quiere pasar del discurso a la acción, por lo tanto, le solicita a su hijo Emilio que estime la ecuación de la recta de regresión a través del método MCO, del que tanto ha hablado. Ante esta justa petición, Emilio responde: Padre, ha llegado el momento de encontrar la verdad; para realizar la estimación de la ecuación de la recta de regresión, utilizaré varios enfoques didácticos, para que al final, tú no tengas ninguna duda acerca de los procedimientos estadísticos utilizados. Entonces, ¡manos a la obra!

Procedimiento 1

Emilio inicia la estimación por MCO utilizando notación matricial para que se capture los detalles del procedimiento. Para ello revisa el Apéndice C del libro *Econometría* de profesor Gujarati, D. (2010: 849-876). Como la ecuación que se pretende encontrar tiene dos variables, entonces para obtener la estimación por MCO de β , primero se escribe la regresión muestral de k -variables:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\mu}_i \quad (2.2)$$

La cual se escribe en forma más compacta en notación matricial como:

$$\hat{y} = X\hat{\beta} + \hat{\mu}_i$$

y en forma matricial como:

$$\begin{array}{c} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \\ \begin{matrix} y \\ n \times 1 \end{matrix} \end{array} = \begin{array}{c} \begin{bmatrix} 1 & X_{11} & \dots & X_{k1} \\ 1 & X_{12} & \dots & X_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & X_{1n} & \dots & X_{kn} \end{bmatrix} \\ \begin{matrix} X \\ n \times k \end{matrix} \end{array} \begin{array}{c} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \\ \begin{matrix} \beta \\ k \times 1 \end{matrix} \end{array} + \begin{array}{c} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_n \end{bmatrix} \\ \begin{matrix} \hat{\mu} \\ n \times 1 \end{matrix} \end{array}$$

Donde $\hat{\beta}$ es un vector columna de k elementos compuesto por los estimadores de MCO de los coeficientes de regresión, y donde $\hat{\mu}$ es un vector columna de $n \times 1$ con n residuos.

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_r \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ 1 & X_3 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

$$\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ X_1 & X_2 & X_3 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_3 \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix}$$

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

En los modelos de dos y tres variables, en el caso de k variables los estimadores de MCO se obtienen al reducir:

$$\sum_i^2 = (Y_i - \beta_0 - \beta_1 - \beta_2 X_{1i} - \dots - \beta_k X_{ki})^2$$

Dónde: \sum_i^2 es la suma de cuadrados residual (SCR). En notación matricial, esto equivale a reducir $\hat{\mu}'\hat{\mu}$, así:

$$\hat{\mu}'\hat{\mu} = \begin{bmatrix} \hat{\mu}_1 & \hat{\mu}_2 & \dots & \hat{\mu}_n \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_n \end{bmatrix}$$

$$\hat{\mu}'\hat{\mu} = \hat{\mu}_1^2 + \hat{\mu}_2^2 + \dots + \hat{\mu}_n^2 = \sum \hat{\mu}_i^2$$

Ahora de $y = \mathbf{X}\hat{\beta} + \hat{\mu}_i$, se obtiene $\hat{\mu} = \mathbf{X}\hat{\beta}$, por consiguiente:

$$\hat{\mu}'\hat{\mu} = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$\hat{\mu}'\hat{\mu} = \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta}$$

Don Pablo al mirar este conjunto de fórmulas y símbolos queda perplejo, exclamando: Emilio, ¿qué intentas hacer?, ¿explicar nuestra diferencia de opiniones, o confundirme más? Emilio, ante la confusión natural de su padre, le pide tener un poco de paciencia y concentración.

Emilio, sabe que este tema no es en nada sencillo, por tanto, trata de explicarlo del siguiente modo: ¡Papá!, considera los datos disponibles en la Tabla 1.1, entre los cuales se encuentran las ventas de libros de la editorial 2009-2018 ($V_t = Y$) y el precio promedio de los ejemplares ($PE_t = X$), así como al tiempo o la variable de tendencia. Al incluir esta última en el modelo, se trata de averiguar la relación de las ventas ($V_t = Y$) con el precio promedio ($PE_t = X$) de la variable de tendencia (que puede representar una multitud de otros factores, como la renta, el momento en el ciclo económico, cambio en gustos o preferencias, etcétera). Para fines empíricos, por consiguiente, el modelo de regresión es:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\mu}_i \quad (2.3)$$

Dónde: \hat{Y}_i = Ventas de la editorial; $\hat{\beta}_0$ = efecto marginal en las ventas cuando el precio promedio de los libros varía en una unidad monetaria; X_i = Precio promedio de los libros y $\hat{\mu}_i$ = residuos estimados de la regresión. En notación matricial, el problema se muestra de la siguiente manera:

Don Pablo enfadado por tanto enredo cometido por Emilio, exclama: Mira hijo, ¿puedes ser más concreto? En verdad, no entiendo nada de lo que has hecho hasta ahora. Emilio ante la impaciencia de su padre, responde: Papá, se cómo te sientes, pero a veces es imposible prescindir de la notación matricial para explicar el MCO. Ahora se va a derivar en forma matricial la ecuación de las ventas de la editorial, para que te des cuenta, que este procedimiento no es nada del otro mundo.

Emilio empieza con la ordenación de los datos de ventas y precios en forma de matrices, del siguiente modo:

$$\begin{bmatrix} 1 & 275,5 \\ 1 & 285,6 \\ 1 & 336,8 \\ 1 & 333,4 \\ 1 & 357,0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 404,9 \\ 1 & 421,3 \\ 1 & 333,7 \\ 1 & 379,1 \\ 1 & 407,0 \end{bmatrix} = \begin{bmatrix} 1 & 98,6 \\ 1 & 96,6 \\ 1 & 102,0 \\ 1 & 91,9 \\ 1 & 105,6 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 93,0 \\ 1 & 83,7 \\ 1 & 93,7 \\ 1 & 94,5 \\ 1 & 97,7 \end{bmatrix} \times \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} + \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \\ \hat{\mu}_4 \\ \hat{\mu}_5 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \hat{\mu}_{36} \\ \hat{\mu}_{37} \\ \hat{\mu}_{38} \\ \hat{\mu}_{39} \\ \hat{\mu}_{40} \end{bmatrix}$$

$y =$ \mathbf{X} $\hat{\beta}$ $+$ $\hat{\mu}$
 40×1 40×2 2×1 40×1

Como ya se mencionó, las relaciones entre las variables económicas deterministas (inexactas), la última matriz los residuos resultantes de la diferencia entre el valor de Ventas = Y el valor pronosticado a partir de la ecuación de la recta de regresión (\hat{Y}); la forma como se calcula este valor será explicada más adelante.

Emilio con el apoyo de Excel se alista a realizar operaciones con las matrices construidas previamente. Se dispone a realizar la multiplicación de la matriz X (precio promedio) por la matriz X traspuesta ($X'X$), así:

$$\begin{bmatrix} 1 & 98,6 \\ 1 & 96,6 \\ 1 & 102,0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 93,7 \\ 1 & 94,5 \\ 1 & 97,7 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 & 1 \\ 98,6 & 96,6 & 102,0 & \cdot & \cdot & \cdot & 93,7 & 94,5 & 97,7 \end{bmatrix}$$

2×40
Matriz traspuesta de X = X'

40×2
Matriz X

Como resultado de la multiplicación $X'X$, obtiene:

$$\begin{bmatrix} 40,00 & 3.870,70 \\ 3.870,700 & 376.376,230 \end{bmatrix}$$

2×2

Emilio, ahora multiplica la matriz traspuesta de X (precio promedio = X') por la matriz y = ventas, es decir, $X'y$:

$$\begin{bmatrix} 1 & 98,6 \\ 1 & 96,6 \\ 1 & 102,0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & 93,7 \\ 1 & 94,5 \\ 1 & 97,7 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 & \cdot & \cdot & \cdot & 1 & 1 & 1 \\ 98,6 & 96,6 & 102,0 & \cdot & \cdot & \cdot & 93,7 & 94,5 & 97,7 \end{bmatrix}$$

2×40
Matriz traspuesta de X = X'

40×2
Matriz X

De la multiplicación $X'y$, se obtiene:

$$\begin{bmatrix} 13.789,7 \\ 1.327.971,8600 \end{bmatrix}$$

2×1
 $X'y$

Ahora se calcula con ayuda de Excel, la matriz inversa de X, o sea, X^{-1} . De la multiplicación de $X'X$, se obtiene la matriz $X'X^{-1}$, así:

$$\begin{bmatrix} 40,00 & 3.870,70 \\ 3.870,700 & 376.376,230 \end{bmatrix}$$

2×2
Matriz $X'X$

Cuyo resultado es la matriz $X'X^{-1}$

$$\begin{bmatrix} 5,174928582 & -0,05321961 \\ -0,05321961 & 0,000549974 \end{bmatrix}$$

2×2

Como $\hat{\beta} = (X'X^{-1})(X'y)$, por tanto, para la estimación de los coeficientes⁸ $\hat{\beta}_0 + \hat{\beta}_1$ se procede a multiplicar la matriz $X'X^{-1}$ por la matriz $X'y$, así:

$$\begin{bmatrix} 5,174928582 & -0,05321961 \\ -0,05321961 & 0,000549974 \end{bmatrix} \times \begin{bmatrix} 13.789,7 \\ 1.327.971,8600 \end{bmatrix}$$

2×2 2×1

Por consiguiente, el valor de los parámetros ($\hat{\beta}_0$ y $\hat{\beta}_1$) de la recta de regresión lineal, serán:

$$\begin{aligned} \hat{\beta}_0 &= \begin{bmatrix} 686,569 \end{bmatrix} \\ \hat{\beta}_1 &= \begin{bmatrix} 3,53245 \end{bmatrix} \end{aligned}$$

2×1

$\hat{\beta}_0$ y $\hat{\beta}_1$ son los valores estimados de β_0 y β_1 , los cuales también podrían representarse simbólicamente como b_0 y b_1 , para evitar el uso del circunflejo, por lo tanto, $\hat{y} = b_0 + b_1X_i$.

Uniendo los coeficientes obtenidos, se logra estimar la recta de regresión lineal: $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1X$, por ende, $\hat{Y} = 686,57 - 3,53X$, donde: \hat{Y} = valor medio de las ventas y X = Precio promedio de los libros. En conclusión, la ecuación de la recta de regresión lineal de la editorial es:

$$\widehat{\text{Ventas}} = 686,57 - 3,53\text{Precio} \quad (2.4)$$

A partir de esta ecuación se podrá explicar la influencia del precio sobre las ventas ($X \rightarrow Y$) y predecir el valor medio de las ventas de la editorial a partir de valores dados del precio promedio de los libros

Como lo sospechaban los dos, existe una relación negativa entre las ventas de los libros y el precio promedio de los mismos. La ecuación arriba descrita indica que si el precio promedio de un libro varía (aumenta/disminuye) en una unidad monetaria (\$1), la ventas variarán (disminuirán/aumentarán) en 3,53 unidades monetarias⁹ (\$3,53). Como parece que esta interpretación no

⁸ El gorro o circunflejo colocado encima de cada parámetro significa un valor estimado del parámetro. Un estimador es un valor que puede calcularse a partir de los datos muestrales y que proporciona información sobre el valor del parámetro (β).

⁹ Se debe recordar que las ventas están representadas en millones de pesos colombianos y el precio promedio en miles de pesos.

es suficientemente clara para Don Pablo, Emilio trata de exponerlo de la siguiente manera: ¡Mira papá!, si la editorial incrementa el precio promedio de un libro en \$1, las ventas se reducirían en \$3,53 y lo contrario. Este análisis permite concluir que las ventas de libros son muy o elásticas al precio, lo cual indica que se debe ser muy prudentes con la fijación de la política de precios, de lo contrario, nuestros ingresos se verán fuertemente menguados.

Procedimiento 2

Para evitar que Don Pablo quede con dudas a estas alturas del ejercicio, Emilio se propone desarrollar la estimación de la ecuación de regresión lineal de un modo alternativo, utilizando la hoja de cálculo Excel organiza la información de ventas y precio promedio en la Tabla 1.4. Para la estimación de los valores de los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$, se procede a calcular la media aritmética o promedio del precio = X y de las ventas = Y, así:

$$\bar{X} = \sum_n \frac{X}{n}$$

Dónde: \bar{X} = promedio (media) de los valores X (precio); $\sum X$ = suma total de la columna precios y n = número de trimestres.

$$\bar{X} = \frac{3.870,7}{40} = \$ 96,8 \text{ miles}$$

$$\bar{Y} = \sum_n \frac{Y}{n}$$

Dónde: \bar{Y} = promedio (media) de los valores Y (ventas); $\sum Y$ = suma total de la columna Ventas y n = número de trimestres.

$$\bar{Y} = \frac{13.789,7}{40} = \$ 344,7 \text{ millones}$$

Primero, se procede a estimar el valor del parámetro de β_1 llamado también, pendiente de la recta de regresión, y posteriormente, el valor de β_0 , conocido como intercepto o constante, partir de las siguientes fórmulas:

$$\beta_1 = \Sigma(X - \bar{X})(\hat{Y} - \bar{Y}) / \Sigma(X - \bar{X})^2 \quad (2.5)$$

Se deduce que la recta de regresión debe pasar por el punto correspondiente a las medias de ambas variables (X y Y) y que debe tener por pendiente ($\hat{\beta}_1$) la covarianza dividida por la varianza de la variable X , lo cual se expresa como $\hat{\beta}_1 = S_{x,y} / S_x^2 = r_{x,y} (S_y / S_x)$.

Reemplazando valores en (2.5) se obtiene: $\hat{\beta}_1 = \frac{6.422,9}{1.818,3} = 3,53$

Tabla 2.1

Relación de ventas trimestrales de la editorial 2009-2018

Trimestre	Y= Ventas	Y= Precio	(X - \bar{X})	(Y - \bar{Y})	(X - \bar{X}) ²	(X - \bar{X}) (Y - \bar{Y})
2009:1	275,5	98,6	1,8	-69,2	3,4	-126,9
2009:2	285,6	96,6	-0,2	-59,1	0,0	9,9
2009:3	336,8	102,0	5,2	-7,9	27,4	-41,6
2009:4	333,4	91,9	-4,9	-11,3	23,7	55,2
2010:1	357,0	105,6	8,8	12,3	78,0	108,3
2010:2	325,2	102,8	6,0	-19,5	36,4	-117,9
2010:3	362,2	102,1	5,3	17,5	28,4	93,1
2010:4	232,0	99,1	2,3	-112,7	5,4	-263,0
2011:1	252,2	100,3	3,5	-92,5	12,5	-326,9
2011:2	322,1	104,8	8,0	-22,6	64,5	-181,9
2011:3	297,3	85,6	-11,2	-47,4	124,7	529,8
2011:4	298,9	103,0	6,2	-45,8	38,8	-285,7
2012:1	246,8	100,4	3,6	-97,9	13,2	-355,8
2012:2	322,4	93,8	-3,0	-22,3	8,8	66,3
2012:3	383,5	99,7	2,9	38,8	8,6	113,7
2012:4	321,8	104,6	7,8	-22,9	61,3	-179,7
2013:1	351,5	100,0	3,2	6,8	10,4	21,8
2013:2	381,1	103,7	6,9	36,4	48,1	252,0
2013:3	412,5	96,7	-0,1	67,8	0,0	-4,6
2013:4	217,5	105,5	8,7	-127,2	76,3	-1111,1
2014:1	229,9	99,2	2,4	-114,8	5,9	-279,4
2014:2	342,9	86,4	-10,4	-1,8	107,5	19,1
2014:3	361,9	92,7	-4,1	17,2	16,5	-69,8
2014:4	301,3	100,9	4,1	-43,4	17,1	-179,5

Continuación Tabla 2.1

Trimestre	Y= Ventas	Y= Precio	(X - \bar{X})	(Y - \bar{Y})	(X - \bar{X}) ²	(X - \bar{X})	(Y - \bar{Y})
2015:1	332,5	101,7	4,9	-12,2	24,3		-60,4
2015:2	343,2	95,3	-1,5	-1,5	2,2		2,3
2015:3	421,9	98,6	1,8	77,2	3,4		141,4
2015:4	401,9	102,5	5,7	57,2	32,9		327,7
2016:1	421,8	101,0	4,2	77,1	17,9		326,1
2016:2	361,7	96,9	0,1	17,0	0,0		2,2
2016:3	393,6	98,3	1,5	48,9	2,3		74,9
2016:4	287,6	91,8	-5,0	-57,1	24,7		283,9
2017:1	380,0	77,8	-19,0	35,3	359,8		-668,7
2017:2	495,6	81,6	-15,2	150,9	230,1		-2288,1
2017:3	452,6	86,6	-10,2	107,9	103,4		-1096,6
2017:4	404,9	93,0	-3,8	60,2	14,2		-226,6
2018:1	421,3	83,7	-13,1	76,6	170,8		-1000,4
2018:2	333,7	93,7	-3,1	-11,0	9,4		33,9
2018:3	379,1	94,5	-2,3	34,4	5,1		-77,9
2018:4	407,0	97,7	0,9	62,3	0,9		58,1
Sumatoria	13.789,7	3.870,7	-	-	1.818,3	-	6.422,9
Promedio	344,7	96,8					

Fuente: Elaboración propia con base en la Tabla 1.2

En $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, se obtiene el coeficiente de $\beta_0 = 344,7 + (-3,53)(96,8) = 686,57$

Por consiguiente: $\widehat{\text{Ventas}} = 686,57 - 3,53\text{Precio}$

Procedimiento 3

Por último, Emilio se dispone a estimar paso a paso la recta de regresión lineal de manera automatizada con la ayuda del software econométrico EViews 10 de amplio uso académico y profesional. El cargue en EViews del archivo de trabajo en excel *Tab. 1* se realiza mediante *File* → *Open* → *EViews Workfile* como se muestra en las Figuras 2.2 y 2.3.

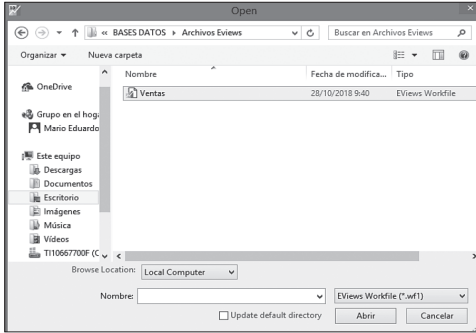


Figura 2.2

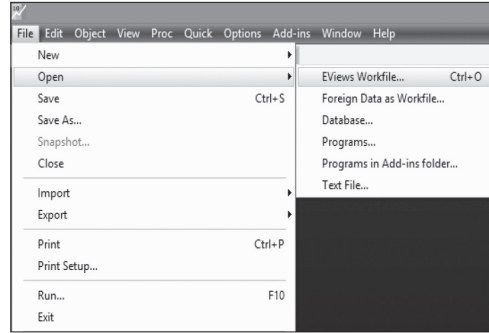


Figura 2.3

Al pulsar Abrir se carga en la memoria la tabla en mención. Si se quiere ver el contenido del conjunto de datos bastará con seleccionar con el puntero del mouse y oprimiendo la tecla CTRL, las variables ventas y precio (en este orden). Posteriormente, se hace clic con el botón derecho del ratón, enseguida se elige las variables en mención y se selecciona Open → As Group (véase Figura 2.4). De este modo, se puede ver la información de las variables del modelo como grupo, como se indica en la Figura 2.5.

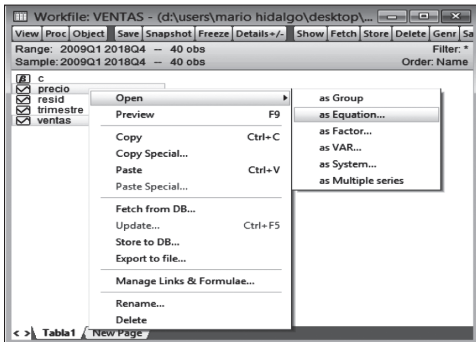


Figura 2.4

The screenshot shows the EViews software data table for 'Group: UNTITLED Workfile: VENTAS:Tabla1'. The table has columns for 'View', 'Proc', 'Object', 'Print', 'Name', 'Freeze', 'Default', 'Sort', 'Edit +/-', 'Smpl +/-', and 'Compare +/-'. The data rows are as follows:

	VENTAS	PRECIO
2009Q1	275.5	98.6
2009Q2	285.6	96.6
2009Q3	336.8	102.0
2009Q4	333.4	91.9
2010Q1	357.0	105.6
2010Q2	325.2	102.8
2010Q3	362.2	102.1
2010Q4	232.0	99.1
2011Q1	252.2	100.3
2011Q2	322.1	104.8
2011Q3	297.3	85.6
2011Q4	298.9	103.0
2012Q1	246.8	100.4
2012Q2	322.4	93.8
2012Q3	383.5	99.7
2012Q4	321.8	104.6
2013Q1	351.5	100.0
2013Q2	381.1	103.7
2013Q3	412.5	96.7
2013Q4	217.5	105.5
2014Q1	229.9	99.2
2014Q2	342.9	86.4
2014Q3

Figura 2.5

Para realizar la regresión se cierra la ventana anterior, pulsando la tecla CTRL más el botón izquierdo del ratón, seleccionamos las variables *ventas* y *precio* en su orden, damos clic derecho y elegimos *Open* → *As Equation*, como se muestra en la Figura 2.6. En el campo *Equation Specification* aparece las variables “ventas precio c”. La variable “c” corresponde al intercepto, término constante u ordenada en el origen. Se puede modificar el orden

en el cual aparecerán las variables en la salida de EViews conservando el orden lógico de la ecuación ($Y = a + bX$; dónde $Y = \text{Ventas}$, $a = \text{intercepto}$, $b = \text{coeficiente}$ y $X = \text{Precio promedio}$) con el propósito de facilitar su lectura y comprensión.

En la ventana de diálogo *Equation specification* se ordena las variables del siguiente modo: *ventas c precio*; enseguida, en el campo *Estimation settings* seleccionamos el método de estimación, en nuestro caso, *LS – Least Squares (NLS and ARMA)*, en español: mínimos cuadrados ordinarios (MCO) y se da Aceptar, como se indica en la Figura 2.7.

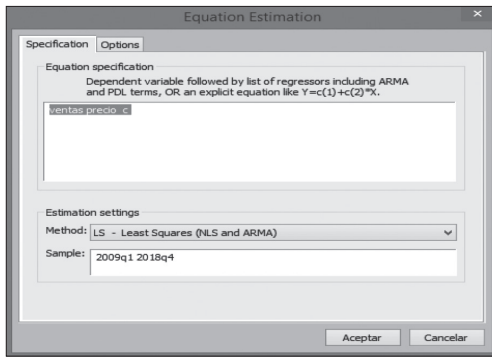


Figura 2.6

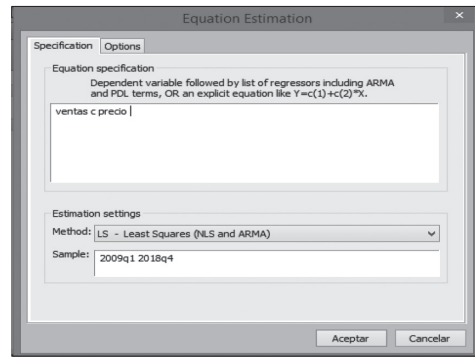


Figura 2.7

Se puede observar la estructura de la ecuación de las ventas de la editorial seleccionando *View* → *Representations* y se da clic (Figura 2.8). EViews en el apartado *Substituted Coefficients* ordena la estimación de la ecuación de la recta de regresión empezando por el término constante y terminando con el coeficiente de la variable explicativa (precio), como se muestra en la Figura 2.9.

	Std. Error	t-Statistic	Prob.
C	135.0660	5.083209	0.0000
PRECIO	1.392403	-2.536944	0.0154

Mean dependent var	344.7425
S.D. dependent var	63.37666
Akaike info criterion	11.05428
Schwarz criterion	11.13873
Hannan-Quinn criter.	11.08482
Durbin-Watson stat	1.241471

F-statistic	6.436083
Prob(F-statistic)	0.015410

Figura 2.8

Estimation Command:
 =====
 LS VENTAS C PRECIO

Estimation Equation:
 =====
 VENTAS = C(1) + C(2)*PRECIO

Substituted Coefficients:
 =====
 VENTAS = 686.568580576 - 3.53244716022*PRECIO

Figura 2.9

Los resultados de la estimación se muestran en la Tabla 2.2.

Tabla 2.2**Resultados MCO de las ventas contra el precio.**

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	686,5686	135,0660	5,083209	0,0000
PE	-3,532447	1,392403	-2,536944	0,0154
R-squared	0,144839	Mean dependent var		344,7425
Adjusted R-squared	0,122335	S.D. dependent var		63,37666
S.E. of regression	59,37385	Akaike info criterion		11,05428
Sum squared resid	133.958,8	Schwarz criterion		11,13873
Log likelihood	-219,0857	Hannan-Quinn criterion		11,08482
F-statistic	6,436083	Durbin-Watson stat		1,241471
Prob(F-statistic)	0,015410			

Aunque la información estadística que proporciona EViews es muy completa, por el momento, solo se tomará de la Tabla 2.2 lo correspondiente a la columna de los coeficientes (*Coefficient*) de las variables de la ecuación (intercepto y coeficiente del precio).

Por consiguiente, la ecuación de la recta de regresión resultante estimada con el paquete EViews genera los mismos resultados obtenidos con el uso de los métodos manuales vistos anteriormente, con la diferencia que la automatización ahorra mucho tiempo y esfuerzo humano, además de ofrecer mayor precisión en los resultados.

Sustituyendo los valores respectivos en la ecuación: $\widehat{Ventas} = \hat{\beta}_0 - \hat{\beta}_1 \text{Precio}$, se obtiene:

$$\widehat{Ventas} = 686,57 - 3,53\text{Precio}$$

La conjetura de Don Pablo respecto del efecto negativo del precio sobre las ventas parece confirmarse. De otro lado, la capacidad explicativa de este modelo a juzgar por el valor del coeficiente de determinación¹⁰ (R^2), parece bastante reducido. Por ello, para poder convencer a Emilio será preciso decidir si el coeficiente estimado proporciona o no un valor fiable (en el sentido de próximo al valor teórico que se quería conocer), lo que implica hacer referencia a la distribución de dicho coeficiente y, en particular, a la dispersión de dicha distribución. Como no es posible cuantificar ésta mientras no se disponga de un estimador para la varianza de las perturbaciones ($\hat{\sigma}^2$), la respuesta, si Don Pablo tiene o no razón en su planteamiento se propone resolverse en los capítulos siguientes. Hasta ahora, Don Pablo siente cierta satisfacción al ver que el precio medio de los libros tiene algún poder explicativo sobre las ventas, como inicialmente intuyó.

2.1.2 Interpretación de los coeficientes.

Como el modelo de regresión simple únicamente tiene dos variables (una dependiente y otra explicativa), el coeficiente de la pendiente β_1 que acompaña la variable explicativa mide la fuerza del efecto marginal de la intensidad de la variable explicativa sobre la variable dependiente, cuando la primera varía en una unidad. Como ya se ha visto, este coeficiente se estima a partir de datos. El coeficiente de regresión es un número que contiene dos tipos de información: además de indicar si la relación propuesta es positiva o negativa, indica la variación de la variable dependiente cuando la variable explicativa varía en una unidad.

¹⁰ El coeficiente de determinación es la proporción de la variación muestral en y_i que es explicada por la línea de regresión de MCO. Por definición R^2 es un número entre 0 y 1. En general, que el R^2 sea pequeño indica que es difícil predecir con mucha exactitud resultados individuales para y . Esto ratifica que en ciencias sociales es muy difícil predecir el comportamiento de los individuos.

En el caso de la editorial, el coeficiente β_1 mide la fuerza del efecto marginal de la intensidad del precio promedio de los libros sobre las ventas de la editorial, en otras palabras, cuantifica el impacto de un cambio unitario en la intensidad del precio promedio sobre las ventas de la editorial.

En la Tabla 2.3 se muestra la interpretación de los coeficientes para los distintos modelos.

Tabla 2.3

Interpretación de los coeficientes.

Modelo	Variable dependiente	Variable explicativa	Interpretación
Nivel-nivel	Y	X	$\Delta y = \beta_1 \Delta x$
Nivel-log	Y	$\ln(X)$	$\Delta y = (\beta_1/100)\% \Delta x$
Log-nivel	$\ln(Y)$	X	$\% \Delta y = (100\beta_1) \Delta x$
Log-log	$\ln(Y)$	$\ln(X)$	$\% \Delta y = \beta_1 \% \Delta x$

Fuente: Pérez (2006: 7).

Modelo nivel - nivel

Se puede interpretar el coeficiente β_1 como el número de unidades que varía en su media Y cuando X_1 varía en una unidad (permaneciendo el resto de las variables constantes o *ceteris paribus*¹¹). El término constante β_0 se interpreta como el pronóstico de Y cuando X_1 es igual a 0.

11 (...) La mayor parte de las cuestiones económicas tienen un carácter *ceteris paribus*. Por ejemplo, cuando se analiza la demanda del consumidor interesa saber el efecto que tiene una modificación en el precio de un determinado bien sobre la cantidad demandada, mientras todos los demás factores —tales como ingreso, precios de los demás bienes y preferencias individuales— se mantienen constantes. Si no permanecen constantes los demás factores, entonces no se puede saber cuál es el efecto de una modificación en el precio sobre la cantidad demandada (Wooldridge, 2009:12). Como se explicará en el capítulo 3, la condición *ceteris paribus* intenta aislar el efecto que provoca cada variable explicativa sobre la variable dependiente para medir su contribución; en nuestro caso, el efecto o razón de cambio del precio sobre las ventas. Si prescindimos de esta condición, el análisis económico se torna prácticamente imposible, dada la interacción de variables múltiples y su simultaneidad.

En nuestro caso, $\widehat{Ventas} = 686,57 - 3,53\text{Precio}$, cuando el coeficiente del precio $\beta_1=3,53$ y el valor de X es igual a cero ($X = 0$), entonces, el valor de Y es igual al valor del intercepto es 686,57, por lo tanto, el pronóstico de las ventas será: $\widehat{Ventas} = \$ 686,57$.

$$\widehat{Ventas} = 686,57 - (3,53)(0)$$

$$\widehat{Ventas} = 686,57 - 0$$

$$\widehat{Ventas} = 686,57$$

Interpretación del término independiente: El término constante representa el valor de la ordenada donde la línea de regresión se intercepta al eje Y. El intercepto o constante paramétrica es el coeficiente que no acompaña ninguna variable; en algunos casos, su magnitud no tiene interpretación económica. En nuestro caso, si el precio de los libros es cero, es decir, en el evento que la editorial los regale a sus clientes (supuesto considerado como absurdo), las ventas de la editorial serían de \$686,57. El valor del intercepto provoca la formulación de varias preguntas: ¿Cómo es posible que las ventas de la editorial sean \$686,57 cuando no se cobra ningún precio por la venta de los libros?; ¿de dónde provienen estos recursos?; si las ventas de la editorial no dependen únicamente del precio, ¿de qué otros factores dependen, considerando que este modelo solo hay una única variable explicativa?

En economía, en muchas ocasiones la interpretación económica del valor de la constante tiene poca importancia por su naturaleza autónoma, y otras veces, por lo contradictorio de su signo. De todos modos, la interpretación del intercepto es algo complejo. ¿Cómo se interpretarían las ventas de la editorial en el modelo de ventas se hubiese obtenido un intercepto con signo negativo, es decir, - \$686,57? En el mundo real no existen ventas negativas, sino variaciones negativas en los montos de ventas.

Por consiguiente, el término independiente, constante o intercepto es el valor que toma la variable dependiente estimada \hat{Y}_i cuando la variable explicativa X_i es igual a cero.

Si se escribe el modelo como nivel-nivel (lineal-lineal):

$$Y = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + \mu_i \quad i = 1, 2, 3, \dots, n$$

$$E(Y_i | X_{1i}, X_{2i}, \dots, X_{ki}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}$$

$$Y = E(Y_i | X_{1i}, X_{2i}, \dots, X_{ki}) = \sigma^2$$

En el caso de que todas las variables, excepto X_i permanezcan constantes, se tiene:

$$\Delta E(Y | X_1, X_2, \dots, X_k) = \beta_i \Delta X_i$$

$$\beta_i = \frac{\Delta E(Y | X_1, X_2, \dots, X_k)}{\Delta X_i}$$

Entonces se puede interpretar el coeficiente β_i como el número de unidades en varía en su media Y cuando X_i varía en una unidad, permaneciendo el resto de las variables constantes.

En nuestro modelo: $\widehat{Ventas} = 686,57 - 3,53 \text{Precio}$, la interpretación es la siguiente: cuando el precio promedio varía en una unidad monetaria, la media de las ventas de la editorial varía en 3,53 unidades.

Modelo nivel – log (lineal logarítmico)

Se propone ahora que en el modelo se escribe la variable explicativa (precio promedio) en logaritmo natural, el cual se denomina modelo nivel-log (lineal-logarítmico), para lo cual se tiene que transformar la variable $X =$ precio promedio en logaritmo natural y luego hacer la respectiva estimación.

$$Y = \beta_0 + \beta_1 \ln X + \mu$$

$$E(\mu | X) = 0 \rightarrow E(Y | X) = \beta_0 + \beta_1 \ln X$$

$$\beta_1 = \frac{\Delta E(Y | X)}{\Delta \ln X} = \frac{\Delta E(Y | X)}{\Delta X / X}$$

Multiplicando y dividiendo por 100 para expresar la variación de X en términos de %, se tiene que

$$\frac{\beta_1}{100} \approx \frac{\Delta E(Y \setminus X)}{100} * \frac{\Delta X}{X}$$

Se puede interpretar entonces que cuando X (precio promedio) varía en uno por ciento (1%), la media de Y (ventas de la editorial) varía en $\beta_1/100$ unidades de Y .

El modelo: $\bar{V} = 1.851,62 - 329,75IP$ puede interpretarse del siguiente modo: cuando el precio promedio de los libros expresado en logaritmo natural varía en 1%, la media de las ventas de la editorial varía en 3,30 ($329,75/100 = 3,30$) unidades monetarias.

Modelo log – nivel (logarítmico lineal)

Se supone ahora que en el modelo se escribe la variable dependiente ventas en logaritmo natural (modelo log-nivel), el cual se denomina modelo log-nivel (logarítmico-lineal), para lo cual se tiene que transformar la variable Y (ventas de la editorial) en logaritmo natural y luego hacer la respectiva estimación.

$$\ln Y = \beta_0 + \beta_1 X + \mu$$

$$E(\mu \setminus X) = 0 \rightarrow E(\ln Y \setminus Y) = \beta_0 + \beta_1 X$$

$$\beta_1 = \frac{\Delta E(\ln Y \setminus X)}{\Delta X} \approx \frac{E\left(\frac{\Delta Y}{Y \setminus X}\right)}{\Delta X}$$

Si se multiplica por 100 para expresar la variación de Y en términos de % se obtiene:

$$\frac{\beta_1}{100} \approx \frac{E\left(100 x \frac{\Delta Y}{Y}\right)}{\Delta X}$$

Podemos interpretar entonces que cuando X varía en una unidad, Y varía en su media en $(\beta_1 \times 100)\%$. El modelo $\widehat{I\hat{V}} = 6,80 - 0,010P$, se interpreta, así: cuando el precio promedio de los libros varía en una unidad monetaria, la media de las ventas expresado en logaritmo natural varía en $0,0001(0,010/100 = 0,0001)$ unidades monetarias.

Modelo log-log (logarítmico-logarítmico)

Se supone ahora que en el modelo se escribe en logaritmos naturales tanto la variable dependiente como la variable explicativa, se tiene entonces el modelo denominado modelo log-log o doble logarítmico.

$$\ln Y = \beta_0 + \beta_1 \ln X + \mu$$

$$E(\mu \setminus X) = 0 \rightarrow \frac{E(\ln Y \setminus X)}{\Delta \ln X} \approx \frac{E\left(\frac{\Delta Y}{Y \setminus X}\right)}{\Delta X/X}$$

Se puede interpretar entonces que cuando X varía en uno por ciento (1%), la media de Y varía en un $\beta_1\%$, es decir, se interpretan los coeficientes en términos de elasticidades. En el modelo $\widehat{I\hat{V}} = 10,13 - 0,94 IP$, cuando el precio promedio de los libros varía en uno por ciento (1%), la media de las ventas de la editorial varía en 0,94%.

Como $\hat{v} = 686,57 - 3,53p$ corresponde a un modelo nivel-nivel (lineal-lineal) no es necesario realizar ninguna transformación de las variables, en cambio, para estimar los modelos nivel-log, log-nivel y log-log es necesario transformar la variable dependiente ventas y la variable explicativa precio en logaritmos naturales y luego realizar la estimación mediante MCO.

Ahora, se automatizará este proceso con el software econométrico EViews¹² versión 10 del siguiente modo: Primero, se procede a transformar

¹² EViews es un paquete estadístico para Microsoft Windows, usado principalmente para análisis econométrico, pronóstico y simulación. Fue desarrollado por Quantitative Micro Software (QMS). Este programa de computador combina la tecnología de hoja de cálculo con tareas tradicionales encontradas en software estadístico tradicional, empleando una interfaz de usuario gráfica. Es el paquete ideal para trabajar con datos de series de tiempo, secciones.

las variables *ventas* y *precio promedio* en logaritmos naturales. Para ello, se abre nuevamente el archivo de trabajo Tabla 1 y en la ventana del Workfile con ayuda del ratón se selecciona *Quick* → *Generate Series* y damos clic (Figura 2.10). En el cuadro de diálogo que se abre, específicamente, en el campo *Generate Series by Equation* se escribe: $lv = \log(v)$ y se da OK (Figura 2.11).

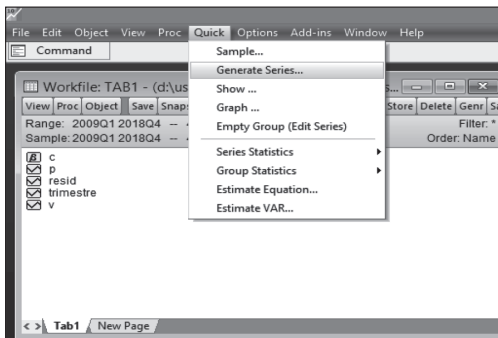


Figura 2.10

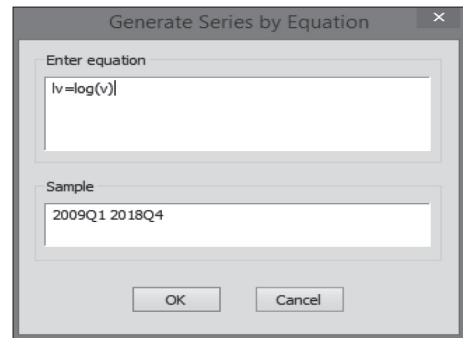


Figura 2.11

Mediante esta orden, la variable “v” se ha transformado en logaritmo natural (lv). Para comprobarlo se abre la variable lv, tal como se muestra en la Figura 2.12. Luego se repite este mismo procedimiento para transformar la variable precio promedio también en logaritmo natural (lp).

transversales o datos longitudinales. EViews permite administrar datos de manera rápida y eficiente, realizar análisis econométricos y estadísticos, generar pronósticos o simulaciones de modelos, y producir gráficos y tablas de alta calidad para publicación o inclusión en otras aplicaciones. Para mayor información se puede consultar: <https://www.software-shop.com/producto/eviews>

Year	Value
2009Q1	5.618598
2009Q2	5.654592
2009Q3	5.819489
2009Q4	5.809343
2010Q1	5.877736
2010Q2	5.784440
2010Q3	5.892197
2010Q4	5.446737
2011Q1	5.530222
2011Q2	5.774862
2011Q3	5.694742
2011Q4	5.700109
2012Q1	5.508578
2012Q2	5.775793
2012Q3	5.949340
2012Q4	5.773930
2013Q1	5.862210
2013Q2	5.943062
2013Q3	6.022236
2013Q4	6.022236

Figura 2.12

Una vez transformadas las variables ventas y precio promedio, con la ayuda de EViews se puede estimar el modelo nivel-log a través de MCO, tomando la variable ventas en su serie original y la variable precio expresada en logaritmo natural (véase Figura 2.12). Para ello hay que situarse en la ventana del *Workfile*, se selecciona las dos variables en su orden respectivo (*v* y *lp*) y abrimos como ecuación mediante *Open* → *as Equation* y se da clic como se muestra en la Figura 2.13. En el cuadro de diálogo que se abre a continuación se verifica que las variables se encuentren en el siguiente orden: *v c lp* y se da Aceptar (véase Figura 2.14).

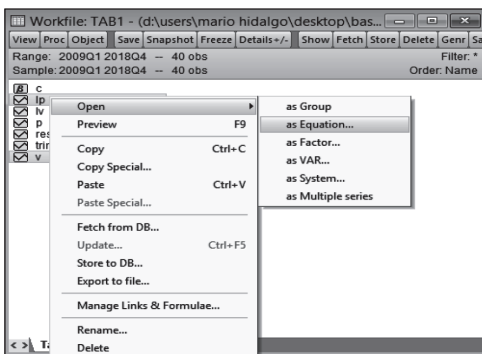


Figura 2.13

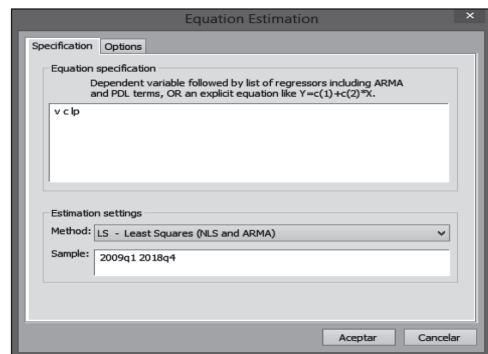


Figura 2.14

Tabla 2.4**Resultados MCO de ventas contra el logaritmo natural del precio.**

Variable Dependiente: V

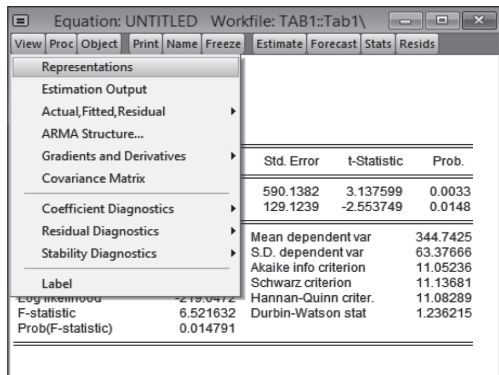
Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

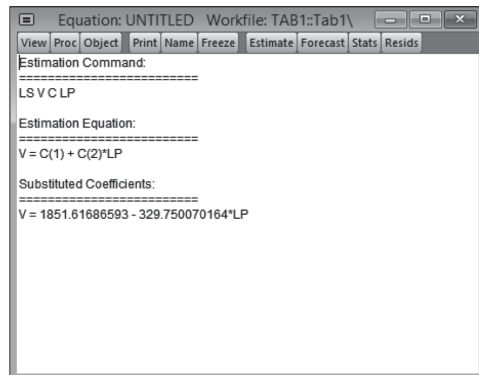
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.851,617	590,1382	3,137599	0,0033
LP	-329,7501	129,1239	-2,553749	0,0148
R-squared	0,146482	Mean dependent var		344,7425
Adjusted R-squared	0,124021	S.D. dependent var		63,37666
S.E. of regression	59,31658	Akaike info criterion		11,05236
Sum squared resid	133.701,4	Schwarz criterion		11,13681
Log likelihood	-219,0472	Hannan-Quinn criterion		11,08289
F-statistic	6,521632	Durbin-Watson stat		1,236215
Prob(F-statistic)	0,014791			

En la ventana del objeto ecuación se selecciona *View* → *Representations* y damos clic como se muestra en la Figura 2.15. De este modo, se obtiene la ecuación de la recta de regresión de las ventas contra el logaritmo natural del precio (véase Figura 2.16).



	Std. Error	t-Statistic	Prob.
590.1382	3.137599	0.0033	
129.1239	-2.553749	0.0148	
Mean dependent var			344.7425
S.D. dependent var			63.37666
Akaike info criterion			11.05236
Schwarz criterion			11.13681
Hannan-Quinn criter.			11.08289
Durbin-Watson stat			1.236215

Figura 2.15



```

Estimation Command:
=====
LS V C LP

Estimation Equation:
=====
V = C(1) + C(2)*LP

Substituted Coefficients:
=====
V = 1851.61688593 - 329.750070164*LP

```

Figura 2.16

El modelo nivel-log estimado $\hat{v} = 1.851,62 - 329,75lp$ se interpreta, así: si el precio promedio de los libros se incrementa en uno por ciento (1%), las ventas de la editorial se reducirán en \$ 329,75 y viceversa, manteniéndose fijas las demás variables que también afectan las ventas.

Ahora se va a estimar el modelo logarítmico lineal (log-nivel) a través de MCO, tomando la variable ventas expresada en logaritmo natural (lv) y la variable precio promedio en su serie original. Para ello se sitúa en la ventana del Workfile, con la tecla CTRL y el puntero del ratón se selecciona las variables lv y p y se abre como ecuación mediante *Open* → *as Equation* y se da clic como se muestra en la Figura 2.17. En el cuadro de diálogo que se abre a continuación se verifica que las variables se encuentren en el siguiente orden: lv c p y se da Aceptar (véase Figura 2.18).

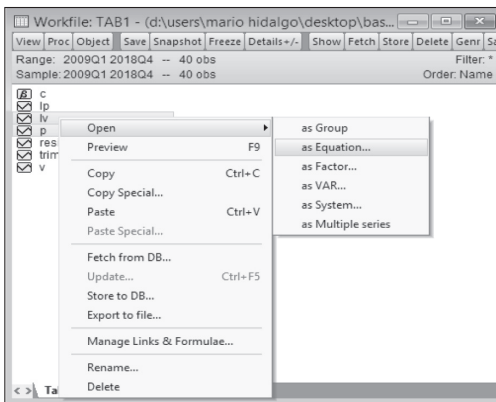


Figura 2.17

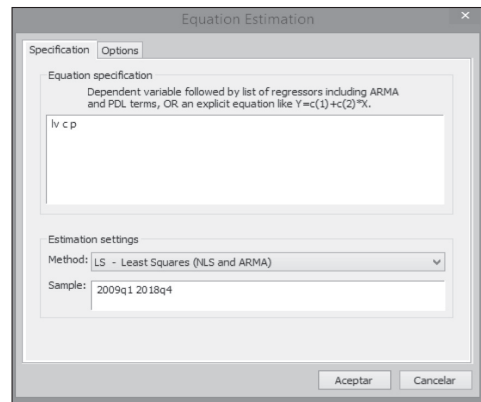


Figura 2.18

El modelo nivel-log estimado $\hat{v} = 1.851,62 - 329,75lp$ se interpreta, así: si el precio promedio de los libros se incrementa en uno por ciento (1%), las ventas de la editorial se reducirán en \$ 329,75 y viceversa, manteniéndose fijas las demás variables que también afectan las ventas.

Ahora se va a estimar el modelo logarítmico lineal (log-nivel) a través de MCO, tomando la variable ventas expresada en logaritmo natural (lv) y la variable precio promedio en su serie original. Para ello se sitúa en la ventana del Workfile, con la tecla CTRL y el puntero del ratón se selecciona las variables lv y p y se abre como ecuación mediante *Open* → *as Equation* y se da clic como se muestra en la Figura 2.17. En el cuadro de diálogo que se abre a continuación se verifica que las variables se encuentren en el siguiente orden: lv c p y se da Aceptar (véase Figura 2.18).

En la Tabla 2.5 se muestran los resultados de la regresión que arroja la salida de EViews.

Tabla 2.5

Resultados MCO del logaritmo natural de ventas contra el precio.

Variable Dependiente: LV
 Método: MCO
 Muestra: 20009Q 2018Q4
 Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6,804526	0,412187	16,50834	0,0000
P	-0,010118	0,004249	-2,381082	0,0224
R-squared	0,129828	Mean dependent var		5,855448
Adjusted R-squared	0,106929	S.D. dependent var		0,191734
S.E. of regression	0,181193	Akaike info criterion		-0,529797
Sum squared resid	1,247579	Schwarz criterion		-0,445353
Log likelihood	12,59595	Hannan-Quinn criterion		-0,499265
F-statistic	5,669550	Durbin-Watson stat		1,249279
Prob(F-statistic)	0,022377			

Enseguida, en la ventana del objeto ecuación se selecciona *View* → *Representations* y se da clic como se muestra en la Figura 2.19. De este modo, se obtiene la ecuación de la recta de regresión de las ventas contra el logaritmo natural del precio (véase Figura 2.20).

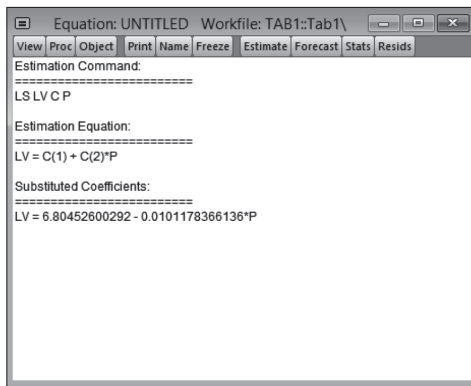


Figura 2.19

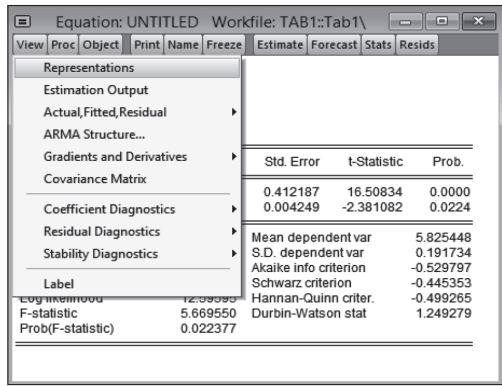


Figura 2.20

El modelo log-nivel estimado $\hat{lv} = 6,805 - 0,010p$ se interpreta, así: si el precio promedio de los libros se incrementa en una unidad monetaria (\$1), la media de las ventas de la editorial expresadas en logaritmo natural se reducen en \$ 0,0001 ($0,010/100 = 0,0001$) y viceversa, manteniéndose fijas las demás variables que también afectan las ventas.

Muchas veces resulta más fácil interpretar el efecto de la variable X sobre la variable Y en términos de porcentajes (elasticidades) que en unidades monetarias, para ello se utiliza el modelo log-log o doble logarítmico. En este caso se estima la ecuación de regresión tomando tanto la variable ventas como la variable precio promedio transformadas en logaritmos naturales. En la ventana del objeto ecuación, en la barra de herramientas se da clic en *Estimate* y en el cuadro de diálogo que se abre a continuación, se edita las variables **lv c lp** y se da Aceptar como se muestra en la Figura 2.21.

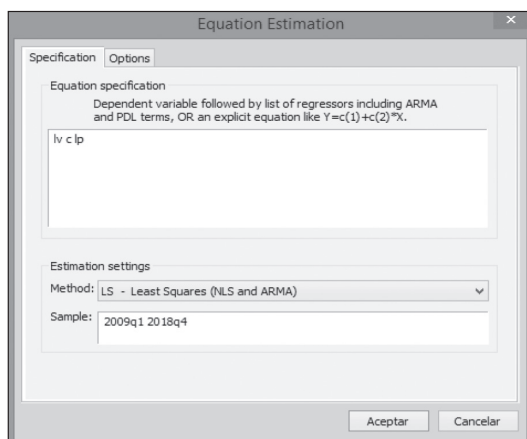


Figura 2.21

De este modo, se estima la ecuación de la recta de regresión a partir de la información proporcionada por EViews disponible en la Tabla 2.6.

Tabla 2.6**Resultados MCO del logaritmo natural de las ventas contra el logaritmo natural del precio.**

Variable Dependiente: LV

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10,13330	1,801677	5,624370	0,0000
LP	-0,942689	0,394212	-2,391324	0,0218
R-squared	0,130801	Mean dependent var		5,825448
Adjusted R-squared	0,107928	S.D. dependent var		0,191734
S.E. of regression	0,181092	Akaike info criterion		-0,530916
Sum squared resid	1,2	Schwarz criterion		-0,446472
Log likelihood	12,61832	Hannan-Quinn criterion		-0,500384
F-statistic	5,718433	Durbin-Watson stat		1,243529
Prob(F-statistic)	0,021844			

De este modo, se obtiene el modelo $\hat{L}V = 10,13 - 0,94lp$. La interpretación es la siguiente: Si el precio promedio de los libros se incrementa en uno por ciento (1%), las ventas de la editorial se reducirán en 0,94% y viceversa, manteniéndose todo lo demás constante.

Siguiendo el mismo procedimiento se obtiene el modelo log-nivel. En la Tabla 2.7 se muestra el conjunto de modelos de ventas de la editorial y sus interpretaciones. En conclusión, los modelos nivel-nivel y log-log son los más fáciles de interpretar.

Tabla 2.7**Interpretación de los coeficientes o parámetros de los modelos de ventas de la editorial.**

Modelo	Modelo econométrico	Interpretación
Nivel-nivel	$\hat{V} = 686,57 - 3,53P$	Cuando el precio promedio de los libros se incrementa en una unidad monetaria, la media de las ventas de la Editorial disminuye en 3,53 unidades monetarias y viceversa.

Modelo	Modelo econométrico	Interpretación
Nivel-log	$\hat{V} = 1.851,62 - 329,75 P$	Cuando el precio promedio de los libros expresado en logaritmo natural aumenta en 1%, la media de las ventas de la Editorial disminuye en 3,30 (329,75/100) unidades monetarias y viceversa.
Log-nivel	$\hat{I}\hat{V} = 6,80 - 0,010P$	Cuando el precio promedio de los libros se incrementa en una unidad monetaria, la media de las ventas de la Editorial expresada en logaritmo natural disminuye en 0,0001 (0,010/100) unidades monetarias y viceversa.
Log-log	$\hat{I}\hat{V} = 10,13 - 0,94IP$	Cuando el precio promedio de los libros expresado en logaritmo natural aumenta en 1%, las ventas de la media de las ventas de la Editorial expresadas en logaritmo natural disminuye en 0,94% y viceversa.

Fuente: Elaboración propia.

Si se analiza detenidamente el modelo log-log se deduce que las ventas de la editorial son muy sensibles al precio de los libros. Para confirmar esta hipótesis se calculara la elasticidad precio de la demanda.¹³ Para ello tomamos de las columnas 2 y 3 de la Tabla 2.1, los siguientes datos:

Ventas 2009:01 → \$ 275,5 millones → Q_1

Ventas 2018:04 → \$ 407,0 millones → Q_2

¹³ La elasticidad-precio de la demanda es la variación porcentual de la cantidad demandada dividida por la correspondiente variación porcentual de su precio, por lo tanto, EPD = variación porcentual de la cantidad demandada/variación porcentual del precio. Se debe tener presente que la elasticidad no es lo mismo que la pendiente de la recta de regresión lineal, dos conceptos de amplio uso en el análisis económico.

Precio promedio 2009:01 → \$ 98,6 miles → P_1

Precio promedio 2018:04 → \$ 97,7 miles → P_2

La elasticidad precio de la demanda se calcula utilizando la siguiente fórmula:

$$EPD = \frac{(Q_2 - Q_1)/(Q_2 + Q_1)/2}{(P_2 - P_1)/(P_1 + P_2)/2}$$

Reemplazando valores en la fórmula obtenemos:

$$EPD = \frac{(4.166 - 2.794)/(2.794 + 4.166)/2}{(97.700 - 98.600)/(98.600 + 97.700)/2} = -43 = |43|$$

Como la elasticidad precio de la demanda se expresa en valor absoluto, $EPD = 42 > 1$. En conclusión, la demanda de libros es elástica al precio, de forma que el precio y el ingreso por ventas varían en sentido contrario. Nótese que si el precio de los libros varía en una unidad monetaria, las ventas de la editorial variarán en 42 unidades monetarias en sentido contrario. Además, debe considerarse que los libros son bienes superiores, cuya demanda está relacionada con el nivel de renta de los individuos.

Don Pablo y Emilio deben considerar este aspecto y ser cautelosos en el momento de fijar precios de los libros.

2.1.3. Coeficiente de determinación (R^2).

El coeficiente de determinación (R^2) mide la proporción del total de la suma de cuadrados explicada (SCE) por la regresión. Este coeficiente da una idea de la bondad de ajuste. Siempre que en el modelo exista término independiente, el R^2 está situado entre cero y uno, cuanto más próximo esté a cero, peor será el ajuste, y cuanto más cercano esté de uno, mejor será el ajuste.

$$SCT = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SCE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

$$SCR = \sum_{i=1}^n \hat{\mu}_i^2$$

Dónde: SCR = Suma de cuadrados de los residuos; SCE = Suma de cuadrados explicada y SCT = Suma de cuadrados total.

Por lo tanto: $SCT = SCE + SCR$

Siguiendo a Wooldridge (2015), la variación total en y_i es la suma de la variación total en \hat{y}_i y en $\hat{\mu}_i$. Suponiendo que la variación total en y no sea cero, como es el caso, a menos que las y_i sean constantes en la muestra se puede dividir $SCT = SCE + SCR$ entre SCT para obtener:

$\frac{SCR}{SCT} + \frac{SCE}{SCT} = 1$, también podría expresarse del siguiente modo:

$R^2 = 1 - \frac{\text{Variación no explicada en Y}}{\text{Variación total en Y}} = \frac{\text{Variación explicada en Y}}{\text{Variación total en Y}}$, por lo tanto:

$$R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}; \quad 0 \leq R^2 \leq 1 \quad (2.5)$$

También se puede mostrar que el coeficiente que R^2 es igual al cuadrado del coeficiente de correlación entre las y_i reales y los valores ajustados de \hat{y}_i , es decir:

$$R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y}))^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\hat{y}_i - \bar{y})^2} \quad (2.6)$$

Aunque Emilio, no está en total desacuerdo con la idea de que el precio tiene una incidencia sobre las ventas, sostiene que la explicación que ofrece su padre es simplista. El efecto causal del precio promedio de los libros sobre las ventas de la Editorial aunque logra explicar cierta variación, ninguno de los dos personajes de la historieta hasta el momento conoce su magnitud. Para dar respuesta a esta inquietud, Emilio se dispone a demostrar a su padre, que si bien es cierto, que el precio ejerce una cierta

influencia sobre las ventas, su efecto es más bien modesto. Emilio, para sostener esta afirmación, calculará el coeficiente de determinación para dos variables (R^2), en un primer momento en notación matricial y después lo hará en EViews.

El coeficiente de determinación mide la bondad de ajuste de la línea de regresión a un conjunto de datos. “(...) es claro que si todas las observaciones cayesen en la línea de regresión, se obtiene un ajuste “perfecto”, pero rara vez se presenta este caso. Por lo general hay algunas μ_i positivas y algunas $\hat{\mu}_i$ negativas. Se tiene la esperanza de que estos residuos alrededor de la línea de regresión sean lo más pequeños posibles” (Gujarati, 2010: 73). El R^2 oscila entre 0 y 1. Cuanto más cerca de 1 se sitúe su valor, mayor será el ajuste del modelo a la variable que estamos intentando explicar. De forma inversa, cuanto más cerca de cero esté, menos ajuste tendrá el modelo, y por tanto, será menos fiable.

El coeficiente de determinación (R^2) es una medida descriptiva del ajuste global del modelo cuyo valor es el cociente entre la variabilidad explicada (suma de cuadrados explicada = SCE) y la variabilidad total (suma de cuadrados total = SCT), o sea, $R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT}$.

La suma de cuadrados total (SCT) es la varianza total de la variable endógena (salvo el factor tamaño muestral), y es por tanto una medida del tamaño de las fluctuaciones experimentadas por dicha variable alrededor de su valor medio. El objeto fundamental de todo modelo econométrico es tratar de explicar dichas fluctuaciones.

$$\text{Suma de cuadrados total (SCT)} = \sum_{i=1}^T (Y - \bar{Y})^2 = Y'Y - T\bar{Y}^2$$

La suma de cuadrados explicada (SCE) es el grado de fluctuación de la variable \hat{Y} , alrededor del promedio de Y. Por tanto, la suma explicada es el nivel de fluctuación de la variable Y que el modelo es capaz de explicar. Es la variación explicada por las variables exógenas.

$$\text{Suma de cuadrados explicada (SCE)} = \sum_{i=1}^T (\hat{Y} - \bar{Y})^2 = \hat{Y}'Y - T\bar{Y}^2$$

La suma de cuadrados de residuos (SCR) o suma residual (SR) es un indicador del nivel de error del modelo en su intento de explicar la evolución temporal de la variable Y .

$$\text{Suma de cuadrados de residuos (SCR)} = \sum_{i=1}^T (Y - \hat{Y})^2 = + \hat{\mu}'\mu$$

$$\text{SCR} = \sum_{i=1}^T (Y - \hat{Y})^2 = + \hat{\mu}'\mu = (Y - X\hat{B})'(Y - X\hat{B}) = Y'Y - \hat{B}'X'Y = Y'Y - \hat{Y}'Y$$

Luego se puede escribir la igualdad $Y'Y = \hat{Y}'Y + \hat{\mu}'\mu$, y si a los dos miembros de esta igualdad le restamos $T\bar{Y}^2$, se obtiene:

$$(Y'Y - T\bar{Y}^2) = (\hat{Y}'Y - T\bar{Y}^2) + \hat{\mu}'\mu, \text{ o sea, } \text{SCT} = \text{SCE} + \text{SCR}$$

Suma de cuadrados total (SCT) = suma de cuadrados explicada (SCE) + suma de cuadrados de residuos (SCR). A estos tres términos se le llama suma de cuadrados (SC). El coeficiente de determinación R^2 se ha definido como:

$$R^2 = \frac{\text{SCR}}{\text{SCT}}, \text{ donde:}$$

SCE: Suma de cuadrados explicada o suma de cuadrados debida a la regresión (debida a la variable explicativa o explicada por esta) $\rightarrow \sum(Y - \hat{Y})^2$

SCT: Suma de cuadrados total o variación de los valores de Y estimados alrededor de su media $\rightarrow \sum(Y - \bar{Y})^2$

SCR: Suma de cuadrados de residuos o variación residual o no explicada de los valores de Y alrededor de la línea de regresión $\rightarrow \sum(\hat{Y} - \bar{Y})^2$

Don Pablo, al ver este montón de fórmulas queda más asustado que antes. Mirando a su hijo, exclama: mira Emilio, he sido muy paciente contigo, pero cada vez me confundes más. Por favor, sé directo y claro en tu explicación.

Ante la petición de Don Pablo, Emilio considera pertinente calcular el R^2 a partir de la elaboración de la Tabla 2.8. Es momento de explicar cómo se estiman las ventas de la editorial (\hat{Y}) y cómo se calculan los residuos, es decir, la distancia existente entre cada punto (observación) y la recta de regresión lineal estimada (Figura 2.1). Las ventas estimadas (\hat{Y}) se calculan a partir de la ecuación de regresión lineal estimada arriba: $\widehat{\text{Ventas}} = 686,57 - 3,53\text{Precio}$. En esta ecuación simplemente se reemplaza la variable Precio = X por cada uno de los valores de la columna "Precio", es decir, de la primera fila de la tabla en mención, se

toma el precio de 98,6 y lo sustituimos en X, así: $\widehat{Ventas} = 686,57 - (3,53)(98,6) = 338,3$; en la segunda fila: $\widehat{Ventas} = 686,57 - (3,53)(96,6) = 345,3$ y así, sucesivamente.

El lector habrá notado que los valores de ventas (Y) y ventas estimadas (\hat{Y}) no son exactamente iguales, por ejemplo, en la primera fila de la Tabla 2.8, $Y = 275,5$ y $(\hat{Y}) = 338,3$, precisamente la diferencia entre las ventas y las ventas estimadas corresponde a un residuo: $Y - (\hat{Y})$, en el caso de la editorial: $275,5 - 338,3 = -62,7744$, por lo tanto, este valor se encuentra, un tanto alejado de la recta de regresión lineal y situado por debajo de dicha recta. El propósito de MCO es minimizar la suma de las distancias verticales entre las respuestas observadas en la muestra y las respuestas del modelo; en otras palabras, minimizar la suma de cuadrados de los residuos, definidos como la diferencia entre el valor observado de la variable que se trata de explicar (ventas) y el valor estimado por la recta ajustada ($\widehat{Ventas} = 686,57 - 3,53\text{Precio}$).

Para calcular el R^2 de la recta de regresión lineal basta con reemplazar en la fórmula (2.6), los valores extractados de la Tabla 2.8, tales como: $SCR = \sum(\hat{Y} - \bar{Y})^2 = 22.688,07$ y la $SCT = \sum(Y - \bar{Y})^2 = 156.647,44$. Por lo tanto, $R^2 = \frac{SCR}{SCT}$.

$$R^2 = \frac{\sum_{i=1}^T(\hat{Y} - \bar{Y})^2}{\sum_{i=1}^T(Y - \bar{Y})^2} \quad (2.7)$$

Al reemplazar valores se obtiene:

$$R^2 = \frac{22.688,07}{156.647,44} = 0,1448 \times 100 = 14,48\%$$

Interpretación:

La variabilidad total de las ventas de la editorial respecto a su media es explicada en un 14,48% por las variaciones en el precio promedio de los libros. Si bien, el precio solo explica el 14,48% de las ventas, el resto de la proporción de la variabilidad de las ventas ($100\% - 14,48\% = 85,52\%$) será explicada por otras variables distintas al precio, las cuales serán incluidas en el modelo en el capítulo 3.

2.1.4. El coeficiente de determinación ajustado (\bar{R}^2).

De acuerdo a la teoría econométrica, un modelo será mejor cuanto mayor sea su coeficiente de determinación (R^2). Como se observa en el caso de la editorial, el coeficiente de determinación depende mucho de nuevas variables introducidas en el modelo, teniendo cuidado de que éstas no empeoren la calidad de la regresión. Este problema se arregla sustituyendo el coeficiente de determinación (R^2) por el coeficiente de determinación ajustado (\bar{R}^2), el cual para muestras grandes no depende del número de variables del modelo. El coeficiente de determinación corregido o ajustado (\bar{R}^2) penaliza la entrada de nuevas variables explicativas al modelo.

$$\bar{R}^2 = 1 - \frac{N-1}{N-K-1} (1 - R^2) \quad (2.8)$$

Dónde:

\bar{R}^2 : Coeficiente de determinación ajustado

N : Tamaño de la muestra

K : Número de variables explicativas o exógenas (en nuestro modelo, la variable Precio = X)

R^2 : Coeficiente de determinación

Tabla 2.8.

Información para la estimación del coeficiente de determinación.

Trimestre	Ventas = Y	Precio X	$(Y - \bar{Y})^2$	(\hat{Y})	$(Y - \hat{Y})^2$	$(Y - \hat{Y})$	$(\hat{Y} - \bar{Y})^2$
2009:1	275,5	98,6	4794,5	338,3	3940,62	-62,7744	41,8368
2009:2	285,6	96,6	3497,8	345,3	3568,77	-59,7392	0,3560
2009:3	336,8	102	63,1	326,3	111,00	10,5358	341,4476
2009:4	333,4	91,9	128,7	361,9	814,61	-28,5414	295,8035
2010:1	357	105,6	150,2	313,5	1888,11	43,4524	973,1243
2010:2	325,2	102,8	381,9	323,4	3,10	1,7617	453,8698
2010:3	362,2	102,1	304,8	325,9	1316,89	36,2890	354,6269
2010:4	232	99,1	12710,9	336,5	10921,96	-104,5082	67,8044
2011:1	252,2	100,3	8564,1	332,3	6411,09	-80,0693	155,5812
2011:2	322,1	104,8	512,7	316,4	32,79	5,7265	804,8013
2011:3	297,3	85,6	2250,8	384,2	7550,84	-86,8956	1556,5439
2011:4	298,9	103	2101,5	322,7	567,95	-23,8318	484,4709
2012:1	246,8	100,4	9592,7	331,9	7244,74	-85,1160	164,5181

Continuación Tabla 2.8.

Trimestre	Ventas = Y	Precio = X	$(Y - \bar{Y})^2$	(\hat{Y})	$(Y - \hat{Y})^2$	$(Y - \hat{Y})$	$(\hat{Y} - \bar{Y})^2$
2012:2	322,4	93,8	499,2	355,2	1077,80	-32,8299	109,9851
2012:3	383,5	99,7	1502,1	334,4	2411,92	49,1113	107,2008
2012:4	321,8	104,6	526,4	317,1	22,28	4,7200	765,2161
2013:1	351,5	100	45,7	333,3	330,19	18,1710	130,2680
2013:2	381,1	103,7	1321,9	320,3	3701,61	60,8409	599,4359
2013:3	412,5	96,7	4591,1	345,0	4558,15	67,5141	0,0593
2013:4	217,5	105,5	16190,7	313,9	9293,11	-96,4008	951,2105
2014:1	229,9	99,2	13188,8	336,2	11290,11	-106,2549	73,7465
2014:2	342,9	86,4	3,4	381,4	1479,91	-38,4696	1341,5474
2014:3	361,9	92,7	294,4	359,1	7,75	2,7845	206,5837
2014:4	301,3	100,9	1887,3	330,1	832,31	-28,8498	212,9457
2015:1	332,5	101,7	149,9	327,3	26,79	5,1761	303,4069
2015:2	343,2	95,3	2,4	349,9	45,31	-6,7313	26,9234
2015:3	421,9	98,6	5953,3	338,3	6993,25	83,6256	41,8368
2015:4	401,9	102,5	3267,0	324,5	5991,07	77,4020	409,8398
2016:1	421,8	101	5937,9	329,8	8464,63	92,0034	223,3799
2016:2	361,7	96,9	287,6	344,3	303,48	17,4206	0,2144
2016:3	393,6	98,3	2387,1	339,3	2944,79	54,2659	29,2510
2016:4	287,6	91,8	3265,3	362,3	5579,30	-74,6947	308,0790
2017:1	380	77,8	1243,1	411,7	1007,95	-31,7483	4489,7746
2017:2	495,6	81,6	22758,0	398,3	9462,39	97,2748	2871,1015
2017:3	452,6	86,6	11633,2	380,7	5174,91	71,9368	1290,2938
2017:4	404,9	93	3618,9	358,1	2194,38	46,8442	177,2440
2018:1	421,3	83,7	5861,1	390,9	923,73	30,3929	2131,1721
2018:2	333,7	93,7	121,9	355,6	478,87	-21,8831	117,5190
2018:3	379,1	94,5	1180,4	352,8	693,94	26,3428	64,2354
2018:4	407	97,7	3876,0	341,5	4296,34	65,5465	10,8174
Sumatoria	13.789,7	3.870,7	156.647,4	13.789,9	133.958,8	- 0,1993	22.688,0728
Promedio	344,7	96,8	SCT		SCE	Residuos	SCR

Fuente: Elaboración propia a partir de los datos de la Tabla 1.2

Se observa que cuando $N \rightarrow \infty$, o sea, para muestras grandes $(N-1)(N-k-1) \rightarrow 1$ y no depende de k , que es el número de variables del modelo. Además, $N \rightarrow \infty \rightarrow \bar{R}^2 \rightarrow R^2$.

Ahora ya se puede considerar a R^2 como una buena medida de la calidad de la regresión. El modelo será tanto mejor cuanto mayor sea el coeficiente de determinación ajustado (\bar{R}^2).

Emilio procede a calcular el \bar{R}^2 para el modelo de las ventas de la editorial, así:

$$R^2 = 0,1448$$

$$N = 40$$

$$k = 1$$

Reemplazando en la fórmula (2.7), se obtiene:

$$\bar{R}^2 = 1 - \frac{40-1}{40-1-1} (1 - 0,1448)$$

$$\bar{R}^2 = 0,1223 \times 100 = 12,23\%$$

Interpretación:

La variabilidad total de las ventas de la editorial respecto a su media es explicada en un 12,23% por las variaciones en el precio promedio de los libros. Si bien, el precio solo explica el 12,23% de las ventas, el resto de la proporción de la variabilidad de las ventas ($100,00\% - 12,23\% = 87,77\%$) es explicada por otras variables distintas al precio.

Donde N es el tamaño de la muestra y k el número de variables explicativas. Por deducción matemática, a valores más altos de k , más alejado estará el \bar{R}^2 del R^2 normal. Y al revés, a valores más bajos de k , más cerca estará de uno la fracción central, por consiguiente, más parecidos serán \bar{R}^2 y R^2 . El lector recordará que k es el número de variables explicativas introducidas en el modelo; por consiguiente, se deduce que si k fuera igual a cero, el coeficiente de determinación sería también cero; es decir, no habría modelo alguno (en un modelo cualquiera, como mínimo se tendría que explicar una variable en función de otra). Dado que k debe ser como mínimo 1, el \bar{R}^2 y el R^2 no pueden tener el mismo valor; el \bar{R}^2 será siempre inferior al R^2 .

En la Tabla 2.2, se observa que EViews al momento de calcular la recta de regresión entre la información que arroja, incluye el R^2 (*R-squared*) y el \bar{R}^2 (*Adjusted R-squared*). Emilio siente satisfacción al observar los coeficientes de determinación del modelo econométrico de la editorial: $\widehat{Ventas} = \$686,57 - \$3,53\text{Precio}$. Emilio mirando a los ojos a Don Pablo y de manera sonriente exclama: ¡Mira papa!, yo tenía razón, el precio no

es una variable suficientemente importante para explicar las ventas; si se analiza el \bar{R}^2 , el precio únicamente explica el 12,23% de las variaciones en las ventas de la editorial y el 14,48% si se toma el R^2 ; es decir, más de 4/5 partes de las ventas no son explicadas por el precio, lo que sugiere que hay otros factores que también influyen en las ventas, las cuales serán consideradas en el modelo que se especifique y estime en el siguiente capítulo.

2.1.5. Coeficiente de correlación.

Don Pablo como es un hombre obstinado, le propone a su hijo que se mida la fuerza o el grado de asociación lineal entre el precio promedio de los libros y las ventas de la editorial. Emilio acepta el reto y decide hacerlo utilizando el coeficiente de correlación. Antes de iniciar con el cálculo de este coeficiente, Emilio considera conveniente establecer las diferencias entre lo realizado con la regresión y el procedimiento actual que se pretende emprender. Emilio, apoyado en el manual de econometría del profesor Gujarati (2010) afirma: “La regresión y la correlación presentan diferencias fundamentales que vale la pena mencionar. En el análisis de regresión hay una asimetría en el tratamiento a las variables dependientes y explicativas. Se supone que la variable dependiente es estadística, aleatoria o estocástica, es decir, que tiene una distribución de probabilidad. Por otra parte, se asume que las variables explicativas tienen valores fijos (en muestras repetidas), (...) En el análisis de correlación, por otra parte, se tratan dos variables cualesquiera en forma simétrica; no hay distinción entre las variables dependiente y explicativa” (p. 20). El coeficiente de correlación lineal puede tomar valores comprendidos entre -1 y 1.

Para el cálculo en forma manual del coeficiente de correlación se desarrolla el siguiente procedimiento, a partir de los datos contenidos en la Tabla 2.9, para ello:

1. Hallamos la medias aritméticas: $\bar{X}_i = 96,8$ y $\bar{Y}_i = 344,7$
2. Calculamos la covarianza: $\sigma_{xy} = \sum(X_i - \bar{X}_i)(Y_i - \bar{Y}_i)/n$

Reemplazando se obtiene: $1.327.971,9/40 - (96,8)(344,7) = -160,57$

3. Calculamos las desviaciones típicas:

$$\sigma_x = \sqrt{\frac{\sum X_i^2}{n} - \bar{X}_i^2} = \sqrt{\frac{376.376,2}{40} - 9.363,9} = 6,7$$

$$\sigma_y = \sqrt{\frac{\sum Y_i^2}{n} - \bar{Y}_i^2} = \sqrt{\frac{4.910.543,1}{40} - 118.847,4} = 62,6$$

4. Por último, se calcula el coeficiente de correlación lineal

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{-160,57}{(6,7)(62,6)} = -0,38$$

Tabla 2.9.

Información para la estimación del coeficiente de correlación lineal.

Trimestre	Ventas = Y	Precio = X	$X_i Y_i$	X_i^2	Y_i^2
2009:1	275,5	98,6	27.164,3	9.722,0	75.900,3
2009:2	285,6	96,6	27.589,0	9.331,6	81.567,4
2009:3	336,8	102,0	34.353,6	10.404,0	113.434,2
2009:4	333,4	91,9	30.639,5	8.445,6	111.155,6
2010:1	357	105,6	37.699,2	11.151,4	127.449,0
2010:2	325,2	102,8	33.430,6	10.567,8	105.755,0
2010:3	362,2	102,1	36.980,6	10.424,4	131.188,8
2010:4	232	99,1	22.991,2	9.820,8	53.824,0
2011:1	252,2	100,3	25.295,7	10.060,1	63.604,8
2011:2	322,1	104,8	33.756,1	10.983,0	103.748,4
2011:3	297,3	85,6	25.448,9	7.327,4	88.387,3
2011:4	298,9	103,0	30.786,7	10.609,0	89.341,2
2012:1	246,8	100,4	24.778,7	10.080,2	60.910,2
2012:2	322,4	93,8	30.241,1	8.798,4	103.941,8
2012:3	383,5	99,7	38.235,0	9.940,1	147.072,3
2012:4	321,8	104,6	33.660,3	10.941,2	103.555,2
2013:1	351,5	100,0	35.150,0	10.000,0	123.552,3
2013:2	381,1	103,7	39.520,1	10.753,7	145.237,2

Continuación Tabla 2.9.

Trimestre	Ventas = Y	Precio = X	$X_i Y_i$	X_i^2	Y_i^2
2013:3	412,5	96,7	39.888,8	9.350,9	170.156,3
2013:4	217,5	105,5	22.946,3	11.130,3	47.306,3
2014:1	229,9	99,2	22.806,1	9.840,6	52.854,0
2014:2	342,9	86,4	29.626,6	7.465,0	117.580,4
2014:3	361,9	92,7	33.548,1	8.593,3	130.971,6
2014:4	301,3	100,9	30.401,2	10.180,8	90.781,7
2015:1	332,5	101,7	33.815,3	10.342,9	110.556,3
2015:2	343,2	95,3	32.707,0	9.082,1	117.786,2
2015:3	421,9	98,6	41.599,3	9.722,0	177.999,6
2015:4	401,9	102,5	41.194,8	10.506,3	161.523,6
2016:1	421,8	101,0	42.601,8	10.201,0	177.915,2
2016:2	361,7	96,9	35.048,7	9.389,6	130.826,9
2016:3	393,6	98,3	38.690,9	9.662,9	154.921,0
2016:4	287,6	91,8	26.401,7	8.427,2	82.713,8
2017:1	380,0	77,8	29.564,0	6.052,8	144.400,0
2017:2	495,6	81,6	40.441,0	6.658,6	245.619,4
2017:3	452,6	86,6	39.195,2	7.499,6	204.846,8
2017:4	404,9	93,0	37.655,7	8.649,0	163.944,0
2018:1	421,3	83,7	35.262,8	7.005,7	177.493,7
2018:2	333,7	93,7	31.267,7	8.779,7	111.355,7
2018:3	379,1	94,5	35.825,0	8.930,3	143.716,8
2018:4	407,0	97,7	39.763,9	9.545,3	165.649,0
Sumatoria	13.789,7	3.870,7	1.327.971,9	376.376,2	4.910.543,1
Promedio	344,7	96,8			

Fuente: Elaboración propia.

Como el coeficiente de correlación se encuentra entre -1 y 1 ($-1 \leq r \leq 1$), el coeficiente de -0,38 se interpreta como una relación lineal entre el precio y las ventas como débil (más cercana a 0) e inversa (signo negativo).

Después de calcular el coeficiente de correlación lineal de manera manual, Emilio procede a ejecutar su cálculo a través de EViews para luego elaborar un gráfico de dispersión que relacione el precio con las ventas (Figura 2.22), así como una matriz de correlación. Para ello se abre el archivo ventas.wfl y se selecciona las variables precio y ventas en este orden, dando clic derecho

en el ratón se elige *Open* → *as Group*. Inmediatamente se abre una ventana con la lista de las variables precio y ventas, tal como se muestra en la Figura 2.23.

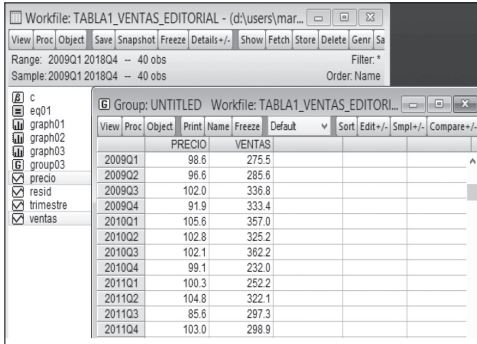


Figura 2.22

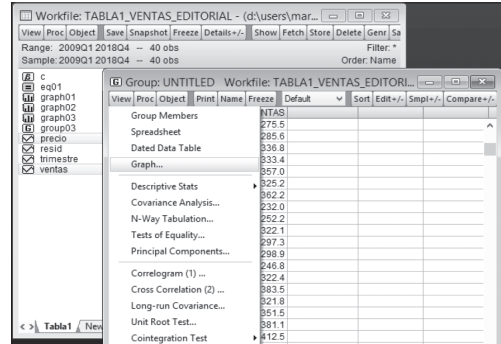


Figura 2.23

Posteriormente, se selecciona *Views* → *Graph*, se elige *Scatter* como se muestra en la Figura 2.24 y pulsamos OK. En la Figura 2.25 se muestra una nube de puntos resultante de la combinación de las variables $X = \text{precio}$ y $Y = \text{ventas}$. El diagrama de dispersión sugiere que existe una relación inversa entre el precio y las ventas; a simple vista se observa que existe una débil relación entre las variables en cuestión.

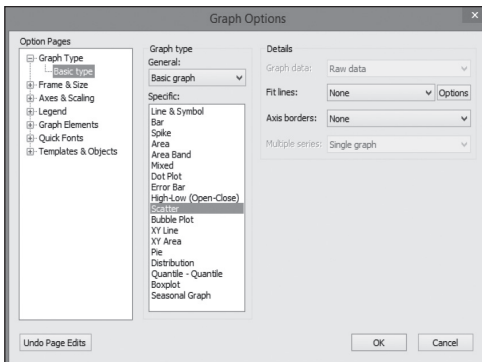


Figura 2.24

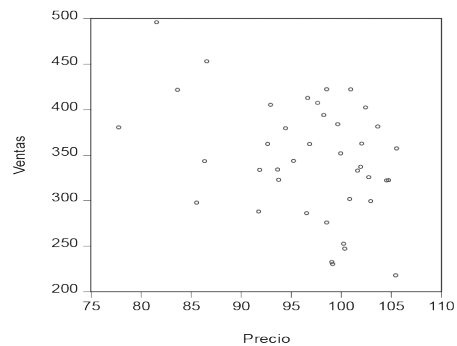


Figura 2.25

Para medir la intensidad o fuerza entre el precio y las ventas, se selecciona las variables precio y ventas en este orden y con el clic derecho del ratón se selecciona *as Group* → *View* → *Covariance analysis* y se da clic (Figura 2.26). Enseguida, se señala *Correlation* como se muestra en la Figura 2.27 y se pulsa OK.

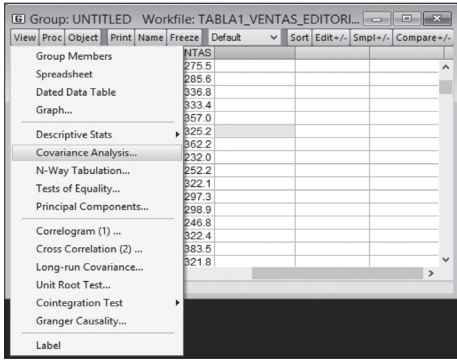


Figura 2.26

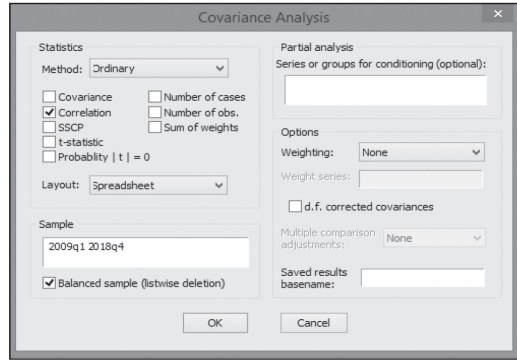


Figura 2.27

Enseguida, se obtiene la matriz de correlación que se muestra en la Tabla 2.10, donde se aprecia que el coeficiente de correlación lineal es de -0,380577.

Tabla 2.10

Coefficiente de correlación lineal

	Precio	Ventas
Precio	1,000000	-0,380577
Ventas	-0,380577	1,000000

En este apartado la discusión inicial entre Don Pablo y Emilio queda planteada en el modelo que aparece a continuación.

$$V_t = \beta_0 + \beta_1 P_t + \mu_t \quad (2.9)$$

A partir de un procedimiento manual o de un proceso automatizado en EViews, la estimación del modelo es la siguiente:

$$\widehat{Ventas} = 686,57 - 3,53 \text{Precio} \quad (2.10)$$

(5,083) (-2,537)

$$R^2 = 14,48\% \quad \bar{R}^2 = 12,23\%$$

El resultado anterior es sólo parte de la información que EViews proporciona para la estimación de este modelo. Si se analiza el conjunto de información (véase Tabla 2.2) que se obtiene con el software, se puede observar, que además, de la columna que contiene los valores de los coeficientes estimados (*Coefficient*), también se proporciona una columna

que recoge las desviaciones típicas estimadas (*Std. Error*) y otra con el valor de la razón $t=\beta/s_{\hat{\beta}}$, (*t-Statistic*), que son los valores situados debajo de cada coeficiente en la ecuación (2.9).

Por su parte, la ecuación (2.10) muestra que existe una relación negativa entre precio y ventas, tal como Don Pablo lo sostenía desde un comienzo. Por otro lado, los valores de R^2 y \bar{R}^2 son bajos, lo que hace suponer que, aunque el precio pueda ser relevante en el comportamiento de las ventas, no es el único factor de interés. Se tiene que concluir, que aunque la variable “*Precio*” debe ser incluida en el modelo econométrico, sería conveniente añadir al modelo otros factores que puedan ser relevantes en la explicación de las ventas de la editorial.

RETO 1:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

**ENCARANDO EL MODELO DE
REGRESIÓN LINEAL GENERAL:
ESPECIFICACIÓN Y ESTIMACIÓN**

Supuesto 2: *Los valores de X son fijos en muestreo repetido: Los valores que toma el regresor X son considerados fijos en muestreo repetido. Técnicamente, X se supone no estocástica.*

Supuesto 3: *El valor medio de la perturbación u_i es igual a cero. Dado el valor de X , la media o el valor esperado del término aleatorio de perturbación u es cero. Técnicamente, el valor de la media condicional de u es cero. Simbólicamente, se tiene que $E(u_i \mid X_i) = 0$*

Supuesto 4: *El número de observaciones n debe ser mayor que el número de parámetros por estimar: Sucesivamente, el número de observaciones n debe ser mayor que el número de variables explicativas.*

Supuesto 5: *La naturaleza de las variables X : No todos los valores X en una muestra determinada deben ser iguales. Técnicamente, $\text{var}(X)$ debe ser un número positivo. Además, no puede haber valores atípicos de la variable X , es decir, valores muy grandes en relación con el resto de las observaciones.*

Supuesto 6: *El modelo de regresión está correctamente especificado. Alternativamente, no hay un sesgo de especificación o error en el modelo utilizado en el análisis empírico.*

En el capítulo anterior en el caso de la Editorial, se establece la relación entre el precio de los libros y las ventas totales, así:

$$\mathcal{V}_t = \beta_0 + \beta_1 \mathcal{P}_t + \mu_t$$

Dónde: \mathcal{V}_t es el número de libros vendidos y \mathcal{P}_t es el precio promedio al que los vende la editorial. Como se mencionó antes, este corresponde a un modelo de regresión simple o bivariado, compuesto por una variable dependiente y una variable explicativa.

Por su parte, el modelo de regresión lineal múltiple es más complejo y de mayor uso en las ciencias; su propósito es explicar el comportamiento de una variable (variable dependiente) a partir de otras (variables explicativas). Su estructura es la siguiente:

$$V_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_i X_{it} + \dots + \beta_k X_{kt} + \mu_t \quad t = 1, 2, 3, \dots, T \quad (3.1)$$

La variable Y_t se denomina variable dependiente o endógena, las variables X_{it} , $i = 1, \dots, K$ son las variables explicativas y μ_t es la perturbación aleatoria, β_{it} son los $(K+1)$ parámetros asociados a cada una de las variables explicativas, llamados también coeficientes de regresión, los cuales miden el impacto de cada variable explicativa en el comportamiento de la variable dependiente. β_0 es el término independiente o parámetro asociado a una variable explicativa que toma el valor de 1 para todas las observaciones, es decir, cuando el valor de las X_i es igual a cero, Y_t es igual al valor que toma β_0 .

En notación matricial el modelo de regresión lineal múltiple se escribe como $Y = X\beta + \mu$, donde:

$$\begin{array}{c}
 \begin{bmatrix} Y_1 \\ \vdots \\ Y_t \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{21} & \dots & X_{k1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{12} & \dots & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & \dots & X_{2n} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} + \begin{bmatrix} \hat{\mu}_1 \\ \vdots \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_n \end{bmatrix} \\
 \begin{array}{ccc}
 \mathcal{Y} & = & X \\
 n \times 1 & & n \times k & & k \times 1 & & n \times 1
 \end{array}
 \end{array}$$

Carrascal et al., (2004), señalan que el modelo de regresión lineal múltiple cumple las siguientes hipótesis clásicas:

- La linealidad de la relación y la constancia de los parámetros a lo largo de la muestra están implícitas en la especificación de la ecuación del modelo.
- No existen relaciones lineales exactas entre las variables explicativas, y estas, no son variables aleatorias.
- Las distribuciones de probabilidad del vector de perturbaciones aleatorias es $\varepsilon \rightarrow N(0, \sigma_t^2, I)$.

Volviendo a la discusión entre padre e hijo, se expondrá ahora la posición de Emilio. Aunque Emilio no está en desacuerdo con la idea de que el precio tiene una incidencia sobre las ventas de la editorial, sostiene que la explicación que da su padre es simplista y que deben añadirse más elementos para poder explicar correctamente lo que ha venido sucediendo. Su afirmación central es que el mundo es más complejo de lo que su padre parece creer. La *competencia* ha sido bastante fuerte en los últimos años y los precios a los que vende la editorial son competitivos, por ende, esta es una variable que debe tenerse muy en cuenta. De la misma manera, la gente es bombardeada constantemente con publicidad, de manera que es necesario tener en cuenta también su efecto. Emilio está convencido de que el gasto en *publicidad* que ha venido realizando la editorial, a pesar de la protesta de su padre, ha sido importante para explicar las ventas que se han logrado. Intuitivamente,

también considera que la renta de los clientes influye de alguna manera en las ventas de la editorial. Todo ello significa que, de acuerdo con su teoría, además del precio, hay otras variables que afectan a las ventas, por lo que es necesario reformular la relación entre variables de la siguiente manera:

$$V_t = \beta_0 - \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t + u_t \quad t = 1, 2, 3, \dots, T \quad (3.2)$$

Donde: V_t representa las ventas de libros, PE_t el precio promedio al que los vende la editorial, PC_t el precio promedio de la competencia (precio al que venden otras editoriales), GP_t el gasto realizado en publicidad por la editorial y R_t la renta o ingreso promedio de los clientes de la editorial, partiendo de la información contenida en la Tabla 3.1. Los signos de los coeficientes señalan la relación que se sospecha exista entre la variable dependiente y cada una de las variables explicativas.

Se procede a realizar sendos gráficos de las variables del modelo para observar su comportamiento en el periodo de análisis. Para la elaboración de gráficos en EViews, se toma la información disponible en el archivo de trabajo *Tab2*; en la ventana del Workfile se selecciona en el siguiente orden las variables V, PE, PC, GP y R oprimiendo la techa CTRL y el puntero del ratón. Enseguida, se abre estas variables como grupo con *View* → *Open* → *as Group* y damos clic como se indica en la Figura 3.1. En la ventana de las variables elegimos *View* → *Graph* y damos clic (Figura 3.2).

Tabla 3.1
Ventas trimestrales de libros 2009-2018 (pesos constantes de 2009).

Trimestre	Ventas (millones de \$)	Precio promedio Editorial (miles de pesos)	Precio promedio competencia (miles de pesos)	Gasto publicidad (miles de pesos)	Renta promedio del consumidor (miles de pesos)
2009:1	275,5	98,6	105,5	1.868,0	895,0
2009:2	285,6	96,6	104,4	2.157,0	957,7
2009:3	336,8	102,0	115,6	2.541,0	1.024,7
2009:4	333,4	91,9	102,7	2.561,0	1.075,9
2010:1	357,0	105,6	106,1	3.103,0	1.129,7
2010:2	325,2	102,8	100,2	2.661,0	1.186,2
2010:3	362,2	102,1	97,2	2.757,0	1.245,5
2010:4	232,0	99,1	93,3	1.403,0	1.307,8
2011:1	252,2	100,3	94,6	1.856,0	1.386,3
2011:2	322,1	104,8	109,4	2.123,0	1.455,6
2011:3	297,3	85,6	94,7	2.181,0	1.528,3
2011:4	298,9	103,0	111,4	2.520,0	1.604,8
2012:1	246,8	100,4	105,0	2.134,0	1.685,0
2012:2	322,4	93,8	118,3	2.473,0	1.786,1
2012:3	383,5	99,7	102,1	3.125,0	1.875,4
2012:4	321,8	104,6	110,6	2.753,0	1.969,2
2013:1	351,5	100,0	97,3	2.869,0	2.067,6
2013:2	381,1	103,7	91,8	3.301,0	2.171,0
2013:3	412,5	96,7	96,8	3.043,0	2.279,6
2013:4	217,5	105,5	102,2	1.538,0	2.370,8
2014:1	229,9	99,2	108,5	1.903,0	2.465,6
2014:2	342,9	86,4	98,1	2.489,0	2.564,2
2014:3	361,9	92,7	122,8	2.770,0	2.666,8
2014:4	301,3	100,9	106,9	2.493,0	2.773,4
2015:1	332,5	101,7	108,3	2.701,0	2.939,8
2015:2	343,2	95,3	114,4	2.497,0	3.116,2
2015:3	421,9	98,6	114,9	3.192,0	3.303,2
2015:4	401,9	102,5	107,5	3.086,0	3.501,4
2016:1	421,8	101,0	108,5	3.533,0	3.676,5
2016:2	361,7	96,9	109,4	2.864,0	3.860,3
2016:3	393,6	98,3	114,6	3.181,0	4.053,3
2016:4	287,6	91,8	118,2	1.855,0	4.256,0
2017:1	380,0	77,8	106,8	2.575,0	4.468,8
2017:2	495,6	81,6	131,4	3.723,0	4.692,2
2017:3	452,6	86,6	101,5	3.268,0	4.879,9
2017:4	404,9	93,0	146,2	3.025,0	5.075,1
2018:1	421,3	83,7	114,2	3.173,0	5.227,4
2018:2	333,7	93,7	119,3	2.387,0	5.384,2
2018:3	379,1	94,5	116,5	2.839,0	5.545,7
2018:4	407,0	97,7	113,5	3.352,0	5.712,1

Fuente: El autor con base a la información proporcionada por Alonso et. al., (2005). En esta tabla se ha incorporado una nueva variable: la renta promedio de los clientes de la editorial.

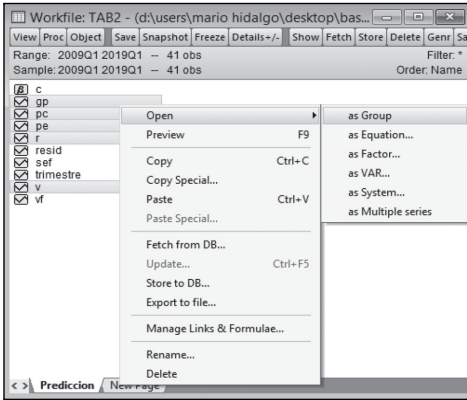


Figura 3.1

Group Members	PE	PC	GP	R
Spreadsheet	98.6	105.5	1888	895.0000
Dated Data Table	96.6	104.4	2157	957.6500
Graph...	102.0	115.6	2541	1024.686
	91.9	102.7	2561	1075.920
	105.6	106.1	3103	1129.716
Descriptive Stats	102.8	100.2	2661	1186.202
Covariance Analysis...	102.1	97.2	2757	1245.512
N-Way Tabulation...	99.1	93.3	1403	1307.787
Tests of Equality...	100.3	94.6	1856	1386.254
Principal Components...	104.8	109.4	2123	1455.567
Correlogram (1) ...	85.6	94.7	2181	1528.346
Cross Correlation (2) ...	103.0	111.4	2520	1604.763
Long-run Covariance...	100.4	105.0	2134	1685.001
Unit Root Test...	93.8	118.3	2473	1786.101
Cointegration Test	99.7	102.1	3125	1875.406
Granger Causality...	104.6	110.6	2753	1969.176
	100.0	97.3	2869	2067.635
	103.7	91.8	3301	2171.017
	96.7	96.8	3043	2279.668
	105.5	102.2	1538	2370.750
	99.2	108.5	1903	2465.581
Label	86.4	98.1	2489	2564.204

Figura 3.2

En el cuadro de opciones de gráficos (*Graph Options*) que se abre (por defecto gráfico de líneas), en la pestaña *Multiple series* se elige *Multiple graph* y damos OK, tal como se indica en la Figura 3.3.

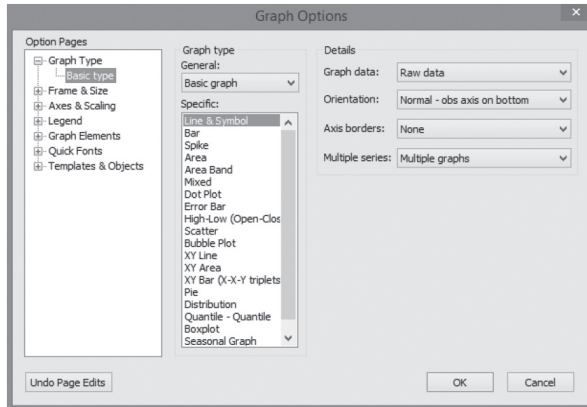


Figura 3.3

De este modo, se generan los gráficos de las series temporales utilizadas en el modelo (véase Figura 3.4). Se aprecia que las ventas de la editorial en el periodo 2009-2018 tienen una tendencia creciente debido posiblemente a la disminución progresiva del precio promedio de los libros (se mueven en dirección contraria) y al aumento del precio medio de la competencia. Por su parte, el comportamiento del gasto en publicidad es muy similar al de las

ventas (se mueven en la misma dirección), lo que permite aproximarse a la conclusión, de que a mayor publicidad, mayor volumen de ventas. La renta personal promedio de los clientes ha tenido un crecimiento notorio, probablemente este factor también ha favorecido las ventas de la editorial.

El análisis gráfico debe ser considerado como una primera aproximación a la comprensión de las relaciones entre las variables del modelo, aunque el análisis numérico será el que permita aceptar o rechazar las hipótesis que vayan surgiendo en el desarrollo de la historia de la editorial.

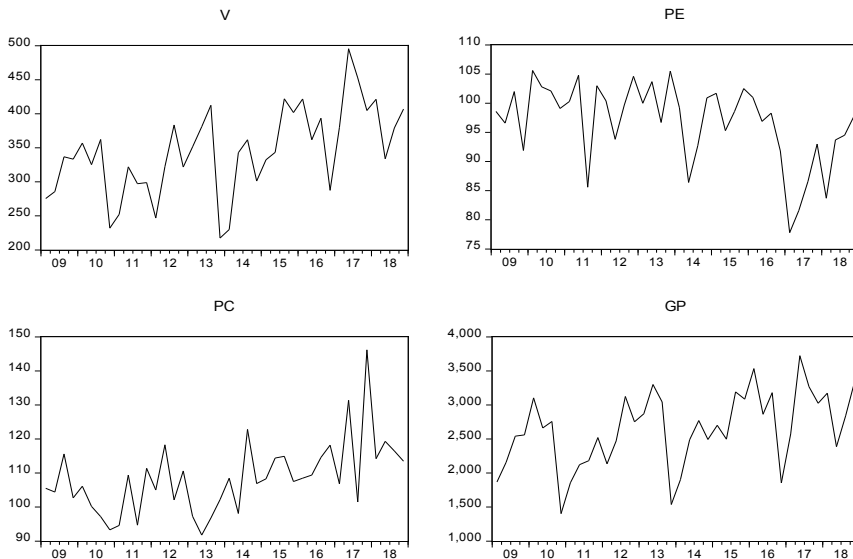


Figura 3.4

También sería interesante calcular la tasa de crecimiento anual media (TCAM)¹⁴ de las variables incorporadas en el modelo para comparar su comportamiento en el transcurso del tiempo. La Tabla 3.2 muestra que en el periodo 2009-2018 las ventas crecieron a una tasa anual media de 3,46% probablemente gracias a la disminución de precios (-1,14%), al aumento de

¹⁴ La tasa de crecimiento anual media se calcula como la derivada del logaritmo natural de las variables en cuestión con respecto al tiempo.

los precios de la competencia (1,69%), a un mayor gasto en publicidad (3,36%) y al buen comportamiento de la renta personal de los clientes de la editorial (19,22%). Nótese que en el análisis preliminar de estas cifras, se están lanzando juicios apresurados carentes de soporte econométrico.

Tabla 3.2

Tasa de crecimiento anual media de las variables del modelo.

Periodo	Ventas	PE	PC	GP	R
2009-2018	3,46%	-1,14%	1,69%	3,36%	19,22%

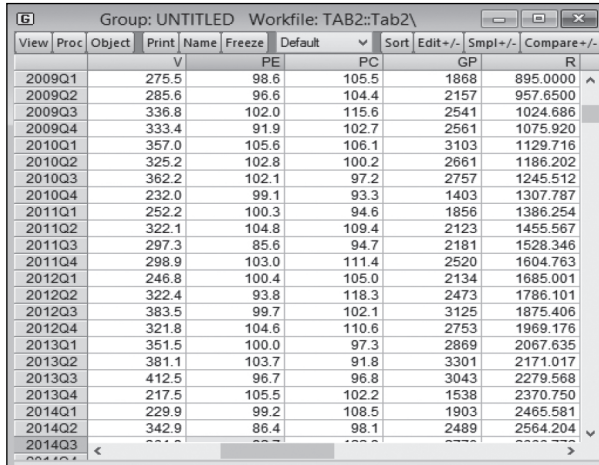
Fuente: Elaboración propia..

En el caso del modelo de regresión lineal simple se observa valores para la variable dependiente Y y para la variable explicativa X , es decir, nuestro problema consistía simplemente ver cómo las ventas variaban ante cambios en el precio promedio de los libros. Ahora, se cuenta con observaciones de un conjunto completo de variables explicativas y hay que enfrentar con la compleja tarea de aislar los efectos de cada una de ellas entre sí, es decir, se debe estimar el efecto individual que cada una de las variables explicativas ejerce sobre las ventas. Este nuevo modelo nos servirá para estimar cómo las ventas de libros dependen de cada una de las variables explicativas; en otras palabras, se debe desarrollar una técnica mediante la cual se puede estimar los valores de los coeficientes de β_0 , β_1 , β_2 , β_3 y β_4 .

3.1 Especificación del modelo.

Antes de especificar un modelo, es aconsejable hacer una inspección gráfica y numérica de las series temporales para analizar el grado y la forma de la relación existente entre las variables. A partir de la información de la Tabla 3.1 y con el apoyo de EViews se procede a realizar la siguiente rutina: al igual que hicimos con el modelo de regresión lineal simple procedemos a cargar el archivo del siguiente modo: Abrimos EViews y elegimos la opción:

Open → an existing EViews workfile, buscamos el archivo de trabajo *Tab2* en la carpeta que haya sido guardado y se procede a abrirlo. Posteriormente, se selecciona en el siguiente orden las siguientes variables: *V*, *PE*, *PC*, *GP* y *R*. se da clic derecho y se selecciona *Open* → *as Group*, se da clic y se visualiza las variables del modelo en columnas (véase Figura 3.5).



View	Proc	Object	Print	Name	Freeze	Default	Sort	Edit+/-	Smpi+/-	Compare+/-
			V	PE	PC	GP	R			
2009Q1			275.5	98.6	105.5	1868	895.0000			
2009Q2			285.6	96.6	104.4	2157	957.6500			
2009Q3			336.8	102.0	115.6	2541	1024.686			
2009Q4			333.4	91.9	102.7	2561	1075.920			
2010Q1			357.0	105.6	106.1	3103	1129.716			
2010Q2			325.2	102.8	100.2	2661	1186.202			
2010Q3			362.2	102.1	97.2	2757	1245.512			
2010Q4			232.0	99.1	93.3	1403	1307.787			
2011Q1			252.2	100.3	94.6	1856	1386.254			
2011Q2			322.1	104.8	109.4	2123	1455.567			
2011Q3			297.3	85.6	94.7	2181	1528.346			
2011Q4			298.9	103.0	111.4	2520	1604.763			
2012Q1			246.8	100.4	105.0	2134	1685.001			
2012Q2			322.4	93.8	118.3	2473	1786.101			
2012Q3			383.5	99.7	102.1	3125	1875.406			
2012Q4			321.8	104.6	110.6	2753	1969.176			
2013Q1			351.5	100.0	97.3	2869	2067.635			
2013Q2			381.1	103.7	91.8	3301	2171.017			
2013Q3			412.5	96.7	96.8	3043	2279.568			
2013Q4			217.5	105.5	102.2	1538	2370.750			
2014Q1			229.9	99.2	108.5	1903	2465.581			
2014Q2			342.9	86.4	98.1	2489	2564.204			
2014Q3					

Figura 3.5

EViews es capaz de realizar un conjunto de diagramas de dispersión muy útiles para el análisis. Si se quiere hacer una inspección gráfica de parejas de variables, en la ventana del Workfile seleccionamos las variables que se quiere chequear pulsando la tecla CTRL y el puntero del ratón empezando por la variable explicativa y luego la variable dependiente y se da clic. Se procede a seleccionar las variables de trabajo en el siguiente orden: PE y V, como se indica en la Figura 3.6. Posteriormente, se selecciona *Open* → *as Group* y se da clic. En la Figura 3.7 se observa que se abren las series de las variables elegidas.

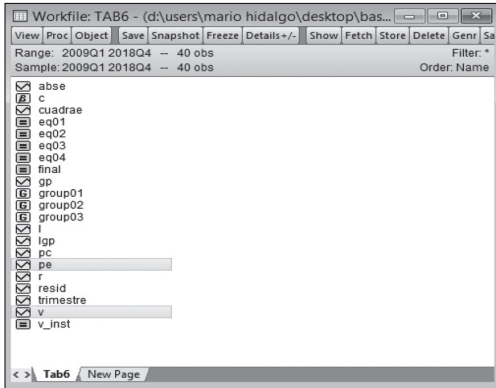


Figura 3.6

The screenshot shows the EViews Group window for 'UNTITLED' with columns for 'PE' and 'V'. The data is as follows:

	PE	V
2009Q1	98.6	275.5
2009Q2	96.6	285.6
2009Q3	102.0	336.8
2009Q4	91.9	333.4
2010Q1	105.6	357.0
2010Q2	102.8	325.2
2010Q3	102.1	362.2
2010Q4	98.1	232.0
2011Q1	100.3	252.2
2011Q2	104.8	322.1
2011Q3	85.6	297.3
2011Q4	103.0	298.9
2012Q1	100.4	246.8
2012Q2	93.9	322.4
2012Q3	99.7	383.5
2012Q4	104.6	321.8
2013Q1	100.0	351.5
2013Q2	103.7	381.1
2013Q3	96.7	412.5
2013Q4	105.5	217.5
2014Q1	99.2	229.9
2014Q2	86.4	342.9
2014Q3	88.7	311.1

Figura 3.7

Enseguida, se selecciona *View* → *Graph* → *Scatter* y en la opción *Fit lines* – *Options* se elige línea de regresión (*Regression line*) y se da OK, como se muestra en la Figura 3.8. La línea de regresión (línea de la recta de regresión) nos permite apreciar la dirección de la pendiente existente entre los pares de variables.

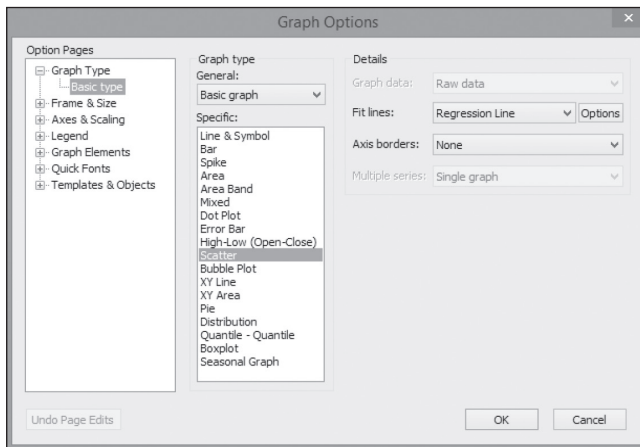


Figura 3.8

Con el propósito de conocer de manera intuitiva el tipo de relación existente entre las ventas y cada una de las tres nuevas variables (a excepción del precio de la editorial) se procede a elaborar los diagramas de dispersión respectivos; por ejemplo, la Figura 3.9 muestra que existe una relación positiva (directa)

entre las ventas y el precio de las editoriales competidoras, es decir, una variación (aumento/disminución) de los precios promedio de la competencia provocaría una variación en el mismo sentido (aumento/disminución) de las ventas de la editorial de propiedad de Don Pablo. Aunque aún es temprano para sacar conclusiones, Emilio siente que su hipótesis es correcta: la variable determinante de las ventas es el precio de la competencia. Don Pablo, por su parte, en su interior reflexiona al respecto: Aunque en el diagrama de dispersión (Figura 3.10) se observa una relación positiva entre las dos variables, esto aún esto no constituye una evidencia contundente.

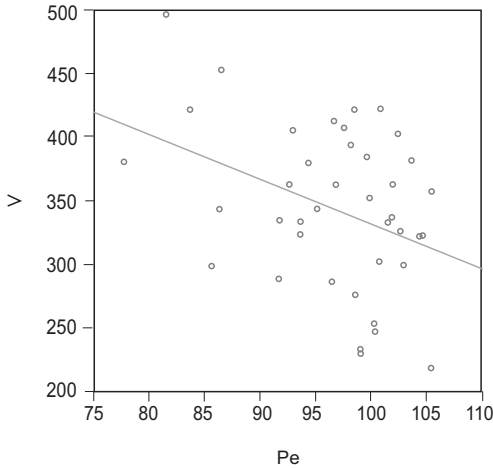


Figura 3.9

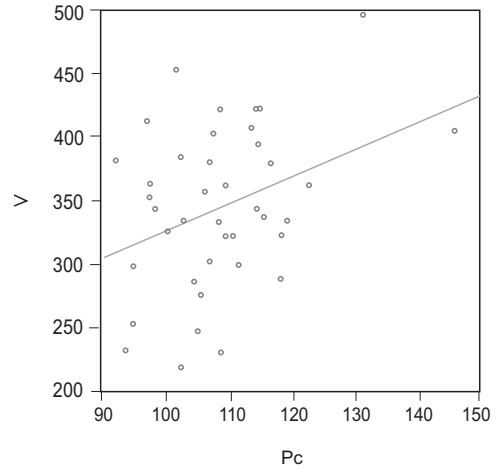


Figura 3.10

La Figura 3.9 muestra la relación negativa (inversa) existente entre el precio de la editorial y las ventas, la que sugiere que a mayor precio menor volumen de ventas y viceversa. Entre el precio de la competencia y las ventas hay una relación positiva (directa), que indica que si la competencia eleva sus precios y si la editorial mantiene constantes los suyos, sus ventas se incrementarán y viceversa (véase Figura 3.10). Entre el gasto en publicidad y las ventas existe una relación positiva (directa) bastante estrecha (los puntos se encuentran muy próximos a la recta de regresión -línea que pasa en medio de los puntos-, por ende, los residuos son más pequeños), lo cual sugiere que

un mayor gasto en publicidad impulsará un mayor volumen de ventas de libros y viceversa (véase Figura 3.11). Por último, la Figura 3.12 muestra una relación positiva (directa) entre la renta personal de los clientes y las ventas de la editorial, la cual indica que si la renta de los clientes aumenta, también lo harán las ventas y lo contrario. Nótese que la inspección gráfica efectuada es un análisis preliminar que permite recoger indicios sobre la dirección en que se mueven las variables, pero no aporta ningún dato sobre la magnitud de dicha relación.

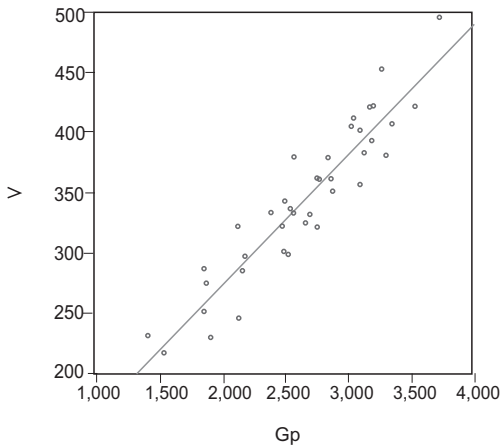


Figura 3.11

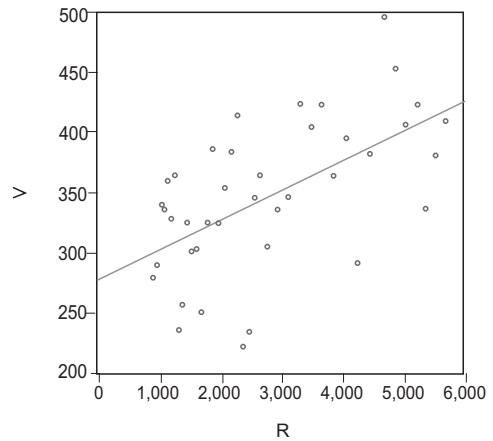


Figura 3.12

A continuación se realiza una inspección numérica apoyada en el análisis del coeficiente de correlación lineal simple (r) entre cada par de variables. Para ello se pulsa la tecla CTRL y con el puntero del ratón se selecciona las variables en el mismo orden anterior; luego se elige *Open* → *As group* y se da clic. En la ventana que se abre, se elige *View* → *Covariance Analysis* y se da clic como se muestra en la Figura 3.13. Posteriormente, en el cuadro de diálogo que se abre, se selecciona *Correlation* y se da OK como se indica en la Figura 3.14.

	PE	PC	GP	R
98.6	105.5	1868	895.0000	
96.6	104.4	2157	957.6500	
102.0	115.6	2541	1024.686	
91.9	102.7	2561	1075.920	
105.6	106.1	3103	1129.716	
102.8	100.2	2661	1186.202	
102.1	97.2	2757	1245.512	
99.1	93.3	1403	1307.787	
100.3	94.6	1856	1386.254	
104.8	109.4	2123	1455.567	
85.6	94.7	2181	1528.346	
103.0	111.4	2520	1604.763	
100.4	105.0	2134	1685.001	
93.8	118.3	2473	1786.101	
99.7	102.1	3125	1875.406	
104.6	110.6	2753	1969.176	
100.0	97.3	2869	2067.635	
103.7	91.8	3301	2171.917	
96.7	96.8	3043	2279.568	
105.5	102.2	1538	2370.750	
99.2	108.5	1903	2465.581	
96.4	98.1	2489	2564.204	

Figura 3.13

Figura 3.14

En la Tabla 3.15 se muestra los coeficientes de correlación lineal, por ejemplo, entre el precio promedio de la editorial (PE) y las ventas (V) existe una relación inversa y débil de $-0,38$, entre el precio de competencia (PC) y las ventas existe una relación directa y débil de $0,36$, en cambio entre el gasto en publicidad (GP) y las ventas existe una relación directa y fuerte de $0,93$, y por último, entre la renta personal y las ventas la relación existe una relación directa y media de $0,58$. En síntesis, mediante el coeficiente de correlación lineal se puede deducir que las ventas de la editorial tienen relación inversa con el precio de la editorial y directa con el precio de la competencia, el gasto en publicidad y con la renta personal. El gasto en publicidad y la renta personal son las variables que tienen mayor correlación lineal con las ventas; en cambio, el precio de la competencia es la variable que tiene menor correlación lineal con las ventas. Esto seguramente, debe preocuparle un tanto a Emilio, quien es el defensor de la hipótesis de que el precio de la competencia es la variable más importante en la explicación de las ventas de la editorial.

Tabla 3.3
Matriz de correlación lineal.

Variables explicativas	PE	PC	GP	R
PE	1,000000	-0,258176	-0,159481	-0,498638
PC	-0,258176	1,000000	0,285806	0,556504
GP	-0,159484	0,285806	1,000000	0,464269
R	-0,498638	0,556504	0,464269	1,000000

Las características descriptivas básicas de cada serie se obtienen a través de la selección de las variables en el orden sugerido anteriormente, con el mouse se selecciona *Open* → *as Group* y se da clic derecho. Inmediatamente se despliega una ventana que contiene las variables del modelo, se da *View* → *Descriptive Stats* → *Common Sample* (véase Figura 3.15). A continuación aparece una ventana que muestra los principales estadísticos de cada una de las variables, tales como: la media (*mean*), mediana (*median*), valor máximo (*maximun*), valor mínimo (*minimun*), desviación típica (*Std. Dev*), etcétera., tal como se muestra en la Tabla 3.4.

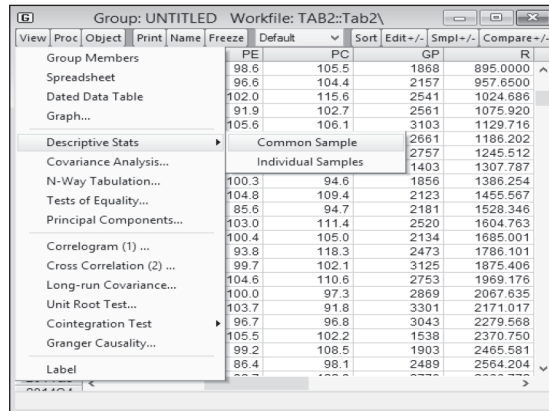


Figura 3.15

Tabla 3.4
Estadísticos descriptivos.

Estadísticos	V	PE	PC	GP	R
Mean	344,7425	96,76750	108,5175	2.646,825	2.779,105
Median	343,0500	98,60000	107,9000	2.681,000	2.418,165
Maximum	495,6000	105,6000	146,2000	3.723,000	5.712,077
Minimum	217,5000	77,80000	91,80000	1.403,000	895,0000
Std. Dev.	63,37666	6,828049	10,65847	545,6214	1.487,668
Skewness	-0,047215	-1,014618	1,125159	-0,317646	0,539739
Kurtosis	2,696249	3,394580	5,414099	2,548437	2,007456
Jarque-Bera	0,168637	7,122489	18,15301	1,012508	3,584032
Probability	0,919139	0,028403	0,000114	0,602749	0,166624
Sum	13.789,70	3.870,700	18,15301	105.873,0	111.164,2
Sum Sq. Dev.	156.647,4	1.818,268	0,000114	11.610.406	86.313.064
Observations	40	40	40	40	40

Fuente: Elaboración propia.

Aunque, Emilio no está en desacuerdo con la idea de que el precio tiene una cierta incidencia sobre las ventas, sostiene que la explicación que da su padre es simplista y que deben añadirse más elementos para poder explicar correctamente lo que ha venido sucediendo. Su afirmación central es que el mundo es más complejo de lo que su padre parece creer. La *competencia* ha sido bastante fuerte en los últimos años y los precios a los que vende la editorial son competitivos (esta es una variable que debe tenerse muy en cuenta). De la misma manera, la gente es bombardeada constantemente con publicidad, de modo que es necesario tener en cuenta también su efecto. Emilio está convencido de que el gasto en *publicidad* que ha venido realizando la editorial, a pesar de las protestas de su padre, ha sido importante para explicar las ventas obtenidas, pero en el análisis tampoco se puede dejar de lado, el ingreso o renta de los clientes de la editorial, considerando que los libros son bienes superiores, muy elásticos al ingreso de los consumidores.

Según la intuición de Emilio, el modelo de ventas de la editorial estaría conformado por una variable dependiente y cuatro variables explicativas, minuciosamente identificadas como se indica en la Tabla 3.5.

Tabla 3.5
Variables del modelo de la Editorial.

Variable Dependiente	Variables Explicativas
$V_t = \text{Ventas}$	$PE_t = \text{Precio promedio de la editorial.}$
	$PC_t = \text{Precio promedio de la competencia}$
	$GP_t = \text{Gasto en publicidad de la editorial}$
	$R_t = \text{Renta personal promedio de los clientes de la editorial}$

Fuente: Elaboración propia.

En términos generales, ahora nuestro objetivo consiste en cuantificar la relación existente entre una variable Y (variable dependiente) y un conjunto de K variables explicativas $X_1, X_2, X_3, \dots, X_K$, mediante un modelo lineal múltiple y una muestra de datos de tamaño T .

$$V_t = \beta_0 - \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_K X_{Kt} + \mu_t \quad t = 1, 2, \dots, T \quad (3.3)$$

Dónde:

$V_t =$ Observación t-ésima de la variable dependiente Y_t

$X_{Kt} =$ Observación t-ésima de la variable explicativa X_k ,

$k = 1, 2, 3, \dots, k$

$\beta_0 =$ Intercepto o término independiente.

$\beta_k =$ Coeficiente o parámetro de la variable explicativa X_k , que mide el impacto de cada variable explicativa en el comportamiento de la variable dependiente.

$\mu_t =$ t-ésimo valor del término de perturbación o de error.

El término de perturbación (μ_t) es una variable aleatoria (no observable) que recoge el efecto de variables no incluidas en el modelo (el número de competidores, los gustos o preferencias, el sexo de los compradores, la religión, entre otras), el comportamiento aleatorio de los agentes económicos, posibles errores de medida, etcétera.

Entonces, el modelo de ventas de la editorial quedaría especificado de la siguiente manera:

$$Ventas = \beta_0 + \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t + u_t \quad t = 1, 2, \dots, T \quad (3.4)$$

3.2 Estimación del modelo.

La estimación por MCO del modelo de ventas de la editorial quedaría, así:

$$\widehat{Ventas} = \hat{\beta}_0 - \hat{\beta}_1 PE_t + \hat{\beta}_2 PC_t + \hat{\beta}_3 GP_t + \hat{\beta}_4 R_t + \hat{\mu}_t \quad t = 1, 2, \dots, T \quad (3.5)$$

El lector recordará que el gorro o circunflejo colocado encima de cada parámetro incluido el término independiente (intercepto) y la perturbación significa valores estimados.

Dado que el modelo de regresión lineal múltiple es más complejo, su estimación se hará con el apoyo de EViews. En la estimación del modelo de regresión lineal múltiple se seleccionan las variables del modelo en el siguiente orden: V, PE, PC, GP y R. Posteriormente, se elige *Open* → *as Equation* y damos clic. En el cuadro de diálogo que se abre (véase Figura 3.16) de forma manual reordenamos las variables en el siguiente orden: V, ¹⁵ PE, PC, GP y R. En la pestaña de método (*Method*), EViews por defecto, elige mínimos cuadrados ordinarios (*LS – Least Squares – NLS and ARIMA*) y se da *Aceptar*.

¹⁵ También llamado término independiente, ordenada en el origen, intercepto o constante.

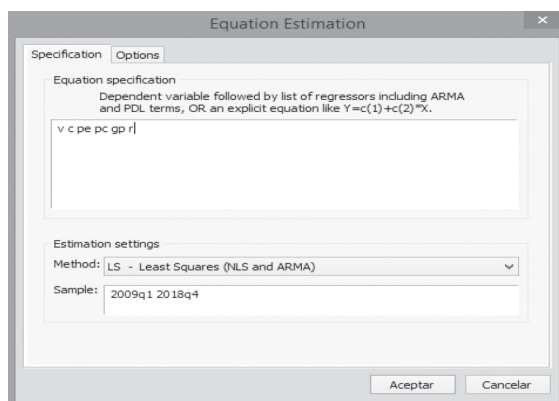


Figura 3.16

A partir de esta rutina, EViews proporciona la información de la regresión múltiple que se muestra en la Tabla 3.6. La salida de la información se distribuye en columnas, tales como: los valores de los coeficientes estimados, sus errores estándar, los valores t , la probabilidad de los valores t y muchos estadísticos más, que enriquecen el análisis econométrico. Infortunadamente, EViews no visualiza el intervalo de confianza al 95% de cada coeficiente de regresión, como si lo hacen otros paquetes como Stata.

Tabla 3.6
Estimación del modelo por MCO.

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	243,2046	59,81264	4,066106	0,0003
PE	-1,913099	0,492254	-3,886406	0,0004
PC	0,126788	0,327776	0,386813	0,7012
GP	0,100304	0,006031	16,63173	0,0000
R	0,002669	0,002832	0,942452	0,3524
R-squared	0,926726	Mean dependent var		344,7425
Adjusted R-squared	0,918351	S.D. dependent var		63,37666
S.E. of regression	18,10938	Akaike info criterion		8,747205
Sum squared resid	11478,23	Schwarz criterion		8,958315
Log likelihood	-169,9441	Hannan-Quinn criterion		8,823536
F-statistic	110,6643	Durbin-Watson stat		1,670121
Prob(F-statistic)	0,000000			

Como EViews no visualiza el intervalo de confianza al 95% de los coeficientes de regresión, se procede a calcularlos manualmente. La tercera columna (*Std. Error*) de la Tabla 3.6 recoge la desviación típica estimada de los parámetros del modelo y puede utilizarse para el cálculo de intervalos de confianza de los parámetros estimados a un nivel de confianza α igual a $[\text{Coefficient} \pm (t_{\alpha/2, T-K-1} \times \text{Std. Error})]$. Como el valor de la t de Student $t_{0,025} = 2,03011$ para 35 grados de libertad ($T-K-1 \rightarrow 40 - 4 - 1 = 35$). Un intervalo de confianza nos va a permitir calcular dos valores alrededor de una media muestral (uno superior y otro inferior). Estos valores van a acotar un rango dentro del cual, con una probabilidad del 95% se van a localizar los parámetros poblacionales. El nivel de confianza y la amplitud del intervalo varían conjuntamente, de forma que un intervalo más amplio tendrá más posibilidades de acierto (mayor nivel de confianza), mientras que para un intervalo más pequeño, que ofrece una estimación más precisa, aumentan sus posibilidades de error.

La probabilidad de que el verdadero valor del parámetro se encuentre en el intervalo construido se denomina nivel de confianza (95%) y la probabilidad de equivocarnos se llama nivel de significancia (5%). Los intervalos de confianza permiten verificar hipótesis planteadas respecto a parámetros poblacionales, como se verá más adelante.

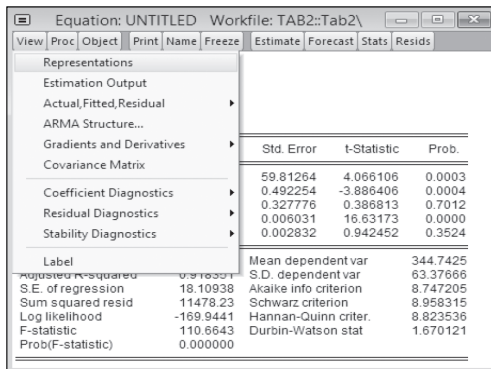
La Tabla 3.7 muestra que los coeficientes de las variables precio de la competencia (PC) y renta personal (R) se encuentran fuera del intervalo de confianza con un nivel de confianza del 95%; en cambio, los coeficientes del precio de la editorial (PE) y el gasto en publicidad (GP) se encuentran dentro de los intervalos de confianza respectivos con un 95% de confianza.

Tabla 3.7
Intervalos de confianza.

Variable	Coficiente	$t_{\alpha/2, T-K-1}$	Error Estándar	Intervalo de confianza al 95%	
C	243,2046	2,03011	59,81264	121,77836	364,63084
PE	-1,913099	2,03011	0,492254	-2,91243	-0,91377
PC	0,126788	2,03011	0,327776	-0,53863	0,79221
GP	0,100304	2,03011	0,006031	0,08806	0,11255
R	0,002669	2,03011	0,002832	-0,00308	0,00842

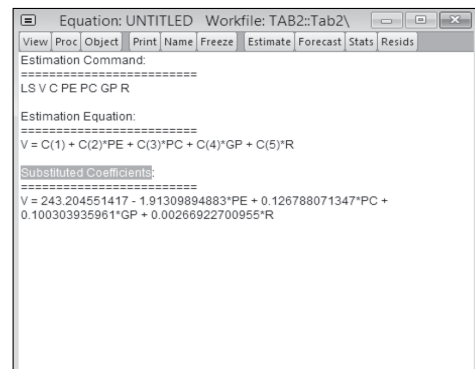
Fuente: Elaboración propia

Posteriormente, con el puntero del ratón se selecciona *View* → *Representations* que muestra la especificación del modelo, así como el modelo finalmente estimado (véase Figura 3.17) y se da clic. En el campo *Substituted Coefficients* aparece la ecuación del modelo estimado, donde cada coeficiente está acompañado de una apreciable cantidad de decimales como sinónimo de mayor precisión en la estimación, tal como se indica en la Figura 3.18.



	Std. Error	t-Statistic	Prob.
59.81264	4.066106	0.0003	
0.492254	-3.886406	0.0004	
0.327776	0.386813	0.7012	
0.006031	16.63173	0.0000	
0.002832	0.942452	0.3524	

Figura 3.17



Estimation Command:
=====

LS V C PE PC GP R

Estimation Equation:
=====

$$V = C(1) + C(2)*PE + C(3)*PC + C(4)*GP + C(5)*R$$

Substituted Coefficients:
=====

$$V = 243.204551417 - 1.91309894883*PE + 0.126788071347*PC + 0.100303935961*GP + 0.00266922700955*R$$

Figura 3.18

La estimación del modelo de ventas de la editorial proporciona el siguiente resultado:

$$\widehat{Ventas} = 243,20 - 1,91PE + 0,127PC + 0,100GP + 0,003R \quad (3.6)$$

$$(59,8) \quad (0,49) \quad (0,33) \quad (0,006) \quad (0,003)$$

$$R^2 = 92,67\%; \quad \bar{R}^2 = 91,84\%$$

Cabe recordar, que los valores entre paréntesis ubicados debajo de cada coeficiente corresponden a la desviación típica o error estándar (*Std. Error*).

Los signos del modelo son coherentes con la teoría económica, así pues, el precio de la editorial está inversamente relacionado con las ventas; en cambio, existe una relación directa entre el precio de la competencia, el gasto en publicidad y la renta personal con las ventas como es apenas lógico.

En el modelo de regresión simple del capítulo 2 se calculó el coeficiente de determinación (R^2) y el coeficiente de determinación ajustado (\bar{R}^2) para medir la bondad de ajuste del modelo, cuyos valores arrojados fueron de 14,48% y 12,23%, respectivamente. En este nuevo modelo, el coeficiente de determinación se ha elevado considerablemente; por ejemplo, si se interpreta el R^2 se dice que el 92,7% de la variabilidad de las ventas en su media son explicadas por PE, PC, GP y R; pero aún queda un faltante de 7,3% ($100\% - 92,7\% = 7,3\%$) de las ventas de la editorial sin ninguna explicación hasta el momento, lo cual sugeriría que podrían incluirse nuevas variables en el modelo, pero esto le restaría simplicidad y utilidad¹⁶.

El signo de los coeficientes confirma que existe una relación negativa (inversa) entre el precio promedio de la editorial y las ventas de la misma, así

¹⁶ Un modelo nunca puede llegar a ser una descripción completamente precisa de la realidad; para describir la realidad, se tendría que desarrollar un modelo tan complejo e incomprensible que éste sería de poca utilidad práctica. En cualquier construcción de modelos es inevitable hacer abstracción o simplificación en alguna medida. El principio de parsimonia, establece que un modelo se debe conservar tan simple como sea posible, esto significa que se deben introducir en el modelo unas pocas variables claves que capturen la esencia del fenómeno bajo estudio relegando toda influencia menor y aleatoria al término de error (μ_i).

como una relación positiva (directa) entre el precio de la competencia, el gasto en publicidad y la renta personal promedio de los clientes con respecto a las ventas. En fin, los signos de los coeficientes son coherentes con los postulados de la teoría económica.

3.3 Estadísticos de la estimación.

EViews al estimar un modelo por MCO proporciona una salida de información muy rica que se muestra en la Tabla 3.5. Aquí se encuentra la estimación de los coeficientes asociados a cada una de las variables explicativas junto con su error estándar o desviación típica estimada, el estadístico t de significación individual y la probabilidad del mismo. En la parte inferior se encuentra un conjunto de estadísticos que permiten evaluar parcialmente la regresión realizada. Recogiendo la exposición de Carrascal et al, (2004), se hace una breve explicación de los principales elementos de la salida de Eviews.

La columna denominada Coeficientes (*Coefficient*) recoge el valor de los estimadores de los parámetros asociados a cada una de las variables explicativas, los cuales se obtienen a partir de la siguiente expresión:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

Si se cumplen las hipótesis clásicas señaladas anteriormente, estos estimadores serán lineales, insesgados, óptimos (ELIO) y consistentes (MELI); por lo tanto, el modelo tendrá mayor capacidad explicativa y predictiva; es decir, mayor poder para explicar de qué dependen las ventas de la editorial, así como para pronosticar las ventas a partir de valores dados de precios, gasto en publicidad y renta personal.

Cada uno de estos coeficientes recoge el cambio que experimenta la variable dependiente ante un cambio unitario de la variable explicativa correspondiente, suponiendo que el resto de las variables permanecen

constantes¹⁷. Así, por ejemplo, los resultados muestran que si el precio promedio de la editorial aumenta en una unidad monetaria, permaneciendo el resto de las variables constantes o fijas (manteniendo constantes el precio de la competencia, el gasto en publicidad y la renta personal), las ventas de la editorial disminuirán en 1,91 unidades monetarias. Mantener el resto de las variables constantes permite aislar el efecto o cambio de cada variable explicativa sobre la variable dependiente. Si se prescinde de esta condición, el análisis de la contribución de cada variable explicativa sobre la variable dependiente se vuelve prácticamente imposible.

La columna de los errores estándar (*Std. Error*) recoge la desviación típica estimada de los estimadores y mide –siempre que los estimadores sean insesgados–, la precisión con la que son estimados los parámetros. También significan el grado de confianza que podemos depositar en nuestras estimaciones.

La matriz de varianzas y covarianzas estimada de los estimadores es $S\hat{\beta}\hat{\beta}' = S^2 (X'X)^{-1}$, donde $S^2 = e'e/T - (K+1) = \sum_{t=1}^T e_t^2 / T - K - 1$ es un estimador insesgado y consistente, bajo las hipótesis clásicas de la varianza de las perturbaciones (σ_ε^2). De esta forma, los errores estándar se obtienen como la raíz cuadrada de los elementos de la diagonal principal de dicha matriz.

El estadístico *t* (*t-statistic*) que se calcula como el cociente entre el estimador y su error estándar $\hat{\beta}_i/\hat{\sigma}_i$ ($\hat{\beta}_i$ para el caso de la variable PE = $\frac{-1,913099}{0,492254} = 0,0004$), permite contrastar la hipótesis de que el coeficiente es igual a cero ($H_0: \beta_i = 0$ frente a $H_1: \beta_i \neq 0$) y que, por tanto, la variable en cuestión no es individualmente significativa para explicar el comportamiento de la variable endógena.

¹⁷ Esto supone que las demás variables explicativas toman el valor de cero (0), que al ser multiplicado por el coeficiente específico de cada variable, el producto es igual a 0, condición que permite analizar por separado el efecto o cambio que ejerce cada variable explicativa sobre la variable dependiente.

La columna de probabilidad (*Prob*) proporciona dos veces el área que el valor absoluto del estadístico *t* deja a su derecha e indica la probabilidad de cometer el error de rechazar la hipótesis nula siendo cierta (error de tipo I). Estos valores están calculados a partir de la distribución *t* de Student con *T-K-1* grados de libertad, siendo *K+1* el número de coeficientes de regresión incluyendo el término constante.

El coeficiente de determinación (*R-squared*) o R^2 , es una medida estadística que sirve para valorar el éxito de la regresión para predecir los valores de la variable dependiente en el periodo muestral y se define como la parte de la varianza de la variable dependiente explicada por las variables independientes.

$$R^2 = 1 - \frac{SCE}{SCT} = \frac{SCR}{SCT}$$

$$\text{Dónde: } SCE = \sum_{t=1}^T e_t^2 ; SCR = \sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 \text{ y } SCT = \sum_{t=1}^T (Y_t - \bar{Y})^2$$

En el modelo de las ventas de la editorial, la capacidad explicativa conjunta de las variables explicativas PE, PC, GP y R es elevada, pues explican el 92,7% de la variabilidad de la variable dependiente ventas (V). Hay que tener bastante cuidado, sin embargo, con la interpretación del coeficiente de determinación. En ocasiones, si la muestra consta de pocas observaciones, quizás, uno o dos residuos elevados pueden generar un coeficiente de determinación reducido y, por ello, conducir a creer que la regresión estimada es mala, cuando excepto por dichas observaciones, el ajuste puede ser excelente. Por otra parte, si la muestra consta de muy pocas observaciones, y ningún residuo es especialmente alto, se tendrá un coeficiente de determinación muy elevado, sin que deba interpretarse como un excelente ajuste, sino más bien como un indicador de escasa información muestral. Otro caso delicado se refiere al uso del coeficiente de determinación con muestras de series temporales que muestran una tendencia similar. En tales casos, el coeficiente de determinación se aproxima a la unidad, aunque la relación entre ambas variables, excepción hecha de sus tendencias, pueda ser pobre.

El coeficiente de determinación ajustado (*adjusted R-squared*) o \bar{R}^2 se obtiene a partir del R^2 ponderado por los grados de libertad.

$$\bar{R}^2 = 1 - \frac{\frac{SCE}{T-K-1}}{\frac{SCT}{T-1}} = 1 - R^2 \frac{T-1}{T-K-1} \quad (3.7)$$

La ventaja que presenta este coeficiente frente al R^2 es que permite comparar la capacidad explicativa de modelos referidos a una misma muestra de la misma variable dependiente con distinto número de variables explicativas, por consiguiente, se elegirá el modelo con mayor \bar{R}^2 . En el modelo de las ventas de la editorial, la capacidad explicativa conjunta de las variables exógenas PE, PC, GP y R es elevada, pues explican el 91,8% de la variabilidad de la variable endógena Ventas (V) en su media.

El error estándar de la regresión (*S.E. of regression*) es otra medida que sirve para analizar la capacidad explicativa del modelo, pues está recogiendo la función objetivo ponderada por sus grados de libertad. Bajo las hipótesis clásicas, el cuadrado de este coeficiente es una estimación insesgada y consistente de la varianza de las perturbaciones.

$$S = \sqrt{S^2} = \sqrt{e'e/(T - K + 1)} = \sqrt{\hat{\sigma}_\varepsilon^2}$$

La suma de los errores al cuadrado (*Sum squared resid*) es el valor de la función objetivo en el mínimo, cuando estimamos por MCO, es decir, cuando obtenemos los estimadores de modo que las diferencias entre el valor observado de Y y el valor estimado de \hat{Y} sea el más pequeño posible.

El logaritmo de la función de verosimilitud (*Log likelihood*) es el valor de la función objetivo en el máximo cuando estimamos por máxima verosimilitud¹⁸, es decir, cuando obtenemos los estimadores de los parámetros

¹⁸ La idea fundamental de este método es tomar como estimación del parámetro estudiado el valor que haga máxima la probabilidad de obtener la muestra observada. El estimador de máxima verosimilitud es el valor de que maximiza $L(\Phi)$.

que maximizan la probabilidad de la muestra (los más verosímiles dada la muestra disponible).

El estadístico de Durbin-Watson (*Durbin-Watson stat*) sirve para contrastar la hipótesis de incorrelación entre las perturbaciones aleatorias frente a la presencia de autocorrelación según un esquema AR(1). Su utilización se verá en el capítulo 10 dedicado a la presencia de autocorrelación.

$$DW = \sum_{t=2}^T (e_t - e_{t-1})^2 / \sum_{t=1}^T e_t^2$$

Las características básicas de la variable dependiente vienen recogidas mediante su media (*Mean dependent var*) y su cuasi desviación típica muestral (*S.D. dependent var*).

El criterio de información de Akaike (*Akaike info criterion*) y el criterio de Schwarz (*Schwarz criterion*), son dos estadísticos que sirven para analizar la capacidad explicativa de un modelo y permiten realizar comparaciones a este respecto entre modelos anidados. La definición de cada uno de ellos es:

$$AIC = -2L/T + 2(K+1)/T; SC = -2L/T + (K+1)(\log T)/T$$

Donde L es el logaritmo de la verosimilitud calculado en los valores obtenidos para los estimadores y cuya expresión viene dado por:

$$L = -T/2 (1 + \log 2\pi + \log(e'e/T))$$

El estadístico F (*F-statistic*) es el estadístico que se construye para contrastar si los parámetros asociados a las variables explicativas del modelo (exceptuando el término independiente) son conjuntamente iguales a cero, es decir, este estadístico permite contrastar la capacidad explicativa conjunta de las variables introducidas en el modelo.

$$\mathcal{F} = \frac{\frac{SCR}{K}}{\frac{SCE}{T-k-1}} = \frac{\frac{R^2}{K}}{\frac{1-R^2}{T-K-1}} \rightarrow F_{T-K-1}^K$$

Dado que el estadístico puede expresarse en términos del R^2 , el contraste puede considerarse como un modo de determinar si el coeficiente de determinación del modelo es suficientemente elevado, estadísticamente como para considerar que la capacidad explicativa del modelo es adecuada.

La probabilidad del estadístico \mathcal{F} (*Prob F-statistic*). Este valor mide la probabilidad de cometer el error de tipo I, la probabilidad de rechazar la hipótesis nula siendo cierta. Para calcularla se parte de una distribución \mathcal{F} de Snedecor con K grados de libertad en el numerador y T-K-1 en el denominador.

La estimación de la matriz de varianzas y covarianzas (*Covariance matrix*) de los estimadores de los coeficientes del modelo se realiza a partir de la raíz cuadrada de los elementos de la diagonal principal. De aquí se derivan las desviaciones típicas (*Std. Error*) de los estimadores (*Coefficient*) que aparecen en los resultados generales de la estimación. La matriz de covarianza de los coeficientes (*Coefficiente covariance matrix*) se obtiene seleccionando *View* → *Covariance Matrix* (véase Figura 3.19) y damos clic. Para comprobar que las desviaciones típicas de los coeficientes se derivan de la raíz cuadrada de los elementos de la diagonal principal (véase Tabla 3.8), tomaremos el valor de 0,242314 que corresponde a la variable precio medio de la editorial (PE), cuya raíz cuadrada es 0,492254. Efectivamente, este valor corresponde a la desviación típica de la variable PE.

	Std. Error	t-Statistic	Prob.
59.81264	4.066106	0.0003	
0.492254	-3.886406	0.0004	
0.327776	0.386813	0.7012	
0.006031	16.63173	0.0000	
0.002832	0.942452	0.3524	

Label	Value	Value	
Adjusted R-squared	0.970307	Mean dependent var	344.7425
S.E. of regression	18.10938	S.D. dependent var	63.37666
Sum squared resid	11478.23	Akaike info criterion	8.747205
Log likelihood	-169.9441	Schwarz criterion	8.958315
F-statistic	110.6643	Hannan-Quinn criter.	8.823536
Prob(F-statistic)	0.000000	Durbin-Watson stat	1.670121

Figura 3.19

Tabla 3.8.
Matriz de coeficientes de Covarianza.

	C	PE	PC	GP	R
C	3,577,552	-24,02121	-9,928135	-0,043883	-0,018478
PE	-24,02121	0,242314	-0,003783	-0,000276	0,000062
PC	-9,928135	-0,003783	0,107437	-0,000069	-0,000425
GP	-0,043883	-0,000276	-0,000069	0,000036	-0,000007
R	-0,018478	0,000617	-0,000425	-0,000007	0,000008

Fuente: Elaboración propia.

RETO 2:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

**ENCARANDO EL MODELO DE
REGRESIÓN LINEAL GENERAL:
RESTRICCIONES Y CONTRASTES**

La discusión entre padre e hijo puede empezar a dilucidarse ya. Si Emilio tiene razón, el valor que los coeficientes estimados toman, no tendrá importancia y si la tendrá en cambio, el saber si puede aceptarse respecto de ellos la hipótesis nula de que cada uno de ellos (o todos conjuntamente, o dos de ellos a la vez) toman el valor cero (0). Si alguna de estas hipótesis se aceptará, la variable que acompaña a ese coeficiente no resultará de interés para explicar las ventas de la editorial. Por lo tanto, llevar a cabo los contrastes adecuados permitirá acabar con esta larga discusión.

El contraste de hipótesis es un tema un tanto confuso que pocas veces se puede comprender tan solo con un repaso de los conceptos. Su entendimiento requiere el uso de ejemplos prácticos y una claridad en las definiciones. Por esta razón, incluiremos un par de conceptos importantes respecto a las hipótesis.

Una hipótesis estadística es un supuesto, concerniente a los parámetros o a la forma de la distribución de probabilidad, correspondiente a una o más poblaciones dadas; en palabras más simples, una hipótesis es un enunciado acerca de un valor estadístico (parámetro) poblacional. La hipótesis se debe formular en forma correcta o lógica y debe ser enunciada antes de obtener los datos muestrales.

En el contraste de hipótesis es común encontrar dos términos singulares: *hipótesis nula* e *hipótesis alternativa*. La hipótesis nula es aquella por medio de la cual se hace una afirmación sobre un parámetro a contrastar con el resultado muestral. De otro lado, la hipótesis alternativa como su nombre lo indica, ofrece una alternativa a la hipótesis nula, afirmando que la hipótesis nula es falsa. En el caso de las ventas de la editorial, a lo largo de este capítulo, se plantearán varias hipótesis; como por ejemplo, Don Pablo cree tener la razón de que el precio es la variable determinante en la explicación de las ventas de la editorial, en cambio, Emilio sospecha que el precio de los libros NO es la única variable determinante de las ventas de la editorial. En otras palabras, la hipótesis nula es la contrapartida de la hipótesis de investigación.

Sampiere et al, (2010) argumenta que “las hipótesis nulas son, en cierto modo, el reverso de las hipótesis de investigación. También constituyen proposiciones acerca de la relación entre variables, sólo que sirven para refutar o negar lo que afirma la hipótesis de investigación” (p.104). Si la hipótesis de investigación de Don Pablo propone que el precio de los libros es la única variable que explica las ventas de la editorial, entonces la hipótesis nula de Emilio propone que el precio NO es la única variable explicativa de las ventas, por ende, la discusión entre padre e hijo en términos de hipótesis quedaría planteada, así:

Don Pablo → Hipótesis nula ($H_0: \beta_1 = 0$): El precio es la única variable que explica las ventas de la editorial.

Emilio → Hipótesis alternativa ($H_1: \beta_1 \neq 0$): El precio NO es la única variable explicativa de las ventas.

El contraste de hipótesis está pensado y creado para someter a prueba este tipo de enunciados para comprobar su exactitud y validez.

4.1 Contraste de significancia individual Modelo 1: Regresión simple.

Alonso et al., (2005:4) afirman que el estadístico t (t -Statistic) de interés para contrastar si una variable explicativa es relevante a la hora de explicar el comportamiento de la variable dependiente es igual al estimador $\hat{\beta}_k$ dividido por el estimador $s_{\hat{\beta}_k}$ de su desviación típica, cociente que, bajo la hipótesis nula, sigue una distribución t de student con $T-K-1$ grados de libertad. Cuando, dada una muestra, se sustituyen dichos estimadores por las correspondientes estimaciones, se obtiene un valor numérico del estadístico. La regla de decisión se adopta entonces en los siguientes términos: si β_k realmente fuera cero, su estimación MCO no debería estar muy alejada de dicho valor y el valor del estadístico t debería ser muy pequeño. Nótese que el coeficiente teórico de β_k puede tomar valores positivos o negativos; ello significa que las desviaciones del estadístico t de cero pueden también tomar valores positivos o negativos. Por lo tanto, una vez elegido el nivel de significación α que determina la región crítica de rechazo de la hipótesis, si el valor absoluto de t supera el valor crítico de $t_{\alpha/2}(T - k - 1)$ extraído de las tablas de la distribución de t Student, entonces se rechaza la hipótesis nula $\mathcal{H}_0: \beta_k = 0$ al nivel de significación elegido (véase Figura 4.1). Obsérvese que la región crítica se divide en dos partes iguales puesto que se trata de discriminar valores próximos a cero de valores alejados de cero tanto positivos como negativos. Debido a que la distribución de referencia es simétrica, el valor absoluto de t que define las dos partes de la región crítica es el mismo. Por ello, el contraste individual a dos colas al nivel de significación de $100 \alpha\%$, el procedimiento a seguir se resume en:

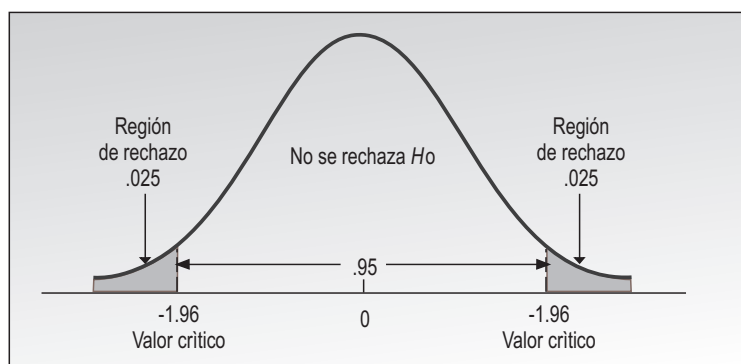


Figura 4.1. Distribución t: nivel de significación α (dos colas)

Fuente: Elaboración propia

Las zonas situadas a lado y lado de la curva rellenas de color negro representan las regiones críticas, mientras que la región no sombreada es la zona de aceptación. Si la hipótesis nula (\mathcal{H}_0) se ubica dentro del intervalo se acepta esta hipótesis, en caso contrario, se acepta la hipótesis alternativa (\mathcal{H}_1).

$$|t| = \left| \frac{\hat{\beta}_k}{S_{\hat{\beta}_k}} \right| > t_{\alpha/2}(T - k - 1) = \text{Rechazar } \mathcal{H}_0$$

Cuando se rechaza la hipótesis nula (\mathcal{H}_0) concluimos que el parámetro es significativamente diferente de cero¹⁹, es decir, la variable explicativa X asociada a dicho parámetro es significativa o relevante a la hora de explicar la

19 Si el valor del coeficiente o parámetro que acompaña a la variable explicativa, por ejemplo, en nuestro modelo estimado, si el coeficiente -1,91 que acompaña a la variable Precio de la editorial (-1,91PE), fuese cero, el producto (0xPE) también sería cero, lo que nos llevaría a concluir que la variable explicativa asociada a dicho parámetro no es relevante a la hora de explicar la variable dependiente. Por esta razón, se requiere que el parámetro sea significativamente diferente de cero. La quinta columna (Probabilidad) de la Tabla 4.1 corresponde al valor-p en las pruebas t. El valor-p es el nivel de significancia de la prueba cuando se usa el estadístico de prueba. Como el valor-p es una probabilidad, su valor es siempre un número desde cero hasta uno. Para calcular el valor-p se necesitan tablas estadísticas de la distribución t extremadamente detalladas –lo que no es muy práctico– (Wooldridge, 2015:133-134). Los programas econométricos como EViews calculan áreas bajo la función de densidad de probabilidad de la distribución t. El valor-p es $P(|T| > |t|)$ donde: T es una variable aleatoria con distribución t y n-k-1 grados de libertad y t es el valor numérico del estadístico de prueba.

variable dependiente Y . En caso contrario, la variable explicativa no es significativa o relevante para explicar Y .

El lector recordará que la discusión inicial entre Don Pablo y su hijo Emilio giraba en torno al precio de los libros como la variable fundamental que explicaba las ventas de la editorial. El contraste de significancia individual será muy útil para zanjar, al menos parcialmente, esta discusión.

La pregunta es si la variable precio (P) incluida en dicho modelo, tal como lo afirma Don Pablo es capaz por sí sola, de explicar las ventas de la editorial. La respuesta puede encontrarse al realizar un contraste de significación individual sobre el coeficiente de β_1 , considerando que el *t estadístico* de la variable P es -2,54 (véase Tabla 4.1) con un nivel de significación de $\alpha = 5\%$ y $T - K - 1 \rightarrow (40 - 1 - 1 = 38)$ grados de libertad.

Tabla 4.1

Modelo 1 - Ventas de la editorial.

Variable	Coefficiente	Error estándar	t-Estadístico	Probabilidad
C	686,5686	135,066	5,083209	0,0000
P	-3,532447	1,392403	-2,536944	0,0154
F- statistic		6,436083		
Prob(F-statistic)		0,015410		

Fuente: Elaboración propia a partir de la salida de Eviews.

Vamos a comprobar si la variable P es individualmente significativa para explicar las ventas de la editorial. Para ello, utilizaremos el resultado del modelo (2.9) junto con la regla de decisión (4.1) para realizar el contraste individual del coeficiente β_1 .

Modelo estimado:

$$\widehat{Ventas} = 686,57 - 3,53P$$

$$(135,07) (1,39)$$

$$R^2 = 14,48\% \quad \bar{R}^2 = 12,23\%$$

El lector recordará que los valores entre paréntesis corresponden al error estándar (*Std. Error*) o desviación típica de los coeficientes (columna 3 de la Tabla 4.1), los cuales son útiles para el contraste de hipótesis.

Ahora procedamos a contrastar las hipótesis de Don Pablo:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El precio <i>NO</i> es una variable que explica las ventas de la editorial.		\mathcal{H}_α : El precio es una variable que explica las ventas de la editorial.
$\mathcal{H}_0: \beta_1 = 0$		$\mathcal{H}_\alpha: \beta_1 \neq 0$

Para ello, utilizaremos la siguiente regla de decisión para comparar las hipótesis:

$$|t_i| = \left| \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \right| \begin{cases} > t_{0,025}(38) \Rightarrow \text{Rechazar } H_0 \\ \leq t_{0,025}(38) \Rightarrow \text{No rechazar } H_0 \end{cases} \quad (4.1)$$

Ahora, tomamos el t estadístico (columna 4 de la Tabla 4.1) resultante del cociente entre el coeficiente del precio medio (-3,53) y el error estándar del precio medio (1,39). El resultado 2,54 (columna 4 – Estadístico t) se compara con el valor 2,024 extraído de la tabla estadística t de Student con 38 grados de libertad.

$$|t_i| = \left| \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} \right| = \left| \frac{3,53}{1,39} \right| = 2,54 > t_{0,025}(38) = 2,024$$

Con el valor de $t_{\alpha/2}(T-K-1)$, dado que $T-K-1 = 40 - 1 - 1 = 38$, con un nivel de significancia de $\alpha = 5\%$, la tabla de la distribución t de Student proporciona un valor de $t_{0,025}(38) = 2,024$, por lo que al comparar el valor obtenido para $|t|$ con el valor anterior, encontramos que $2,54 > 2,024$, lo cual nos lleva a rechazar la hipótesis nula de que $\beta_1 = 0$. En conclusión, *el precio de los libros es una variable significativa al 5%*. Claro está, que esto no supone que el precio sea la única variable determinante a la hora de explicar las ventas, tal como lo afirma Don Pablo.

Realicemos ahora el contraste del parámetro Precio con la ayuda de EViews. Para ello utilizaremos el test de Wald de la siguiente manera: en la ventana del objeto ecuación seleccionamos *View* → *Coefficient Diagnostics* → *Wald Test – Coefficient Restrictions* (véase Figura 4.1) y en el cuadro de diálogo *Wald Test* escribimos la restricción $(C(1)=0)$ del parámetro o coeficiente asociado a la variable P, como se indica en la Figura 4.2 y damos OK.

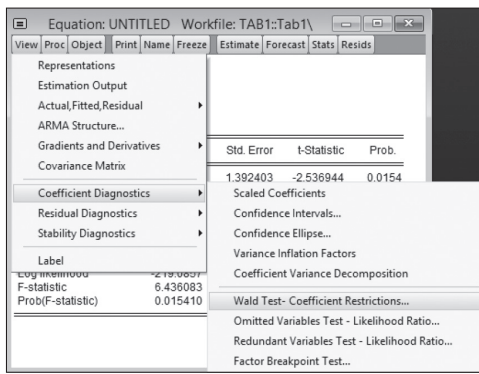


Figura 4.1

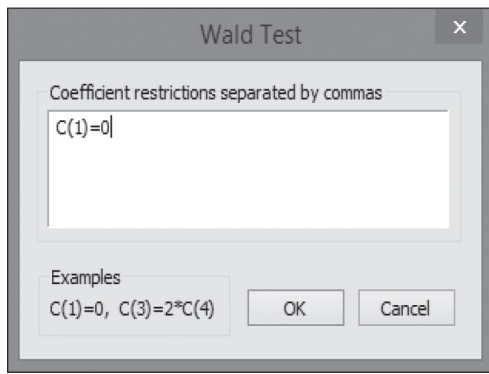


Figura 4.2

El resultado obtenido en la Tabla 4.2 arroja una probabilidad (*Probability*)²⁰ con el t estadístico (*t-statistic*) muy baja de 0,0000 ($0,0000 < 0,05$), por consiguiente, rechazamos la hipótesis nula, lo cual significa, que el precio de la editorial es una variable relevante en la explicación de las ventas.

²⁰ Indica la probabilidad de rechazar la hipótesis nula, siendo cierta, que tiene asociado el valor muestral del estadístico.

Tabla 4.2
Prueba de Wald.

Test Statistic	Value	df	Probability
t-statistic	5,083209	38	0,0000
F-statistic	25,83902	(1, 38)	0,0000
Chi-square	25,83902	1	0,0000
Null Hypothesis: C(1)=0			
Null Hypothesis Summary:			
Normalized Restriction (=0)		Value	Std. Error
C(1) → P		686,5686	135,0660

Restrictions are linear in coefficients.

4.2 Contraste de significancia individual Modelo 2: Regresión múltiple.

Debemos preguntarnos si cada una de las variables incluidas en el modelo tiene por sí la misma capacidad para explicar las ventas de la editorial, cuando las otras variables forman parte del modelo. Esto supone llevar a cabo un contraste de significación individual sobre cada uno de los coeficientes de la pendiente (β_1 , β_2 , β_3 y β_4). Como se puede apreciar en los datos de la Tabla 4.3, los valores del estadístico t para los coeficientes de las cuatro (4) variables incluidas en el modelo son, respectivamente: -3,89, 0,39 y 16,63 y 0,94. Por lo tanto, y dado que para llevar a cabo un contraste de significación individual sobre uno cualquiera de los coeficientes en este modelo, eligiendo el nivel de significación de $\alpha = 5\%$, el procedimiento a seguir, es el siguiente:

Tabla 4.3
Modelo 2 - Ventas de la editorial.

Variable	Coficiente	Error estándar	t-Estadístico	Probabilidad
C	243,2046	59,81264	4,066106	0,0003
PE	-1,913099	0,492254	-3,886408	0,0004
PC	0,126788	0,327776	16,63173	0,7012
GP	0,100304	0,006031	16,63173	0,0000
R	0,002669	0,002832	0,942452	0,3524
F- statistic		110,6643		
Prob(F-statistic)		0,000000		

Fuente: Elaboración propia a partir de la salida de Eviews.

$$\mathcal{H}_0: \beta_i = 0 \text{ Frente a } \mathcal{H}_a: \beta_i \neq 0$$

$$|t_i| = \left| \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}} \right| \begin{cases} > t_{0,025}(35) \Rightarrow \text{Rechazar } H_0 \\ \leq t_{0,025}(35) \Rightarrow \text{Aceptar } H_0 \end{cases} \quad (4.2)$$

El lector curioso se preguntará de dónde salen los valores 0,025 y 35. El valor 0,025 proviene del cociente $\frac{0,05}{2} = 0,025$ (distribución t de Student a dos colas) y el valor de 35 corresponde a los grados de libertad: $T - K - 1 \rightarrow (40 - 4 - 1 = 35)$. K representa las variables explicativas del modelo ($X_1 = PE$, $X_2 = PC$, $X_3 = GP$ y $X_4 = R$).

El coeficiente de β_0 no es de interés por no estar asociado a ninguna variable explicativa. Si dicho parámetro fuese cero, solo nos estaría indicando que cuando todas las variables explicativas toman el valor de cero, se espera que el valor medio de Y sea también cero, o lo que es lo mismo, $E = (Y \setminus X_1 = X_2 = \dots = X_K = 0) = 0$.

Vamos a comprobar si las variables seleccionadas en el modelo para explicar las ventas de la editorial son individualmente significativas. Para ello, utilizaremos el resultado del modelo (3.6) junto con la regla de decisión (4.2) para realizar el contraste individual de los coeficientes.

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El precio de la editorial <i>NO</i> es una variable que explica las ventas de la editorial.		\mathcal{H}_a : El precio de la editorial es una variable que explica las ventas de la editorial.
$\mathcal{H}_0: \beta_1 = 0$		$\mathcal{H}_a: \beta_1 \neq 0$

Ahora, comparamos el valor de:

$$|t_i| = \left| \frac{\hat{\beta}_1}{s_{\hat{\beta}_1}} \right| = \left| \frac{-1,91}{0,49} \right| = 3,90 > t_{0,025}(35) = 2,030$$

Con el valor de $t_{\alpha/2}$ (T-K-1), dado que T-K-1 = 40 - 4 - 1 = 35, si elegimos un nivel de significancia de $\alpha = 5\%$, la tabla de la distribución t de Student proporcionará un valor de $t_{0,025}(35) = 2,030$, por lo que al comparar el valor obtenido para $|t|$ con el valor anterior encontramos que $3,90 > 2,028$, lo cual nos lleva a rechazar la hipótesis nula de que $\beta_1 = 0$ (\mathcal{H}_0 : el precio de los libros *NO* es una variable que explique las ventas de la editorial), por ende, el precio promedio de los libros es una variable significativa al 5%, es decir, el precio es una variable explicativa de las ventas de la editorial.

El valor crítico de comparación 2,030 de las tablas de t Student también puede hallarse de manera muy sencilla a través del uso del software econométrico Gretl²¹ de uso gratuito, muy potente y de fácil instalación y manipulación, disponible en <http://gretl.sourceforge.net/es.html>.

Simplemente, se requiere una vez instalado en programa en el disco duro del PC, abrir y seleccionar *Herramientas* → *Tablas estadísticas* y damos clic

²¹ Gretl es un paquete de cálculo econométrico gratuito que incluye una biblioteca compartida, un programa cliente de línea de instrucciones y un interfaz gráfico para el usuario (GUI). Gretl ofrece un interfaz intuitivo para el usuario; es muy fácil de instalar y de usar para realizar análisis econométrico. El paquete está muy ligado a los manuales de Econometría de Ramu Ramanathan, Jeffrey Woolridge y James Stock y Mark Watson lo cual permite ofrecer muchos archivos de datos para ejercicios prácticos así como ficheros con lotes de instrucciones (scripts). Estos son fácilmente accesibles y vienen acompañados de muchos comentarios. Para mayor información descargar en manual de usuario disponible en: <http://gretl.sourceforge.net/es.html>.

como se muestra en la Figura 4.3. En la ventana de diálogo que se abre, seleccionamos la tabla “t” y en la casilla grados de libertad (gl) digitamos 35 ($40-4-1$) y en la casilla probabilidad en la cola derecha digitamos con punto (no con coma): 0.025 ($\alpha = \frac{0,05}{2} = 0,025$) y damos Aceptar como se muestra en la Figura 4.4, y de este modo, obtenemos el valor crítico de $2,03011 \approx 2,030$ (véase Figura 4.5).

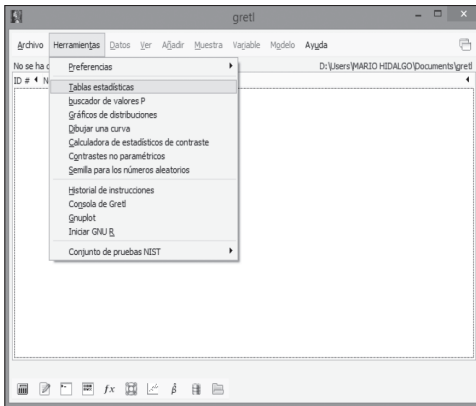


Figura 4.3

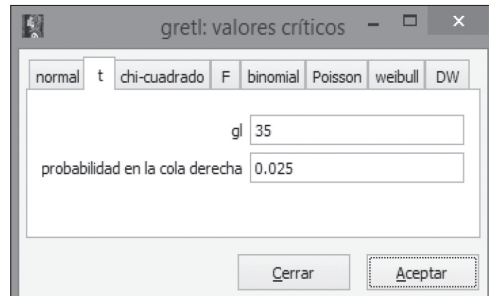


Figura 4.4

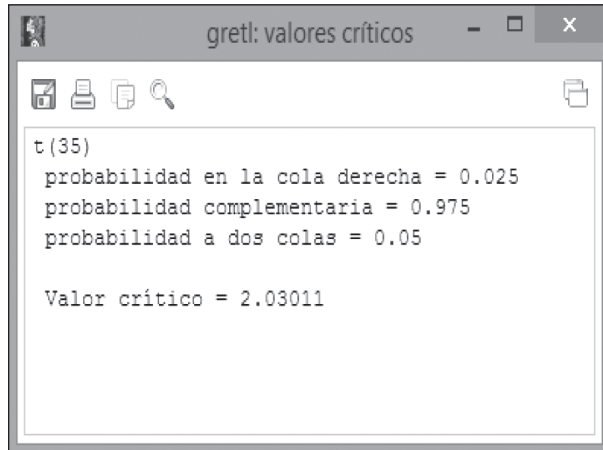


Figura 4.5

El contraste individual para los coeficientes de las variables PC, GP y R se llevará a cabo de la misma manera. Como los tres coeficientes han sido estimados en el mismo modelo con el mismo número de observaciones, el

valor crítico al 5% es igual a 2,030 para las tres variables. Basta, por lo tanto, calcular los respectivos estadísticos t :

Ahora realizaremos el contraste del parámetro PE con la ayuda de EViews. Para ello utilizaremos el test de Wald de la siguiente manera: en la ventana del objeto ecuación seleccionamos *View* → *Coefficient Diagnostics* → *Wald Test* – *Coefficient Restrictions* (Figura 4.6) y en el cuadro de diálogo *Wald Test* digitamos la restricción del parámetro o coeficiente asociado a la variable PE, como se muestra en la Figura 4.7 y damos OK.

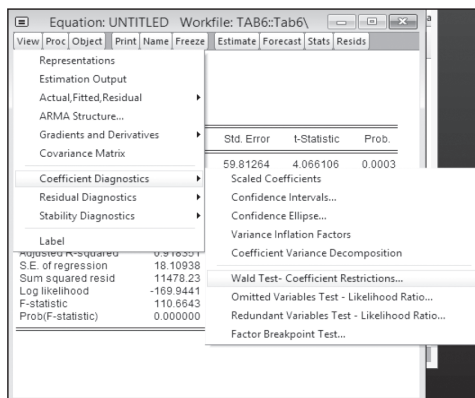


Figura 4.6

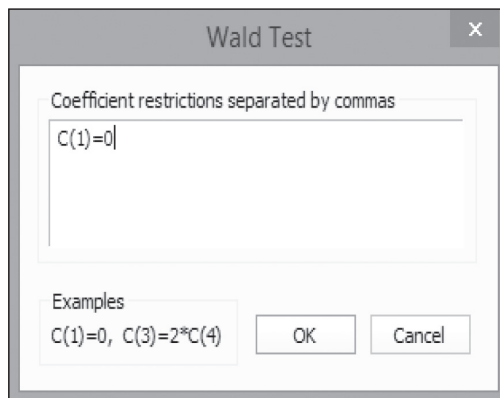


Figura 4.7

El resultado obtenido en la Tabla 4.4 arroja una probabilidad con el t estadístico (t -statistic) muy baja de 0,0004 ($0,0004 < 0,05$) por lo tanto, rechazamos la hipótesis nula, lo cual significa que el precio de la editorial es una variable relevante en la explicación de las ventas.

Tabla 4.4
Prueba de Wald.

Test Statistic	Value	df	Probability
t-statistic	-3,886406	35	0,0004
F-statistic	15,10415	(1, 35)	0,0004
Chi-square	15,10415	1	0,0001
Null Hypothesis: $C(1) = 0$			
Null Hypothesis Summary:			
Normalized Restriction (=0)		Value	Std. Error
C(1) → PE		-1,913099	0,492254

Restrictions are linear in coefficients.

Contraste individual de la variable *Precio de la competencia (PC)*:

La hipótesis de investigación propuesta por Emilio es que el precio de la competencia es la variable determinante de las ventas de la editorial, por lo tanto, la hipótesis nula se postula, así:

Para contrastar la hipótesis:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El precio de la editorial <i>NO</i> es una variable que explica las ventas de la editorial.		\mathcal{H}_a : El precio de la editorial es una variable que explica las ventas de la editorial.
$\mathcal{H}_0: \beta_1 = 0$		$\mathcal{H}_a: \beta_1 \neq 0$

Ahora, comparamos el valor de:

$$|t_i| = \left| \frac{\hat{\beta}_2}{s_{\hat{\beta}_2}} \right| = \left| \frac{0,127}{0,33} \right| = 0,38 < t_{0,025}(35) = 2,030$$

Como $0,38 < 2,030$ por lo tanto, no se rechaza la hipótesis nula, por ende, el precio de la competencia no es una variable significativa al 5% (esta variable individualmente no explica las ventas).

Hasta aquí, esta afirmación entristece un poco a Emilio y eleva el orgullo de Don Pablo, quien se siente parcialmente ganador en la discusión.

Para la restricción del parámetro asociado a la variable PC con EViews, seleccionamos *View* → *Coefficient Diagnostics* → *Wald Test – Coefficient Restrictions*. En el cuadro de diálogo *Wald Test* digitamos la restricción respectiva como se muestra en la Figura 4.8 y damos OK.

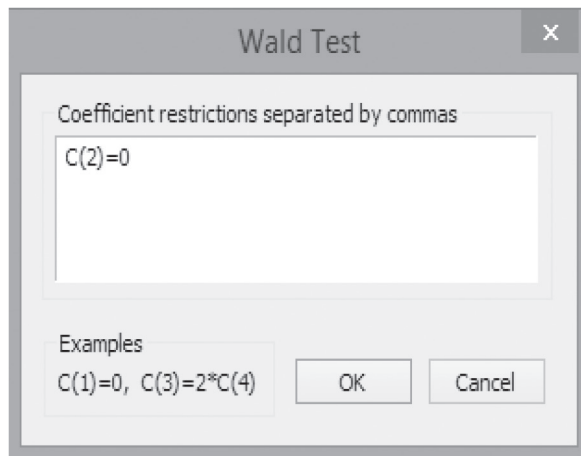


Figura 4.8

El resultado obtenido en la Tabla 4.5 arroja una probabilidad con el t estadístico (t -statistic) de 0,7012 ($0,7012 > 0,05$), por ende, no rechazamos la hipótesis nula, lo cual significa que el precio de la competencia no es una variable relevante en la explicación de las ventas.

Tabla 4.5**Prueba de Wald.**

Test Statistic	Value	df	Probability
t-statistic	0,386813	35	0,7012
F-statistic	0,149624	(1, 35)	0,7012
Chi-square	0,149624	1	0,6989
Null Hypothesis: C(2) = 0			
Null Hypothesis Summary:			
Normalized Restriction (=0)	Value	Std. Error	
C(2) → PC	0,126788	0,327776	

Restrictions are linear in coefficients.

Contraste individual de la variable *Gasto en publicidad (GP)*:

Otra de las hipótesis de investigación propuesta por Emilio es que el gasto en publicidad es una variable determinante de las ventas de la editorial, por lo tanto, la hipótesis nula se postula, así:

Para contrastar la hipótesis:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : La renta personal <i>NO</i> es una variable que explica las ventas de la editorial.		\mathcal{H}_a : La renta personal es una variable que explica las ventas de la editorial.
$\mathcal{H}_0: \beta_4 = 0$		$\mathcal{H}_a: \beta_4 \neq 0$

Ahora, comparamos el valor de:

$$|t_i| = \left| \frac{\hat{\beta}_3}{s_{\hat{\beta}_3}} \right| = \left| \frac{0,100}{0,006} \right| = 16,67 > t_{0,025}(35) = 2,030$$

Como $16,67 > 2,030$ por lo tanto, no se rechaza la hipótesis nula, por consiguiente, el gasto en publicidad es una variable significativa al 5% (esta variable individualmente explica las ventas).

Para la restricción del parámetro asociado a la variable GP con EViews, seleccionamos *View* → *Coefficient Diagnostics* → *Wald Test – Coefficient Restrictions*, y en el cuadro de diálogo *Wald Test* digitamos la restricción respectiva como se muestra en la Figura 4.9 y damos OK.

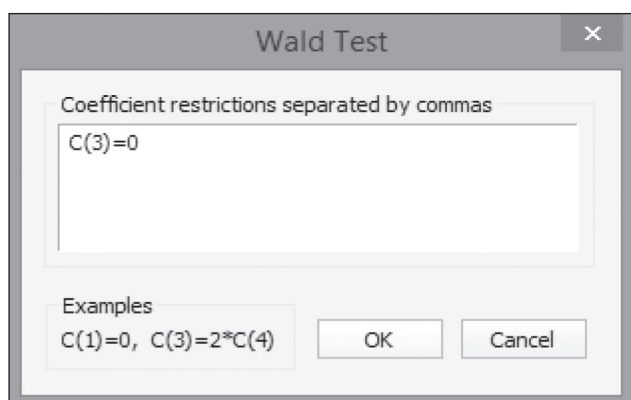


Figura 4.9

El resultado obtenido en la Tabla 4.6 arroja una probabilidad con el *t* estadístico (*t-statistic*) de 0,0000 ($0,0000 < 0$), por lo tanto, rechazamos la hipótesis nula, lo cual significa que el gasto en publicidad es una variable muy relevante en la explicación de las ventas de la editorial.

Tabla 4.6

Prueba de Wald.

Test Statistic	Value	df	Probability
t-statistic	16,63173	35	0,0000
F-statistic	276,6145	(1, 35)	0,0000
Chi-square	276,6145	1	0,0000
Null Hypothesis: $C(3) = 0$			
Null Hypothesis Summary:			
Normalized Restriction (=0)		Value	Std. Error
$C(3) \rightarrow GP$		0,100304	0,006031
Restrictions are linear in coefficients			

La última hipótesis de investigación propuesta por Emilio es que la renta o ingreso personal es una variable determinante de las ventas de la editorial, por lo tanto, la hipótesis nula se postula, así:

Para contrastar la hipótesis:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_0 : El gasto en publicidad <i>NO</i> es una variable que explica las ventas de la editorial.	frente	\mathcal{H}_a : El gasto en publicidad es una variable que explica las ventas de la editorial.
$\mathcal{H}_0: \beta_3 = 0$		$\mathcal{H}_a: \beta_3 \neq 0$

Ahora, comparamos el valor de:

$$|t_i| = \left| \frac{\hat{\beta}_4}{s_{\hat{\beta}_4}} \right| = \left| \frac{0,003}{0,0028} \right| = 1,07 < t_{0,025}(35) = 2,030$$

Como $1,07 < 2,030$, por lo tanto, no se rechaza la hipótesis nula, por consiguiente, la renta personal no es una variable significativa al 5% (esta variable individualmente no explica las ventas).

Hasta ahora, el contraste de hipótesis ha sido muy útil para concluir que efectivamente el precio de la editorial como lo afirmó inicialmente Don Pablo y el gasto en publicidad como lo hizo Emilio, son las variables determinantes de las ventas de la editorial. Por su parte, el precio de la competencia y la renta personal no demostraron individualmente tener significancia estadística.

Para la restricción del parámetro asociado a la variable R con Eviews, seleccionamos *View* → *Coefficient Diagnostics* → *Wald Test – Coefficient Restrictions* y en el cuadro de diálogo *Wald Test* digitamos la restricción respectiva como se muestra en la Figura 4.10 y damos OK.

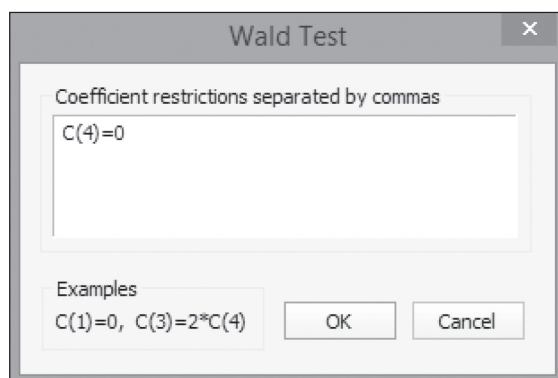


Figura 4.10

El resultado obtenido en la Tabla 4.7 arroja una probabilidad con el t estadístico (t -statistic) de 0,3524 ($0,3524 > 0,05$), por lo tanto, no rechazamos la hipótesis nula, lo cual significa que la renta personal no es una variable relevante en la explicación de las ventas de la editorial.

Tabla 4.7

Prueba de Wald.

Test Statistic	Value	df	Probability
t-statistic	0,942452	35	0,3524
F-statistic	0,888215	(1, 35)	0,3524
Chi-square	0,888215	1	0,3460
Null Hypothesis: $C(4) = 0$			
Null Hypothesis Summary:			
Normalized Restriction (=0)		Value	Std. Error
$C(4) \rightarrow R$		0,002669	0,002832

Restrictions are linear in coefficients.

4.3 Contraste conjunto del modelo 2

La hipótesis nula para el contraste conjunto de que todos los coeficientes que acompañan a las variables explicativas, excluyendo el intercepto (β_0) son cero, puede escribirse en los siguientes términos:

$\mathcal{H}_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$ frente a $\mathcal{H}_a: \beta_1 \neq 0, \beta_2 \neq 0 \dots \beta_K \neq 0$

O alternativamente, como: $\mathcal{H}_0: \beta^* = 0$ frente a $\mathcal{H}_a: \beta^* \neq 0$

Donde β^* representa el vector (K x 1) que incluye a los coeficientes β_1, \dots, β_K , pero excluye el intercepto.

$$\mathcal{F} = \frac{\hat{y}'\hat{y}}{\hat{\mu}'\hat{\mu}/(T-K-1)} = \frac{RSS/K}{SCR/(T-K-1)} \sim \mathcal{F}(K, T-K-1)$$

Dónde: RSS = Residual Sum of Squares (suma de cuadrados de los residuos)

Una expresión equivalente para este estadístico se puede escribirlo en términos de R^2 dividiendo el numerador y el denominador por la suma de los cuadrados total.

$$\mathcal{F} = \frac{R^2/K}{(1-R^2)/(T-K-1)} \sim \mathcal{F}(K, T-K-1)$$

Esta última expresión del estadístico de contraste de significación global es la más intuitiva y sencilla para realizar el contraste de que todos los coeficientes que acompañan a las variables incluidas en un modelo son conjuntamente cero, o, equivalentemente, el contraste para determinar si las variables explicativas en su conjunto son o no significativas a la hora de explicar la variable dependiente.

Conocida la distribución bajo la hipótesis nula del estadístico \mathcal{F} de interés, y dada la información muestral que nos permite obtener un valor específico de dicho estadístico \mathcal{F} , la regla de decisión se adopta en los siguientes términos. Si la hipótesis nula fuera cierta, el valor de $\hat{\beta}^*$ no debería estar muy alejado de cero, lo que proporcionaría un valor reducido para el numerador de la fracción. Por tanto, se rechaza la hipótesis nula $\beta^* = 0$, al nivel de significación del 5%, comparado con el valor crítico de $\mathcal{F}_{0,05}(K, T-K-1)$

obtenido de las tablas de la \mathcal{F} de Snedecor. En general, en el contraste global al nivel de significación del 100 $\alpha\%$, se sigue la siguiente regla de decisión:

$$\mathcal{F} = \frac{R_2/k}{(1-R_2)/(T-k-1)} > \mathcal{F}\alpha (K, T-K-1) \rightarrow \text{Rechazar } \mathcal{H}_0$$

En el caso de la editorial...

En primer lugar, puede comprobarse que el ajuste del segundo modelo es muy bueno ($\bar{R}^2 = 91,8$); el gasto en publicidad y el precio de la editorial, en particular, explican en buena medida las ventas. Por su parte, el precio de la competencia y la renta personal, mediante los contrastes individuales han demostrado no tener significancia estadística al 5%. Por último, un contraste de significación conjunta de los coeficientes de pendiente nos permitirá comprobar la capacidad explicativa del modelo en cuestión.

Para contrastar la hipótesis:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_0 : PE, PC, GP y R <i>NO</i> explican las ventas de la editorial.	frente	\mathcal{H}_a : PE, PC, GP o R explican las ventas de la editorial.
\mathcal{H}_0 : $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$		\mathcal{H}_a : $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 \neq 0, \beta_4 \neq 0$

Utilizamos la siguiente regla de decisión:

$$\mathcal{F} = \frac{R^2/K}{(1-R^2)/(T-k-1)} > \mathcal{F}_{0,05} (4, 35) \quad (4.3)$$

Comparamos el valor de:

$$\mathcal{F} = \frac{0,927/4}{(1 - 0,927)/(40 - 4 - 1)} = 111,11 > 2,64^{22} = \mathcal{F}_{0,05} (4, 35)$$

22 Corresponde al valor hallado en la tabla estadística F de Snedecor con 4 grados de libertad en el numerador, 35 ($T - K - 1 \rightarrow 40 - 4 - 1 = 35$) grados de libertad en el denominador y una probabilidad en la cola derecha de 0,05.

El valor crítico de comparación 2,030 de las tablas de la F de Snedecor también puede hallarse de manera muy sencilla a través Gretl, de la siguiente forma: seleccionamos Herramientas \rightarrow Tablas estadísticas y en la ventana de diálogo que se abre seleccionamos “F”; en la casilla grados de libertad en el numeral (gln) digitamos 4 ($K=4$ que corresponde al número de variables explicativas utilizadas en el modelo), en grados de libertad en el denominador (gld) editamos 35 ($T-K-1 \rightarrow 40 - 4 - 1 = 35$) y en la casilla probabilidad en la cola derecha digitamos con punto (no con coma) 0.05 y damos Aceptar, como se muestra en la Figura 4.11. De este procedimiento se obtiene el valor de $2,64147 \approx 2,64$ (véase Figura 4.12).

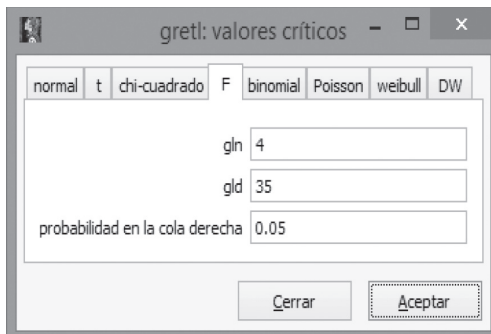


Figura 4.11

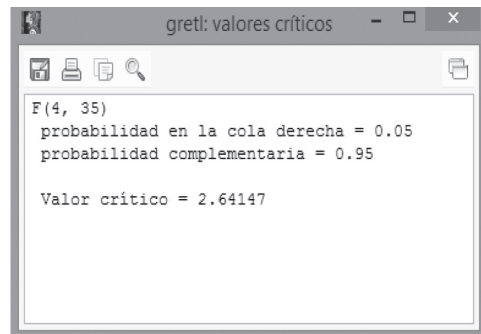


Figura 4.12

Se confirma, por tanto, que globalmente las cuatro variables tienen capacidad explicativa sobre las ventas de la editorial.

Para realizar con EViews el contraste global de las hipótesis sobre los parámetros de las variables explicativas, sobre la ventana del objeto ecuación seleccionamos: *View* \rightarrow *Representations* y damos clic. En la segunda fila de la Figura x aparece representada la ecuación estimada (*Estimation Equation*) del modelo de regresión múltiple que venimos trabajando, donde: $C(1)$, $C(2)$, $C(3)$ y $C(4)$ corresponden a los coeficientes de la pendiente de las variables explicativas PE, PC, GP y R, respectivamente. Esta simbología es importante tenerla en cuenta al momento de editar las restricciones en el cuadro de diálogo del contraste de Wald, como se indica en la Figura 4.13.

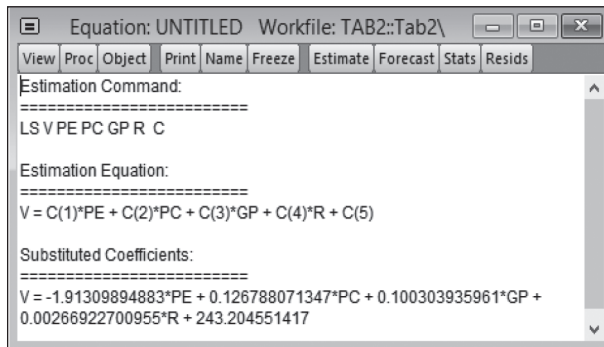


Figura 4.13

EViews permite contrastar restricciones en los parámetros a partir del contraste de Wald mediante *View* → *Coefficient Diagnostics* → *Wald Test – Coefficient Restrictions* (Figura 4.14). Rellenando la pantalla *Wald Test* como se indica en la Figura 4.15 se contrasta la significatividad conjunta de los parámetros estimados por esta vía (forma alternativa al contraste de la distribución \mathcal{F} de Snedecor realizado manualmente con anterioridad).

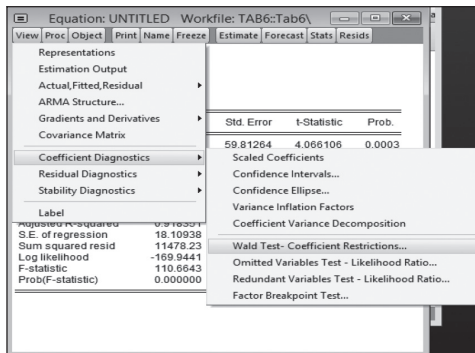


Figura 4.14

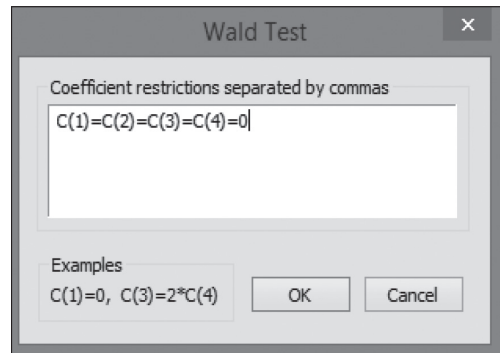


Figura 4.15

En la Tabla 4.8 se observa el rechazo de la hipótesis nula con una probabilidad del estadístico \mathcal{F} (*F-statistic*) de 0,0000 ($0,0000 < 0,05$), es decir, se acepta la significatividad conjunta de los parámetros del modelo, por consiguiente, el conjunto de variables incorporado al modelo son relevantes en la explicación de las ventas de la editorial.

Tabla 4.8
Prueba de Wald.

Test Statistic	Value	df	Probability
F-statistic	110,6643	(4, 35)	0,0000
Chi-square	442,6572	4	0,0000
Null Hypothesis: C(1) = C(2) = C(3) = C(4) = 0			
Null Hypothesis Summary:			
Normalized Restriction (=0)	Value	Std. Error	
C(1) → PE	-1,913099	0,492254	
C(2) → PC	0,126788	0,327776	
C(3) → GP	0,100304	0,006031	
C(4) → R	0,002669	0,002832	

Restrictions are linear in coefficients.

4.4 Contraste de significancia para un subconjunto de coeficientes.

Dado que el modelo contiene como variables explicativas dos precios, el de la propia editorial y el de la competencia, merece la pena contrastar si ambos precios tienen el mismo impacto sobre las ventas (de signo contrario, claro) así como la relevancia de ambos precios conjuntamente para explicar las ventas.

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El precio de la editorial y el precio de la competencia explican conjuntamente las ventas de la editorial.		\mathcal{H}_a : El precio de la editorial y el precio de la competencia NO explican conjuntamente las ventas de la editorial.
$\mathcal{H}_0: \beta_1 = -\beta_2$		$\mathcal{H}_a: \beta_1 \neq -\beta_2$

Lo que puede reformularse como:

$$\mathcal{H}_0: \beta_1 + \beta_2 = 0 \text{ frente a } \mathcal{H}_a: \beta_1 + \beta_2 \neq 0$$

Con EViews seleccionamos: *View* → *Coefficient Diagnostics* → *Wald Test* – *Coefficient Restrictions*, en el cuadro de diálogo de *Wald Test* escribimos la restricción de los parámetros como se muestra en la Figura 4.16 y damos OK.

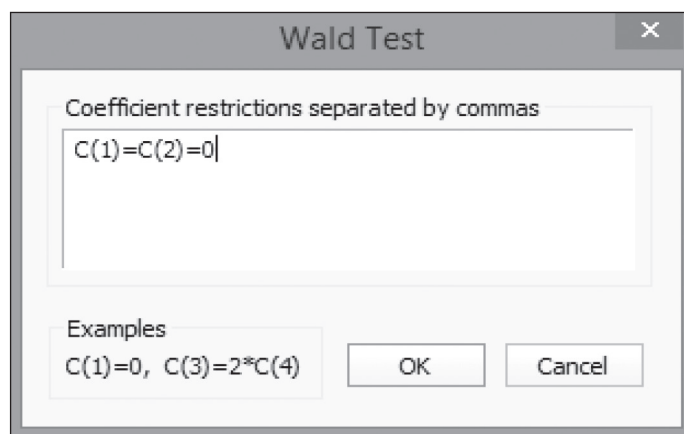


Figura 4.16

El contraste con el estadístico \mathcal{F} arroja una probabilidad de 0,0018 ($0,0018 < 0,05$) como muestra la Tabla 4.9, por lo tanto, rechazamos la hipótesis nula, y por ende, el precio de la editorial y el precio de la competencia son relevantes en la explicación de las ventas de la editorial.

Tabla 4.9

Prueba de Wald.

Test Statistic	Value	df	Probability
F-statistic	7,595819	(2, 35)	0,0018
Chi-square	15,19164	2	0,0005
Null Hypothesis: $C(1) = C(2) = 0$			
Null Hypothesis Summary:			
Normalized Restriction (=0)		Value	Std. Error
C(1) → PE		-1,913099	0,492254
C(2) → PC		0,126788	0,327776

Restrictions are linear in coefficients.

4.5 Contraste de hipótesis informativas

Respecto de la discusión anterior, un primer paso para decidir quién tiene razón en este debate se obtiene contrastando si la inclusión de las tres nuevas variables (PC, GP y R) ha sido o no acertada. Aunque el valor del coeficiente

de determinación del Modelo 2 es $R^2 = 0,927$ resulta ya bastante expresivo, debido al gran aumento que ha experimentado respecto del que se obtuvo con el modelo 1 con un $R^2 = 0,144$. La manera correcta de responder a esta pregunta consiste en contrastar conjuntamente los coeficientes de las tres nuevas variables añadidas. Este contraste puede realizarse a través de la diferencia entre los coeficientes de determinación del modelo de regresión lineal múltiple y el coeficiente de determinación del modelo de regresión simple, del siguiente modo:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : PC, GP y R explican las ventas de la editorial.		\mathcal{H}_α : PC, GP y R <i>NO</i> explican las ventas de la editorial.
\mathcal{H}_0 : $\beta_2 = \beta_3 = \beta_4 = 0$		\mathcal{H}_α : $\beta_2 \neq 0$, $\beta_3 \neq 0$ y $\beta_4 \neq 0$

Utilizamos la siguiente regla de decisión:

$$\mathcal{F} = \frac{(R^2 - R_R^2)/q}{(1 - R^2)/(T - k - 1)} > \mathcal{F}_{0,05}(4, 35) \quad (4.4)$$

Comparamos el valor de:

$$\mathcal{F} = \frac{(0,927 - 0,144)/2}{(1 - 0,927)/35} = 187,71 > 2,64 \rightarrow \mathcal{F}_{0,05}(4, 35)$$

Como $187,71 > 2,64$, por consiguiente, se rechaza la hipótesis nula en este contraste, poniendo en evidencia que las variables PC, GP y R que se han añadido en el modelo son conjuntamente significativas al 5%, es decir, son relevantes en la explicación de las ventas de la editorial.

A lo largo de este capítulo se han mostrado los procedimientos que deben seguirse para responder a las preguntas que Don Pablo y su hijo Emilio se hacían mutuamente al comienzo del mismo. Tras realizar los contrastes que se han presentado en este capítulo, debemos ser capaces ahora de responder a las preguntas pendientes: *¿Tiene razón Emilio o la tiene Don Pablo?, ¿o ambos tienen razón en parte?, y si, es así, ¿en qué parte aciertan cada uno de ellos?*

Con estos resultados, podemos responder a la discusión siguiente mantenida entre el padre e hijo.

Don Pablo: Lo ves, ya te lo decía yo. Ahí tienes la influencia del precio, con signo negativo.

Emilio: Admito que tiene un valor negativo, pero yo nunca he dicho que el precio no tenga influencia; sólo he dicho que no es el único factor que influye, y quizás, ni siquiera sea el más importante.

Don Pablo: Pero ¿es que no ves que ese es el único factor importante? No tienes más que mirar a su coeficiente, que no sólo es negativo, sino que es el que mayor valor que toma en el conjunto de parámetros (-1,91).

Emilio: ¡Qué tiene que ver el valor que toma! Lo que importa no es el valor del coeficiente, sino saber si hay mucha o poca evidencia a favor de ese valor. Podría darse que, a pesar de que la cifra obtenida sea alta, su importancia real sea inexistente.

Don Pablo: Tú solamente quieres discutir, sin admitir que yo tengo la razón. ¿Cómo puede ser que no tenga importancia? Si quieres seguir discutiendo de esta cuestión tendrás que demostrarme que lo que afirmas es verdad.

Los contrastes individuales se realizarán comparando los valores obtenidos para el estadístico con el valor $t_{0,025}(35) = 2,030$, de donde se deduce que se rechazan dos hipótesis nulas, $\mathcal{H}_0: \beta_1 = 0$ y $\mathcal{H}_0: \beta_3 = 0$ mientras que no se rechaza la hipótesis $\mathcal{H}_0: \beta_2 = 0$ y $\mathcal{H}_0: \beta_4$. Es decir, de acuerdo con la muestra utilizada, hay evidencia de que las variables precio de la editorial (PE) y gasto en publicidad (GP) son relevantes para explicar las ventas, mientras que el precio de la competencia (PC) y la renta personal (R) no parecen ser significativas en el modelo en discusión.

Los resultados de los contrastes ponen de manifiesto que, en la discusión entre Don Pablo y su hijo Emilio, ambos tienen parte de razón. Acierta Don Pablo al decir que el precio (PE) al que venden sus libros es una variable

importante para explicar las ventas de la editorial, pero queda claro también que no es el único factor a considerar. El gasto realizado en publicidad (GP) aparece como una nueva variable importante para explicar las ventas, dando así la razón a Emilio. Sin embargo, el precio a que la competencia vende su producto y la renta personal de los clientes no resultan ser factores relevantes en el comportamiento de las ventas de la editorial, ya que no hay evidencia en la muestra de que tengan un efecto significativo. Este resultado, por cierto, permite a Don Pablo entender lo que su hijo pretendía decirle; aunque el precio de la competencia tiene un coeficiente estimado (0,127) ligeramente mayor que el coeficiente del gasto en publicidad (0,100) y mucho mayor que el coeficiente estimado de la renta personal (0,003), resultan ser estas dos variables (PC y R) poco relevantes, mientras que es claramente significativo el gasto en publicidad.

Por otra parte, el contraste conjunto de las cuatro variables arroja un resultado un tanto diferente al contraste individual. Globalmente el conjunto de variables incorporadas en el modelo tienen capacidad explicativa sobre las ventas de la editorial, por consiguiente, padre e hijo deciden zanjar sus diferencias, conservando el modelo inicial estimado en el modelo (3.6). Por el momento, ninguno se atreve a eliminar ninguna variable, aunque la discusión prosigue en el siguiente capítulo.

RETO 3:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

**ENCARANDO EL MODELO DE
REGRESIÓN LINEAL GENERAL:
DIAGNÓSTICO ESTRUCTURAL**

En los capítulos anteriores se pudo establecer que los valores de la variable dependiente Y puede escribirse como: $Y = X\beta + \mu$ o bien como $Y = X\hat{\beta} + \hat{\mu}$, por lo tanto, los residuos serán: $\hat{\mu} = Y - X\hat{\beta}$.

Para Alonso et al., (2010:121), si utilizamos $\hat{\beta}$ como la mejor aproximación, dada la muestra, a los coeficientes desconocidos β_i , las igualdades anteriores sugieren que los residuos $\hat{\mu}$ serán también aproximaciones adecuadas a las perturbaciones no observables μ , donde $\hat{\mu}$ será denominada residuos mínimo cuadráticos o residuos MCO.

Tanto las perturbaciones como los residuos mínimo cuadráticos siguen distribuciones centradas en cero, esto es, $E(\mu) = E(\hat{\mu}) = 0$. Esta propiedad de los residuos se deriva de la distribución de las perturbaciones, que se suponen centradas en cero, dada su definición. En efecto, en el caso de que el modelo esté correctamente especificado, la parte sistemática X del mismo, será capaz de explicar las variaciones de Y y únicamente quedarán sin explicar los factores puramente aleatorios e impredecibles.

En todo caso, los residuos contienen evidencia útil sobre el modelo; si las hipótesis realizadas en su especificación son correctas, el comportamiento observado de los residuos debería ser el esperado. Desviaciones de este comportamiento esperado puede así alertar sobre posibles errores en la especificación del modelo.

El investigador que realiza trabajo empírico suele cometer errores de varios tipos a lo largo del proceso de elaboración de su modelo. La etapa de especificación de un modelo econométrico es la etapa de mayor dificultad y en la que con más frecuencia pueden cometerse errores de juicio (Ibíd., 149).

Como a Don Pablo no le gusta perder en las discusiones, aunque reconoce que su hijo le ha convencido con evidencia clara de que la variable *Gasto en publicidad* ha tenido su influencia, no olvida que también Emilio insistía en que el precio de la competencia y la renta personal eran importantes y han resultado no serlo. Por eso, no está dispuesto todavía a dar su brazo a torcer. Además, como ha pedido explicaciones detalladas de todo lo que se ha hecho para alcanzar esas conclusiones, está informado de que ha sido preciso hacer determinadas hipótesis. De manera que decide ponerlas en cuestión.

Don Pablo: De acuerdo. Me has convencido de que la publicidad que hemos hecho ha tenido un efecto sobre nuestras ventas. Pero también decías que eran importantísimos el precio de la competencia y la renta personal y ha resultado, como yo decía, que no era así. Además, por lo que me has explicado, ha habido que hacer bastantes hipótesis que no tienen por qué cumplirse y que a lo mejor han influido en los resultados.

Emilio: Hay que ver qué terco eres. Nunca estás dispuesto a aceptar que puedes estar equivocado. Sí, claro que ha habido que hacer hipótesis para llegar a los resultados, pero eso no cambia nada.

Don Pablo: ¿Y por qué tengo yo que creer eso que dices? ¿No me has dicho antes que son hipótesis simplificadoras y que pueden no ser muy realistas? Pues en caso de no cumplirse podrían estar dando una imagen incorrecta de cuáles son las variables que influyen en las ventas de la editorial.

Emilio: Bien, ya veo que no hemos acabado todavía con esta cuestión. Estoy dispuesto a comprobar si hay alguna evidencia de que las hipótesis que hemos hecho no se cumplen y a convencerte del resultado.

5.1 Selección de variables explicativas: error de especificación.

Siguiendo a Alonso et al, (2005), la investigación en economía parte de alguna teoría relacionada con la determinación de la variable dependiente objeto de estudio y, en primer lugar, trata de especificar un conjunto de variables explicativas que describan adecuadamente dicha teoría. Ahora bien, los modelos que propone la teoría económica son modelos que simplifican un mundo más complejo, extrayendo de él únicamente los rasgos que se consideran más importantes para el análisis. Es bastante natural, por esto, que la etapa de especificación del modelo econométrico correspondiente no esté libre de errores.

Aunque, en sentido amplio, error de especificación sería cualquier error que se cometa en la especificación de un modelo, tradicionalmente se denomina error de especificación meramente al error cometido en la selección de las variables explicativas consistente en utilizar un conjunto de regresores incluidos en la matriz X que no es el correcto. En este sentido, es posible cometer dos tipos de errores: bien por defecto omitiendo variables relevantes (cuyo coeficiente poblacional β no es cero), o bien por exceso, incluyendo variables irrelevantes (cuyo coeficiente poblacional realmente es cero).

A pesar de toda la evidencia reunida, la discusión entre padre e hijo no se ha acabado. Don Pablo admite que el modelo (3.6) está bien, pero como sólo le interesa conocer el efecto del precio sobre las ventas, por tanto, se propone utilizar para su análisis el modelo (2.9), a lo que Emilio presenta ciertas objeciones.

Don Pablo sigue sin entender del todo la postura de su hijo sobre incluir otras variables explicativas además del precio para explicar las ventas de la editorial. Sigue empeñado en que su propuesta de incluir sólo el precio como variable explicativa es suficiente para tomar las decisiones que le interesan sobre su política de precios.

Don Pablo: Yo propongo un modelo simple donde el precio promedio es la única y principal variable explicativa de las ventas de la editorial, por consiguiente, mi modelo propuesto es el siguiente:

$$\widehat{Ventas} = 686,57 - 3,53\text{Precio}$$

Emilio: Mi modelo en ningún momento desconoce el poder explicativo del precio sobre las ventas, pero no estoy de acuerdo, en que esta variable sea el único factor explicativo. Mi tesis se centra en que, además del precio promedio, las ventas de la editorial también dependen del precio de la competencia, el gasto en publicidad y la renta personal de nuestros clientes, por consiguiente, propongo el siguiente modelo:

$$\widehat{Ventas} = 243,20 - 1,91PE_t + 0,127PC_t + 0,100GP_t + 0,003R_t$$

Don Pablo: En el modelo que yo propongo se ha obtenido un coeficiente estimado igual a -3,53 y un $|t| = 2,54$ ¿No es esto suficiente?

Emilio: ¡Papá!, estas cifras no son relevantes en estas condiciones. Las consecuencias de no incluir las otras variables son muy graves.

Don Pablo: ¿En qué sentido son muy graves? No entiendo el porqué de esa gravedad si el coeficiente estimado asociado a los precios no cambia tanto. En tu modelo adopta el valor de -1,91 y en el mío -3,53.

Emilio: Te voy a demostrar de nuevo, que los valores numéricos no son lo único importante, tan importante es la confianza o fiabilidad que nos merecen dichos valores.

5.2 Errores de especificación en las variables explicativas.

El proceso de construcción de un modelo econométrico se inicia con la especificación de la relación a estimar y la formulación de un conjunto de hipótesis. Por ello, es conveniente someter al modelo elaborado y estimado a diversas pruebas estadísticas que permitan comprobar su validez y calidad antes de utilizarlo en el trabajo empírico.

Para Carrascal et al, (2004), existen diferentes pruebas de especificación y diagnosis relativas a la especificación de la parte sistemática del modelo (selección de las variables explicativas, forma funcional que relaciona las variables) y a las propiedades que debe cumplir la perturbación aleatoria. Las consecuencias de los diversos errores son conocidas y la solución consiste, en muchas ocasiones, en re-especificar el modelo incorporando las correcciones oportunas. Uno de los principales problemas a los que se enfrenta un investigador es determinar qué variables se incluyen como explicativas en un modelo. Aunque en la práctica no existen procedimientos para poder determinar de forma inequívoca el número de variables a introducir, los errores que pueden cometerse en la especificación de las variables explicativas son dos: omitir variables relevantes o incluir variables irrelevantes (pp.185-186).

5.2.1 Omisión de variables relevantes.

La omisión de variables relevantes hace que el término de perturbación recoja su efecto, por lo que este deja de cumplir las hipótesis clásicas planteadas en el modelo de regresión. El resultado es que tanto los estimadores de los coeficientes de regresión (excepto en el caso de ortogonalidad entre las variables omitidas en incluidas) como el de la varianza de las perturbaciones dejan de ser insesgados y consistentes.

Por un instante, vamos a suponer que el gasto en publicidad (GP_t) no es relevante en la explicación de las ventas de la editorial, es decir, las hipótesis se plantearían de la siguiente manera:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_0 : El gasto en publicidad es una variable relevante en el modelo.	frente	\mathcal{H}_a : El gasto en publicidad no es una variable relevante en el modelo.

Para ello, abrimos en EViews el archivo *Tab2*. En la ventana del Workfile seleccionamos con la tecla CTRL y el ratón las variables V y PE. Posteriormente, abrimos estas variables como ecuación (*Open* → *as Equation*) y en la ventana de diálogo de estimación de la ecuación (*Equation Estimation*) y damos Aceptar. En la ventana del objeto ecuación resultante, seleccionamos *View* → *Coefficient Diagnostic* → *Omitted Variables Test – Likelihood ratio* y damos clic como se muestra en la Figura 5.1. En el cuadro de diálogo que se abre escribimos GP y damos OK, como se muestra en la Figura 5.2.

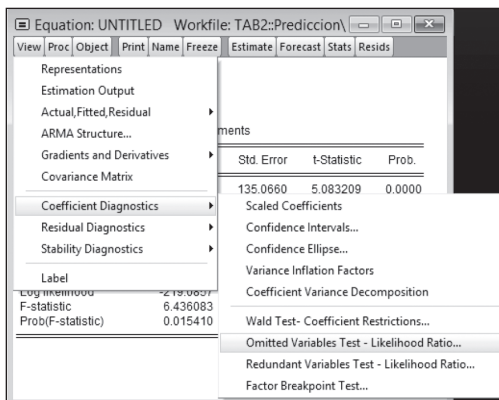


Figura 5.1

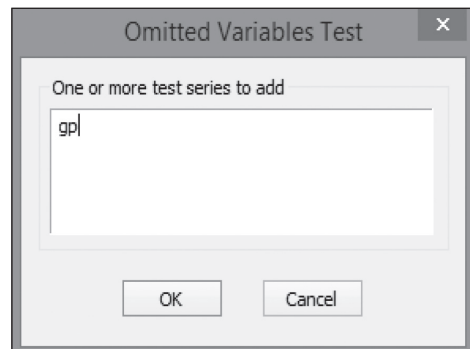


Figura 5.2

En la Tabla 5.1 se visualizan los resultados de la prueba de variables omitidas, cuya hipótesis nula es que el gasto en publicidad no es una variable relevante en la explicación de las ventas de la editorial. Los estadísticos t, F y Likelihood ratio tienen una probabilidad de 0,0000 al 5% de significación estadística ($0,0000 < 0,05$). Por su parte, la probabilidad del estadístico de máxima verosimilitud (*Likelihood ratio*) es inferior al 5% ($0,0000 < 0,05$), por lo tanto, no rechazamos la hipótesis nula del contraste, pues la probabilidad de equivocarnos si lo hiciéramos es de 0% ($0,0\% < 5\%$).

De otro lado, en la parte inferior de la misma tabla, se observa que el modelo estimado donde se incluye el gasto en publicidad arroja un coeficiente de determinación ajustado de 91,89%, muy superior al coeficiente de determinación ajustado de 12,24% del modelo de regresión de las ventas contra el precio promedio de la editorial (2.9) analizado en el capítulo 2.

Tabla 5.1
Prueba de variables omitidas.

Null hypothesis: GP are jointly significant

Specification: V C PE

Omitted variables: GP

Test Statistic	Value	df	Probability
t-statistic	19,34961	37	0,0000
F-statistic	374,4074	(2, 37)	0,0000
Likelihood ratio	96,34664	1	0,0000

F-test summary:

	Sum of Sq.	df	Mean Squares
Test SSR	121.911,2	1	121.911,2
Restricted SSR	133.958,8	38	3.525,231
Unrestricted LogL	12.047,6	37	325,6110

LR test summary:

	Value
Restricted LogL	-219,0857
Unrestricted LogL	-170,9124

Unrestricted Test Equation

Dependent Variable: V

Method: Least Squares

Sample: 2009Q1: 2018Q4

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	283,827	46,02427	6,166899	0,0000
PE	-2,209645	0,422866	-5,154749	0,0000
GP	0,103799	0,005364	19,34961	0,0000
R-squared	0,923091	Mean dependent var		344,7425
Adjusted R-squared	0,918934	S.D. dependent var		63,37666
S.E. of regression	18,04469	Akaike info criterion		8,695619
Sum squared resid	12.047,61	Schwarz criterion		8,822285
Log likelihood	-170,9124	Hannan-Quinn criterion		8,741417
F-statistic	222,0439	Durbin-Watson stat		1,696297
Prob(F-statistic)	0,000000			

5.2.2 Inclusión de variables irrelevantes.

La inclusión de variables irrelevantes supone no incorporar en la especificación del modelo una hipótesis cierta, es decir, que el parámetro de la variable irrelevante en cuestión es igual a cero (0). Como consecuencia, aunque los estimadores mínimo cuadráticos mantienen sus propiedades, se estiman con menor precisión, pues no tienen en cuenta dicha información.

Don Pablo ha estado muy atento al desarrollo econométrico realizado por Emilio, pero en verdad, no comprende varias de las conclusiones a las que su hijo ha llegado después de efectuar los contrastes tanto individual como global, por ello, decide preguntar: Emilio, me parece que te confundes en tus explicaciones, por un lado, dices que el precio de la competencia y la renta personal de nuestros clientes no son individualmente significativas en la explicación de las ventas de la editorial (PC: $0,7012 > 0,005$ y R: $0,3524 > 0,05$); y luego, afirmas que todas las variables del modelo (incluidas estas dos) tienen significatividad conjunta, es decir, que son relevantes en la explicación del modelo. Hijo, ahora dime: el precio de la competencia y la renta personal, ¿deben o no ser incluidas en el modelo que tanto defiendes?

Emilio bastante seguro del procedimiento realizado, le responde a su padre: Mira papá, por supuesto, que todas las variables incorporadas al modelo son relevantes para explicar las ventas de la editorial, pero para quedes satisfecho con mi respuesta, realizaré un último procedimiento con la ayuda de EViews. Usaremos la opción de variables redundantes, la cual sirve para contrastar la significación de una o varias variables explicativas incluidas en el modelo de regresión; si dichas variables son irrelevantes, entonces serán eliminadas del modelo.

Sobre la ventana del objeto ecuación seleccionamos *View* → *Coefficient Diagnostics* → *Redudant Variables Test – Likelihood Ratio* y damos clic como se muestra en la Figura 5.3; en el cuadro de diálogo digitamos separadas por un espacio, los nombres de las variables individualmente no significativas, tales como el precio de la competencia (PC) y la renta personal (R) como se muestra en la Figura 5.4 y damos OK. EViews realiza un contraste en el que

la hipótesis nula especifica que los coeficientes de las variables redundantes son iguales a cero.

Contraste de hipótesis:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El precio de la competencia y la renta personal son conjuntamente irrelevantes en el modelo.		\mathcal{H}_a : El precio de la competencia y la renta personal son conjuntamente relevantes en el modelo.

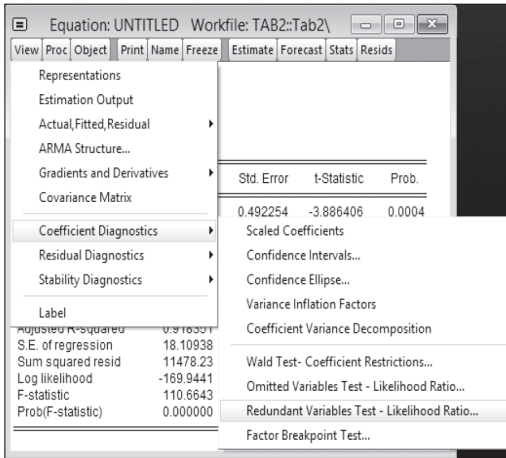


Figura 5.3

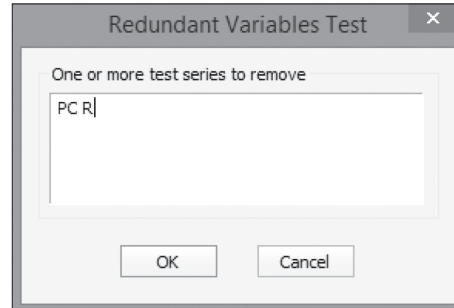


Figura 5.4

Si analizamos el estadístico de máxima verosimilitud (*Log Likelihood ratio*) que se muestra en la Tabla 5.2, no rechazamos la hipótesis nula del contraste, pues la probabilidad de equivocarnos si lo hiciéramos es de 0,3797, o sea, (37,97% > 5%), por lo tanto, podemos considerar que el precio de la competencia (PC) y la renta personal (R) no son variables redundantes en el modelo que explica las ventas de la editorial. Además, podemos observar que el coeficiente de determinación ajustado del modelo sin variables redundantes es de 91,89%, levemente superior (cuando el precio de la editorial y el gasto en publicidad son las únicas variables explicativas) al coeficiente de determinación ajustado del modelo original de 91,84% (91,89% - 91,84% = 0,05%), lo que indica que en nuestro modelo no estamos incluyendo ninguna

variable irrelevante; o dicho en otras palabras, las variables incluidas en el modelo hasta el momento son relevantes. Por otra parte, podemos comprobar que las variables PE y GP que permanecen en la ecuación son individualmente significativas, ya que la probabilidad de rechazar la hipótesis nula de que sus coeficientes son iguales a cero siendo cierta es inferior a 0,05 (PE \rightarrow Prob. 0,0000 > 0,05 y GP \rightarrow Prob. 0,0000 > 0,05).

Tabla 5.2***Prueba de variables redundantes.***

Null hypothesis: PC R are jointly insignificant

Specification: V C PE PC GP R

Redundant variables: PC R

Test Statistic	Value	df	Probability
F-statistic	0,868081	(2, 35)	0,4286
Likelihood ratio	1,936543	2	0,3797
F-test summary:			
	Sum of Sq.	df	Mean Squares
Test SSR	569,3737	1	284,6869
Restricted SSR	12.047,61	37	325,6110
Unrestricted LogL	11.478,23	35	327,9495

LR test summary:

	Value
Restricted LogL	-170,9124
Unrestricted LogL	-169,9441

Restricted Test Equation

Dependent Variable: V

Method: Least Squares

Sample: 2009Q1: 2018Q4

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2,209645	0,428662	-5,154749	0,0000
PE	0,103799	0,005364	19,34961	0,0000
GP	283,8270	46,02427	6,166899	0,0000
R-squared	0,923091	Mean dependent var		344,7425
Adjusted R-squared	0,918934	S.D. dependent var		63,37666
S.E. of regression	18,04469	Akaike info criterion		8,695619
Sum squared resid	12.047,61	Schwarz criterion		8,822285
Log likelihood	-170,9124	Hannan-Quinn criterion		8,741417
F-statistic	222,0439	Durbin-Watson stat		1,696297
Prob(F-statistic)	0,000000			

Emilio: ¡Papá!, ¿comprendes ahora por qué los resultados del modelo donde se incluye solamente la variable precio no son relevantes? Ni la estimación, ni el valor de la t son fiables.

Don Pablo: Comprendo, pero entonces, también tu modelo tiene problemas, ya que las variables precio de la competencia de la competencia y renta personal no son relevantes individualmente.

Emilio: Si bien es cierto, que el precio de la competencia y la renta personal no son individualmente significativas, conjuntamente si lo son. Veo que tú estás confundiendo las cosas nuevamente. Te lo demostraré.

5.3 Análisis de la estabilidad estructural.

Siguiendo a Carrascal et al, (2004: 187), una de las hipótesis que suponemos cumple el modelo de regresión especificado es que los coeficientes se mantienen constantes para todo el periodo muestral. Sin embargo, es posible que existan submuestras para las que el comportamiento del modelo y su estructura sea diferente, siendo necesario contrastar esta posibilidad. EViews ofrece varias alternativas para analizar la estabilidad de los parámetros del modelo a partir de los contrastes de Chow y la estimación recursiva.

5.3.1 Contraste de Chow.

El test de Chow es un contraste estadístico para detectar un posible cambio estructural, es decir, un cambio profundo en el sistema que se trata de representar a través de un modelo y que provoque que los coeficientes del modelo dejen de ser constantes, incumpléndose de esta forma una de las hipótesis del modelo básico de regresión lineal. A través de este contraste, lo que se pretende es comprobar si los parámetros se mantienen constantes a lo largo de toda la muestra o si cambian de un valor a otro.

El hecho de que no haya estabilidad en los parámetros genera el mismo efecto que la falta de linealidad, pues el modelo estaría mal especificado y las

estimaciones serían sesgadas e inconsistentes, lo cual significa que el método de estimación deja de ser válido, por ese motivo detectar la inestabilidad de parámetros es fundamental.

La hipótesis nula de estabilidad estructural plantea un solo modelo para el conjunto de las observaciones (el modelo restringido), mientras que la hipótesis alternativa establece un comportamiento diferente en cada grupo en el que se divide la muestra. Por lo tanto, el modelo sin restringir es aquel en el que los parámetros pueden cambiar de una submuestra a otra.

La hipótesis nula de estabilidad estructural y la hipótesis alternativa de cambio estructural en términos de los parámetros se definen como:

$$H_0 = \begin{bmatrix} \beta^{1_0} \\ \beta^{1_1} \\ \vdots \\ \beta^{1_k} \end{bmatrix} = \begin{bmatrix} \beta^{2_0} \\ \beta^{2_1} \\ \vdots \\ \beta^{2_k} \end{bmatrix} \quad H_1 = \begin{bmatrix} \beta^{1_c} \\ \beta^{1_1} \\ \vdots \\ \beta^{1_k} \end{bmatrix} \neq \begin{bmatrix} \beta^{2_c} \\ \beta^{2_1} \\ \vdots \\ \beta^{2_k} \end{bmatrix}$$

Contraste de hipótesis

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_0 : No hay cambio estructural en el modelo.	frente	\mathcal{H}_a : Hay cambio estructural en los parámetros del modelo.

Para realizar el contraste de Chow en EViews, abrimos el archivo de trabajo Tabla 3.1, pulsando la tecla CTRL y con el botón izquierdo del ratón seleccionamos las variables V, PE, PC, GP y R en dicho orden. Posteriormente, abrimos el grupo de variables como ecuación, obteniendo la estimación del modelo.

Para analizar la estabilidad del modelo, situados en la ventana del objeto ecuación seleccionamos la opción *View* → *Stability Diagnostics* → *Chow Breakpoint Test* y damos clic como se muestra en la Figura 5.5. En el cuadro de diálogo que se abre editamos el número 2013, valor que corresponde a la mitad del periodo de la muestra analizada (2009-2018) y damos OK como se

indica en la Figura 5.6. Lo anterior genera una partición del conjunto total de observaciones en dos: una submuestra de 2009 a 2013 y otra de 2014 a 2018.

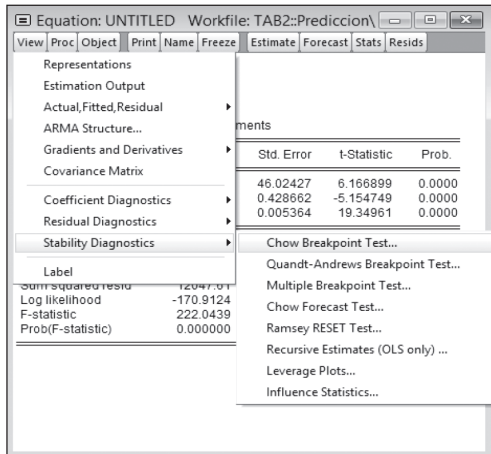


Figura 5.5

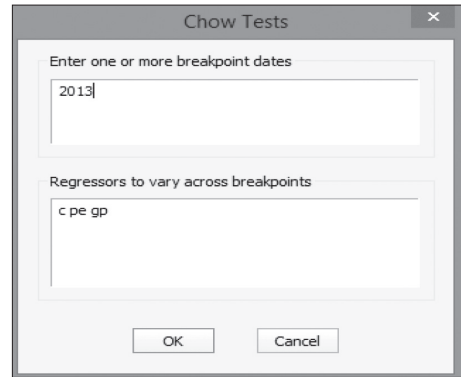


Figura 5.6

En la Tabla 5.3 aparecen los estadísticos que calcula EViews (*F-statistic*, *Log likelihood ratio* y *Wald statistic*) para contrastar un cambio estructural en 2013. El valor del estadístico F y su probabilidad indican el no rechazo (aceptación) de la hipótesis nula de estabilidad estructural, puesto que la probabilidad de rechazar dicha hipótesis siendo cierta supera el 0,05 ($F\text{-statistic} \rightarrow \text{Prob. } F(5, 30) \rightarrow 0,2556 > 0,05$). Igualmente, los estadísticos Log likelihood ratio y Wald statistic son superiores al 5%; por ende, concluimos que en el modelo estimado no hay ningún cambio estructural en el año 2013.

Tabla 5.3
Prueba de Chow.

Null hypothesis: No braks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2009Q1 2018Q4

F-statistic	2,349641	Prob. F(3, 34)	0,0898
Log likelihood ratio	7,536164	Prob. Chi-Square (3)	0,0566
Wald Statistic	7,048924	Prob. Chi-Square (3)	0,0704

5.3.2 Contraste de predicción de Chow.

La idea de este contraste consiste en comparar si las últimas observaciones muestrales disponibles presentan cambios respecto a las anteriores. Los estadísticos del contraste son dos: el estadístico F y el estadístico χ^2 de razón de verosimilitud. El primero compara la suma cuadrada de los residuos ($\mu' \mu$) del modelo estimado con el total de observaciones (T) con la suma cuadrada de los residuos ($\mu^1 \mu^1$) del modelo que utiliza solo una parte (T1) de la muestra.

Para realizar este contraste, en la ventana del objeto ecuación seleccionamos *View* → *Stability Diagnostics* → *Chow Forecast Test* y damos clic como se muestra en la Figura 5.7. En el cuadro de dialogo que se abre a continuación editamos nuevamente el año 2013 y damos OK (véase Figura 5.8).

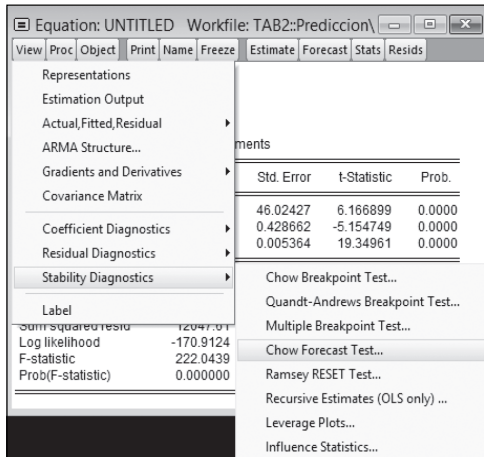


Figura 5.7

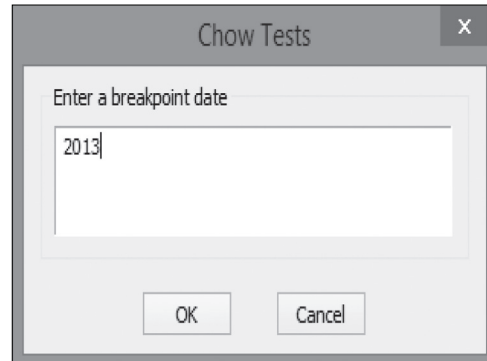


Figura 5.8

Los resultados de la Tabla 5.4 incluyen los valores de los estadísticos de contraste de la submuestra 2013Q1 a 2018Q4 y la estimación del modelo para la submuestra 2009Q1 a 2012Q4. Las probabilidades de los estadísticos F y del estadístico χ^2 de razón de verosimilitud son superiores a 0,05 (F-statistic → 0,7512 > 0,05 y Likelihood ratio → 0,0791 > 0,05); por consiguiente, no se rechaza la hipótesis nula de estabilidad estructural. En conclusión, en el modelo estimado no se produce un cambio estructural en el año 2013.

Tabla 5.4***Prueba de pronóstico de Chow.***

Specification: V C PE GP

Test predicciones for observaciones from 2013Q1 to 2018Q4

Test Statistic	Value	df	Probability
F-statistic	0,735816	(24, 13)	0,7512
Likelihood ratio	34,31985	24	0,0791
F-test summary:			
	Sum of Sq.	df	Mean Squares
Test SSR	6.939,290	24	289,1371
Restricted SSR	12.047,61	37	325,6110
Unrestricted SSR	5.108,32	13	392,9474
LR test summary:			
	Value		
Restricted LogL	-170,9124		
Unrestricted LogL	-153,7525		

Unrestricted log likelihood adjusts test equation results to account for observations in forecast sample.

Restricted Test Equation

Dependent Variable: V

Method: Least Squares

Sample: 2009Q1: 2012Q4

Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	158,6530	96,11745	1,650616	0,1228
PE	-0,526595	0,999896	-0,526650	0,6073
GP	0,085096	0,011424	7,449110	0,0000
R-squared	0,815971	Mean dependent var		309,5438
Adjusted R-squared	0,787659	S.D. dependent var		43,01798
S.E. of regression	19,82290	Akaike info criterion		8,978913
Sum squared resid	5.108,316	Schwarz criterion		9,123774
Log likelihood	-68,83131	Hannan-Quinn criterion		8,986331
F-statistic	0,000017	Durbin-Watson stat		1,949101
Prob(F-statistic)				

5.3.3 Estimación de residuos recursivos.

Una buena manera de empezar a comprobar si los residuos muestran o no un comportamiento adecuado, es a través de su análisis gráfico. En este apartado se analizará el comportamiento de los residuos del modelo estimado (3.6).

Otra alternativa para analizar la estabilidad del modelo es la estimación recursiva. Esta técnica es adecuada cuando trabajamos con series temporales y se desconoce el momento en que se ha producido un cambio estructural, por lo que el análisis de la estabilidad del modelo se realiza en forma general.

La forma en que EViews presenta los resultados de la estimación recursiva y el análisis de la estabilidad del modelo que de ella se deriva es mediante las representaciones gráficas de los coeficientes y de los residuos recursivos. Para ello seleccionamos *View* → *Stability Diagnostics* → *Recursive Estimates (OLS only)* y damos clic como se muestra en la Figura 5.9. En la ventana de diálogo que se abre a continuación, elegimos la opción *Recursive Residuals* y damos OK (Figura 5.10).

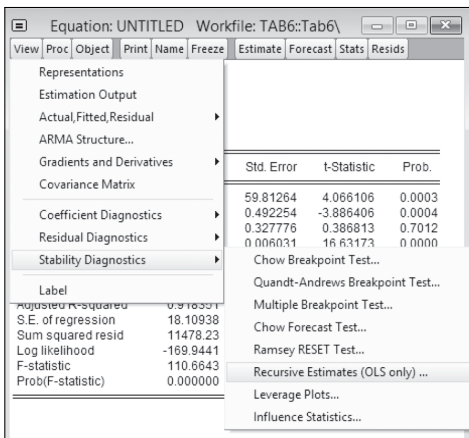


Figura 5.9

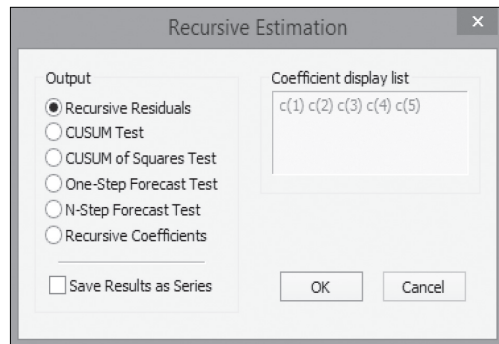


Figura 5.10

El gráfico obtenido (Figura 5.11) muestra la variación de los residuos recursivos. La serie de residuos toma valores positivos y negativos, y que, a primera vista, no parece haber ninguna pauta inestable en su comportamiento, lo que indica en cierto modo, que los residuos recursivos han sido constantes, manteniéndose dentro de los límites (bandas de ± 2 veces su desviación estándar), a excepción del cuarto trimestre de 2011, y el tercer trimestre de 2013, donde los residuos se salen ligeramente de las bandas. En conclusión, los residuos obtenidos parecen comportarse como variables aleatorias incorrelacionadas.

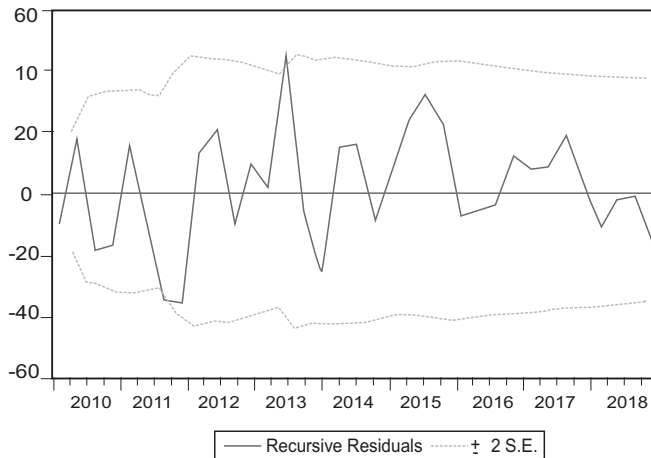


Figura 5.11

5.3.4 Estimación de coeficientes recursivos.

El gráfico de los coeficientes recursivos representa el comportamiento de cada uno de los estimadores al ir añadiendo observaciones a la muestra con la que se realiza la estimación. Las series de coeficientes se muestran junto con sus bandas de confianza (± 2 veces su desviación estándar) lo que permite intuir la presencia de inestabilidad en el modelo si los coeficientes sufren grandes cambios al ir variando la muestra, es decir, si no se mantienen aproximadamente constantes.

EViews realiza el gráfico de los coeficientes recursivos seleccionando en la ventana del objeto ecuación *View* \rightarrow *Stability Diagnostics* \rightarrow *Recursive*

Estimates (OLS only) y damos clic. Posteriormente, en el cuadro de opciones que se abre, elegimos la opción *Recursive Coefficients* y damos OK (Figura 5.12). En la ventana de la derecha del cuadro *Coefficients Display List* podemos escribir la lista de gráficos que queremos ver representados, aunque por defecto EViews incluye todos.

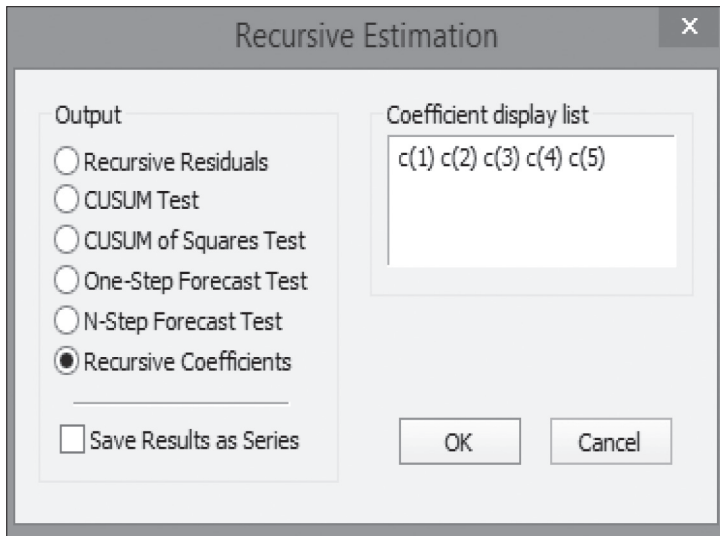


Figura 5.12

La representación gráfica de los coeficientes recursivos para el modelo estimado que aparecen a continuación, nos permite realizar un análisis interesante. El examen de dichos gráficos, en los que es necesario fijarse en la escala de medida del eje de las ordenadas (línea vertical) para apreciar la magnitud de las variaciones, indica que existe constancia de los valores de los estimadores al ir añadiendo nuevas observaciones, lo que permite concluir que el modelo estimado (3.6) es estable.

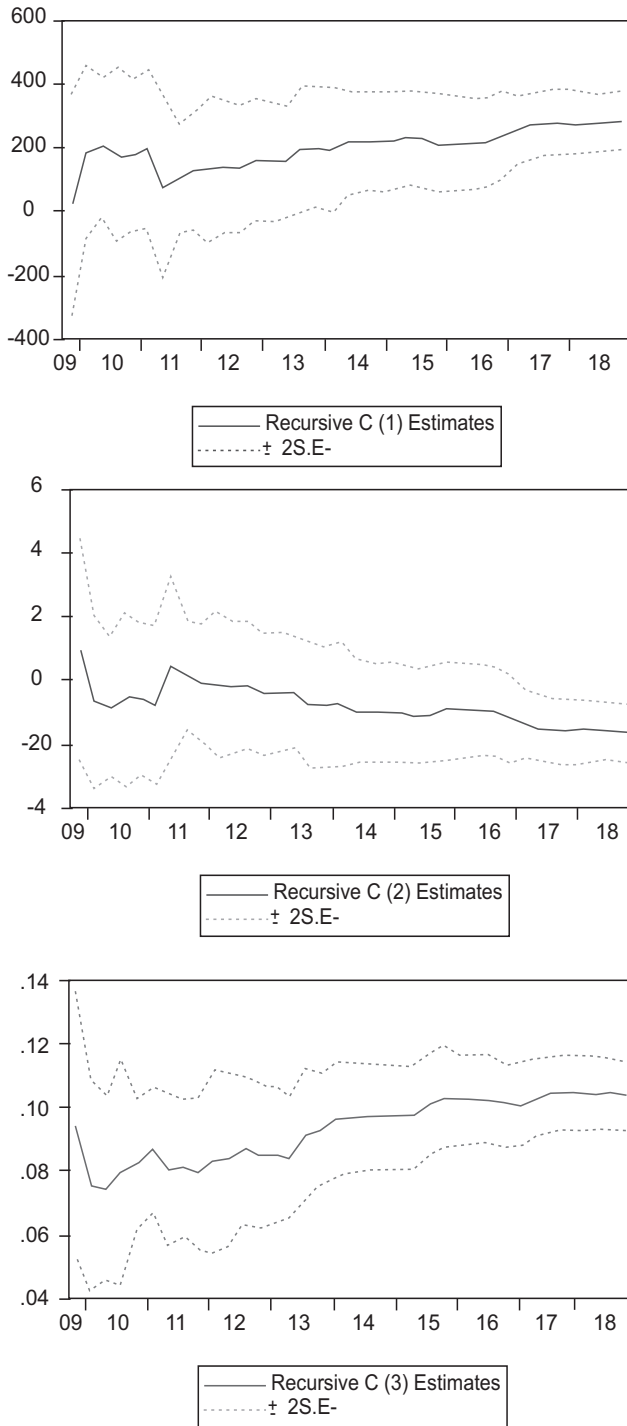


Figura 5.13

5.3.5 El contraste de CUSUM de mala especificación.

Según Alonso et al, (2005), la incorrecta especificación de un modelo afecta a la elección de la parte sistemática del modelo y no a los supuestos sobre el comportamiento de las perturbaciones. En ocasiones un análisis cuidadoso de los residuos puede no ser muy informativo respecto de la idoneidad del modelo, por lo que parece lógico pensar en utilizar instrumentos más potentes que el análisis gráfico. Para superar esta limitación, la econometría ha desarrollado un conjunto de pruebas como el contraste de CUSUM.

El gráfico del valor del estadístico CUSUM se calcula en EViews al 95% de nivel de confianza; este se obtiene seleccionando en la ventana del objeto ecuación *View* → *Stability Diagnostics* → *Recursive Estimates (OLS only)* → *CUSUM* y damos clic como se muestra en la Figura 5.14. En el cuadro de opciones de estimación recursiva (*Recursive Estimation*) elegimos la opción *CUSUM Test* y damos OK (Figura 5.15).

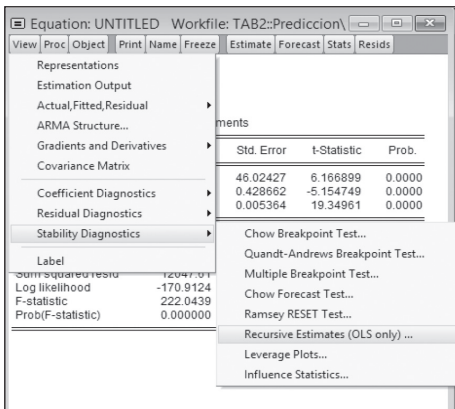


Figura 5.14

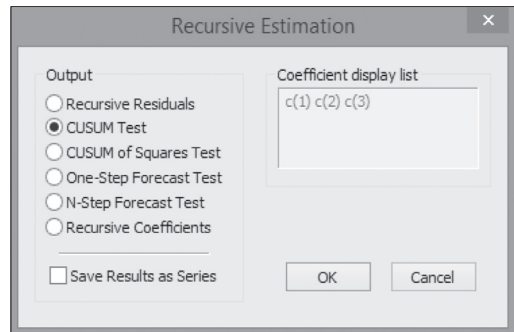


Figura 5.15

A continuación se visualiza el gráfico del estadístico CUSUM, el cual indica que los residuos tienen un comportamiento normal, en ningún momento se salen de los límites de control, es decir, del espacio definido por las bandas con línea punteada (véase Figura 5.16). Por consiguiente, hay evidencia de que no existen problemas de especificación en el modelo (3.6).

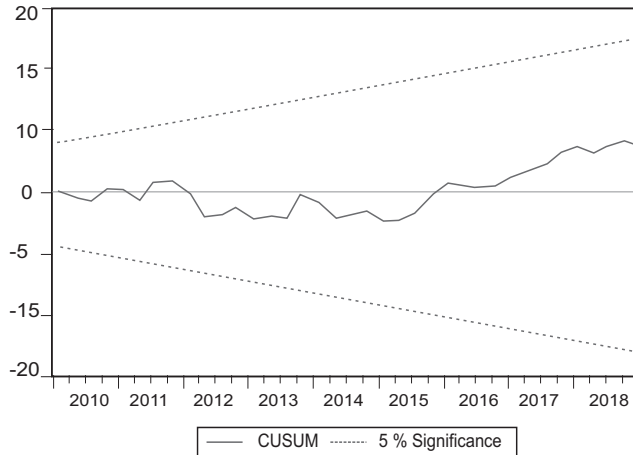


Figura 5.16

Nuevamente seleccionamos *View* → *Stability Diagnostics* → *Recursive Estimates (OLS only)* → *CUSUM* y damos clic. En el cuadro de opciones de estimación recursiva (*Recursive Estimation*) elegimos *CUSUM of Squares Test* y damos OK como se muestra en la Figura 5.17.

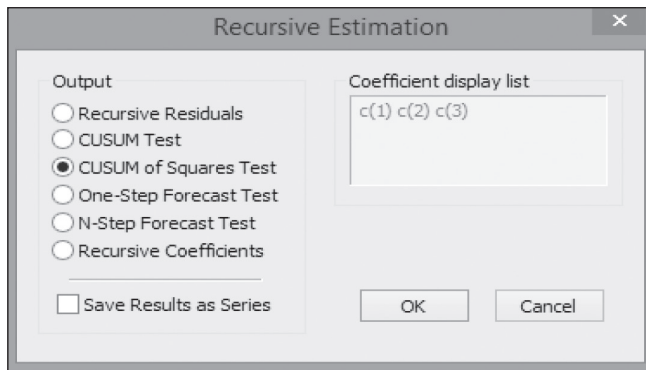


Figura 5.17

El gráfico resultante de esta prueba (véase Figura 5.18) indica que a un nivel de significancia estadística de 5%, los residuos se mueven dentro de los límites definidos por las líneas punteadas, por lo tanto, el modelo no tiene ningún problema de especificación.

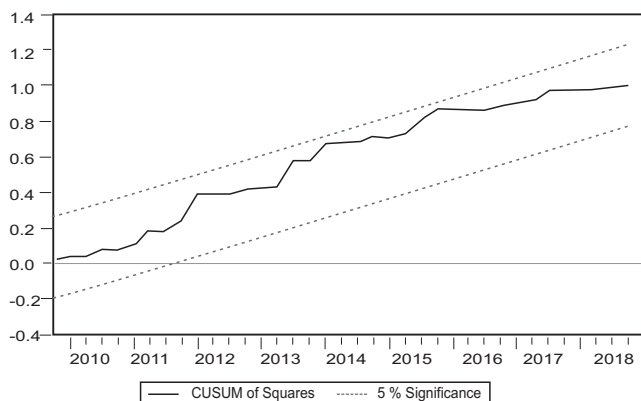


Figura 5.18

5.3.6 Error de especificación de la forma funcional.

Los errores de especificación debidos a la formulación incorrecta de una expresión lineal para el modelo de regresión pueden ser estudiados utilizando el contraste RESET propuesto por Ramsey en 1969. En realidad, se trata de un contraste general para detectar errores de especificación de un modelo que puede utilizarse en particular para analizar un error en la formulación lineal del modelo.

El análisis de la hipótesis de linealidad del modelo se realiza contrastando la restricción de que el coeficiente de la potencia de la variable dependiente estimada es igual a cero. Si la hipótesis nula se rechaza, la conclusión es que la forma funcional lineal del modelo inicialmente especificado es incorrecta y debe incluir la no linealidad detectada.

Para ejecutar el contraste RESET en EViews procedemos a realizar el siguiente procedimiento: en la ventana del objeto ecuación seleccionamos *View* → *Stability Diagnostics* → *Ramsey RESET Test* y damos clic como se muestra en la Figura 5.19. A continuación, en la ventana de diálogo que se

abre, indicamos el número de potencias de la variable endógena ajustada (Number of fitted terms) a incluir empezando por el cuadrado, en nuestro caso por defecto dejamos 1 (se añadirá el cuadrado de la variable dependiente) y damos OK, como se indica en la Figura 5.20.

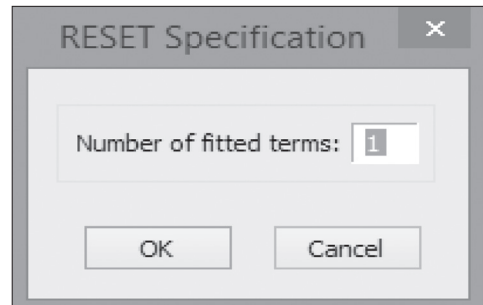
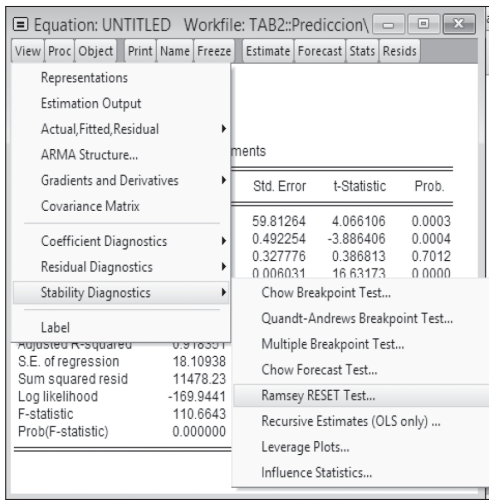


Figura 5.20

Figura 5.19

En la aplicación del contraste RESET formulamos las siguientes hipótesis:

Hipótesis nula	frente	Hipótesis alternativa
H_0 : El modelo es lineal es sus parámetros (el error es ruido blanco).		H_a : Existe alguna variable independiente de orden superior al de la lineal que influye en la variable dependiente.

El resultado de la Tabla 5.5 nos muestra los valores del estadístico F usual y del estadístico χ^2 de razón de verosimilitud junto con la estimación de la ecuación en la que se incorporan las nuevas variables. La probabilidad del estadístico F (*F-statistic*) de 0,2554 es mayor que 0,05. Por ende, el resultado obtenido nos permite no rechazar la hipótesis nula de linealidad a un nivel de significancia de 5%, aportando evidencia de que el modelo (3.6) es lineal es sus parámetros.

Tabla 5.5
Contraste de RESET de Ramsey.

Specification: V C PE PC GP R

Omitted Variables: Squares of fitted values

Test Statistic	Value	df	Probability
t - statistic		34	0,2554
F-statistic		(1, 34)	0,2554
Likelihood ratio		1	0,2140
F-test summary:			
	Sum of Sq.	df	Mean Squares
Test SSR	434,6896	1	434,6896
Restricted SSR	11.478,23	35	327,9495
Unrestricted SSR	11.043,54	34	324,8101
LR test summary:			
	Value		
Restricted LogL	-169,9441		
Unrestricted LogL	-169,172		

Unrestricted Test Equation

Dependent Variable: V

Method: Least Squares

Sample: 2009Q1: 2018Q4

Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	200,1609	70,19775	2,851387	0,0073
PE	-0,977096	0,945852	-1,033033	0,3089
PC	0,075221	0,329235	0,228472	0,8206
GP	0,054509	0,040039	1,361392	0,1823
R	0,000816	0,003242	0,251808	0,8027
FITTED^2	0,000689	0,000596	1,156844	0,2554
R-squared	0,929501	Mean dependent var		344,7425
Adjusted R-squared	0,919133	S.D. dependent var		63,37666
S.E. of regression	18,02249	Akaike info criterion		8,758599
Sum squared resid	11.043,540	Schwarz criterion		9,011931
Log likelihood	-169,172	Hannan-Quinn criterion		8,850196
F-statistic	89,65479	Durbin-Watson stat		1,840823
Prob(F-statistic)	0,000000			

5.3.7 Normalidad de las perturbaciones.

La hipótesis de normalidad de las perturbaciones no es necesaria para la obtención de los estimadores de los coeficientes de regresión por el método de mínimos cuadrados ordinarios, pero es fundamental para la aplicación de otros métodos de estimación y para la realización de inferencia en el modelo. Debido a ello, es imprescindible un análisis del cumplimiento de esta hipótesis al analizar los resultados de la estimación.

Dado que las perturbaciones del modelo no son observables, el estudio de la normalidad se lleva a cabo a partir del análisis de los residuos del modelo. Si la hipótesis se cumple, la distribución empírica de dichos residuos debería presentar características similares a las de la distribución normal. EViews recoge entre sus posibilidades la obtención del histograma de los residuos y el cálculo de ciertos estadísticos descriptivos para dicha serie, entre los que está el estadístico Jarque-Bera que contrasta la normalidad de una variable. En el primer caso, si aproximadamente la distribución de los residuos es una normal, la representación del histograma debería parecer una campana de Gauss simétrica y con un apuntamiento similar al de la distribución normal. En el segundo, como los momentos poblacionales de orden impar de una variable normal son cero (y por lo tanto, su coeficiente de asimetría también) y su coeficiente de kurtosis o apuntamiento es 3, tendríamos que encontrar valores similares a estos al calcular los momentos muestrales de los residuos.

La Figura 5.21 que representa el histograma de los residuos del modelo permite apreciar que éstos se distribuyen como una distribución normal casi simétrica, asemejándose a una campana de Gauss.

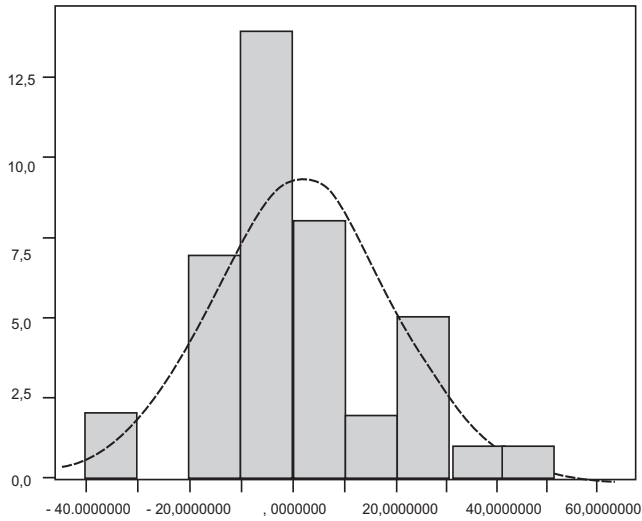


Figura 5.21

El contraste de Jarque-Bera plantea en la hipótesis nula la existencia de normalidad de las perturbaciones y en la alternativa la no normalidad. El estadístico del contraste es el de los multiplicadores de Lagrange, y se construye a partir de los coeficientes de asimetría (*Skewness*) y de kurtosis mencionados arriba.

Hipótesis a contrastar:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : Existencia de normalidad de las perturbaciones del modelo.		\mathcal{H}_a : Existencia de no normalidad de las perturbaciones del modelo.

Para la ejecución de este contraste en EViews, situados en la ventana del objeto ecuación seleccionamos *View* → *Residual Diagnostics* → *Histogram Normality Test* y damos clic como se muestra en la Figura 5.22.

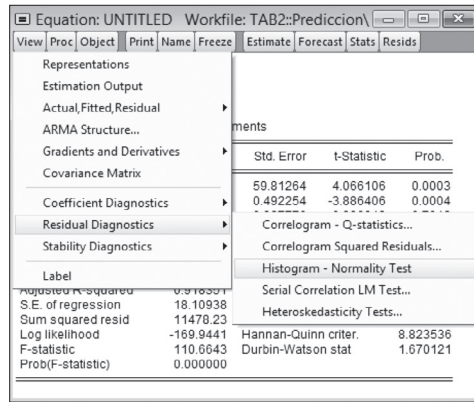


Figura 5.22

El resultado del contraste de normalidad de las perturbaciones se visualiza en la Figura 5.23, el cual se acompaña de una serie de estadísticos como los nombrados arriba. El análisis de los resultados nos permite comprobar que la representación gráfica de los residuos mediante su histograma presenta una serie de observaciones en las colas izquierda y derecha que no concuerda con la escasa probabilidad de que estas zonas se asemejen a una distribución normal. Por otro lado, el valor del coeficiente de asimetría (*Skewness*) de 0,454737 se aproxima a cero y el coeficiente de apuntamiento o Kurtosis muestral es de 3,547986, ligeramente superior a 3. Finalmente, el estadístico de Jarque-Bera de 1,879051 y su probabilidad de 0,390813 es superior al 5% ($0,390813 > 0,05$), lo cual permite No rechazar la hipótesis nula de normalidad, por lo tanto, las perturbaciones del modelo tienen una distribución normal.

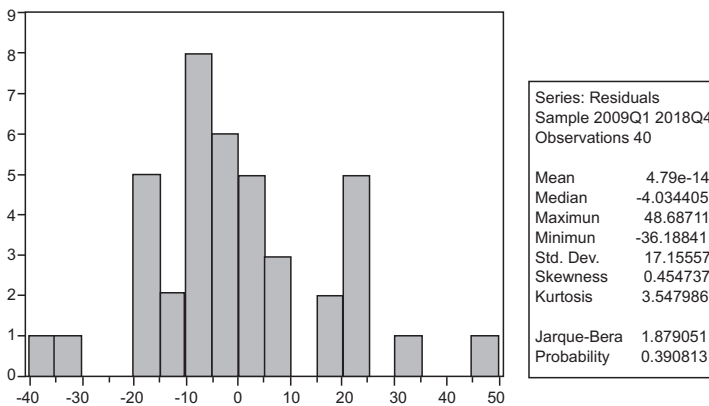


Figura 5.23

Don Pablo se siente bastante sorprendido frente a los resultados de los diferentes contrastes a los que ha sido sometido el modelo de ventas de la editorial para probar su estabilidad. Hasta ahora, el modelo en discusión ha superado las pruebas de diagnóstico, lo que produce cierta satisfacción en Don Pablo, para quien la discusión aún no ha finalizado. Para él, Emilio ha ganado una batalla más, pero no la guerra. Estaremos atentos a lo que suceda en los próximos capítulos.

RETO 4:

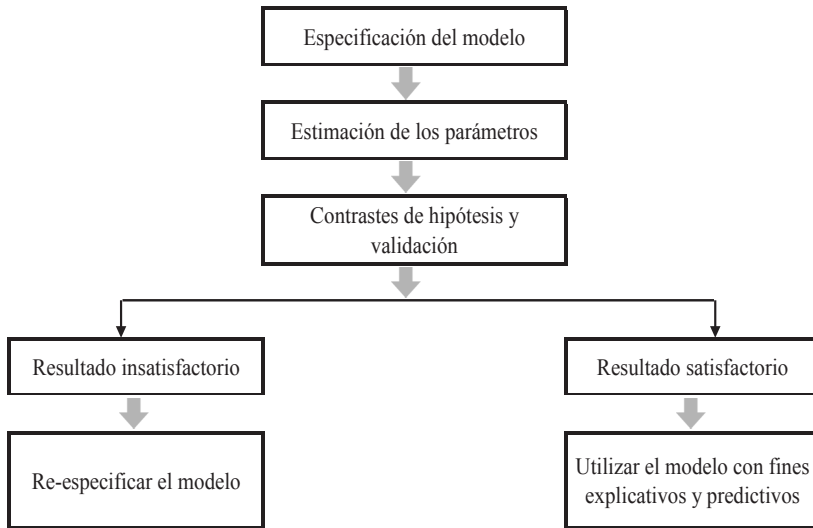
Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

ENCARANDO EL MODELO DE REGRESIÓN LINEAL GENERAL: PREDICCIÓN

Uno de los principales objetivos de la construcción de un modelo econométrico es la realización de predicciones, es decir, el pronóstico del valor que va a tomar la variable dependiente para unos valores dados de las variables explicativas fuera de las observaciones muestrales.²³

La predicción es fiable si el modelo está correctamente especificado, la relación causal establecida en el modelo se mantiene para las observaciones fuera de la muestra y conocemos con exactitud el valor de las variables explicativas en el periodo de predicción (valores conocidos de X). Según el valor que queramos predecir se distinguen dos tipos de pronóstico: a) predicción de un valor medio y b) predicción de un valor individual o predicción por punto. Partimos de la base de que hemos construido un modelo econométrico satisfactorio del acuerdo al desarrollo del siguiente proceso:

²³ En el caso de la editorial, las observaciones muestrales corresponden al periodo 2009:01 a 2018:04, es decir, 40 observaciones. Para la predicción se añadirán observaciones en las columnas de las variables explicativas para el trimestre 2019:01.

Figura 6.1. Proceso de construcción de un modelo econométrico

Fuente: Elaboración propia a partir de Alonso et. al., (2005:135).

A estas alturas, Don Pablo ya se ha resignado a no tener toda la razón en sus planteamientos. En cambio, como es persona curiosa, se preguntará ¿qué puede hacerse con un modelo como el especificado hasta ahora? Por eso, interroga a su hijo Emilio:

Don Pablo: Es evidente que este modelito que te has inventado te ha servido para demostrarme que tenías algo de razón en lo que afirmabas sobre las ventas de la editorial. Pero supongo que no lo habrás hecho sólo para convencerme, ¿no? Me imagino que alguna otra utilidad se le podrá encontrar a tanto tiempo y esfuerzo invertidos.

Emilio: Ya veo que no estás convencido del todo. A mí me parece que a través del modelo hemos aprendido muchas cosas. No sólo sabemos que nuestra conjetura inicial sobre el efecto del precio de la editorial y la publicidad en las ventas es correcta sino que podemos ponerle números a esa influencia. Sabemos también que el precio de la competencia y la renta personal tienen, según la evidencia disponible, una relevancia muy reducida y, sabemos por último, que hemos tenido en cuenta la mayor parte de los factores importantes puesto que el modelo es capaz de explicar aproximadamente el 92% ($R^2 = 0,9184$) de las oscilaciones que han registrado nuestras ventas en los últimos diez años. Acaso, ¿te parece poco?

Don Pablo: No es que me parezca poco, pero yo esperaba más. Al fin y al cabo, todo esto que dices se refiere a lo que ya ha pasado. ¿No puede este modelo ayudarnos a saber qué va a pasar en el futuro?

Emilio: Si me dices a qué precio quieres vender y cuánto dinero te quieres gastar en publicidad, y si me dices qué esperas que pase con el precio de la competencia y la renta personal de nuestros clientes, puedo usar el modelo para decirte cuánto puedes esperar vender.

Don Pablo: Pues mira, a mí me gustaría saber cuánto podemos vender si el precio de la competencia permanece fijo (valor correspondiente al último trimestre de 2018) y, si nosotros, mantenemos fijo nuestro último precio, pero en cambio, disminuimos en un 10% el gasto en publicidad, manteniendo constante la renta personal de nuestros clientes.

Emilio: Muy bien, haré los cálculos respectivos y te lo diré cuál será el valor de las ventas de la editorial en el primer trimestre de 2019.

Por consiguiente, se trata de usar el modelo para llevar a cabo una predicción del valor que se espera alcancen las ventas cuando las otras variables explicativas toman valores específicos. La predicción por punto se obtiene a través de la expresión mientras que puede obtenerse un intervalo de confianza para la predicción a través de

$$Pr(\hat{Y}_p - \hat{\sigma}_e t_{\alpha/2} \leq Y_p \leq \hat{Y}_p + \hat{\sigma}_e t_{\alpha/2}) = 1 - \alpha \quad (6.1)$$

En nuestro caso, los últimos valores conocidos de las variables explicativas correspondientes al cuarto trimestre de 2018 son:

$$PE = 97,7 \quad PC = 113,5 \quad GP = 3.352,0 \quad R = 5.712,1$$

Como Don Pablo está muy interesado en reducir el gasto en publicidad, en al menos, el 10%, por lo tanto, en la predicción de las ventas se utilizará un valor de GP de \$ 3.016,8 (\$3.352 - \$3.352 x 0,10 = \$ 3.016,8). Esto significa que el pronóstico que Don Pablo quiere conocer para el primer trimestre de 2019 se efectúa a partir de los valores de las variables explicativas correspondientes al cuarto trimestre de 2018, así:

$$X'_p = [97,7 \quad 113,5 \quad 3.016,8 \quad 5.712,1]$$

Los valores de las variables explicativas se reemplazan en el modelo obtenido, por ende, la predicción por punto de las ventas de la editorial resulta ser:

$$\hat{Y}_p = 243,20 - 1,91x97,7 + 0,127x113,5 + 0,100x3.016,8 + 0,003x5.712,1 = 388,53$$

Con un intervalo de:

$$IC(Y_p)_{(1-\alpha)} = [\hat{Y}_p \pm \hat{\sigma}_e t_{\alpha/2}]$$

$$IC(Y_p)_{0,95} = [388,53 \pm (19,73)(2,030)]$$

$$= [348,48 \quad 428,58]$$

El valor de 19,73 corresponde al error estándar de la predicción para el primer trimestre de 2019 (2019Q1), como se mostrará más adelante cuando se utilice EViews para el pronóstico. Por su parte, el valor de 2,030 se extrae de la tabla estadística t de Student con 35 grados de libertad y una probabilidad en la cola derecha de 0,025.

Los valores [348,48 428,58] establecen como los límites entre los que se espera que esté el volumen de ventas de la editorial en el caso de que las variables explicativas del modelo tomen los siguientes valores: PE = 97,7; PC = 113,5; GP = 3.016,8 y R = 5.712,1, al tiempo que el valor central 388,53 ($348,48 + 428,58 / 2 = 388,53$), constituye la mejor aproximación puntual al valor que se espera obtener.

Don Pablo: Mira hijo, estoy muy ansioso por conocer el valor proyectado de las ventas de la editorial en el primer trimestre de 2019. Por favor, dílo de una vez.

Emilio: Mira papá, el valor pronosticado de las ventas de la editorial en el primer semestre de 2019 será de \$388,53 millones, con una disminución de los gastos en publicidad del 10% con respecto al cuarto trimestre de 2018, manteniendo las demás variables constantes. Como tú solo vives preocupado

por el comportamiento de los gastos de publicidad, debes tener presente que una reducción del 10% en este tipo de gastos traerá consigo una disminución en las ventas de libros de 4,54%. Papá, ¿acaso, esto es lo que tú quieres?

Don Pablo: Pero si disminuimos el 10% en gastos de publicidad nos ahorraremos un montón de dinero, así caigan las ventas en 4,54%. ¡Creo que esta podría ser una buena decisión!

Emilio: Otra vez, veo que cometes ligerezas en tu análisis financiero. El 10% de ahorro en publicidad equivale a \$0,34 millones, mientras que una caída de 4,54% en las ventas generará una disminución de \$18,5 millones en las ventas. Ahora, ¿si entiendes, cuál es la importancia de un buen modelo econométrico para la toma de decisiones empresariales acertadas?

Don Pablo: Perdóname hijo, reconozco tu capacidad gerencial y tu excelente desempeño técnico. Agradezco tu interés por ayudarme a tomar buenas decisiones.

Ahora bien, el tipo de predicción que calcula EViews es la predicción individual. El procedimiento concreto para realizar predicciones en una ecuación estimada debe iniciarse incrementando el tamaño del rango de las observaciones e incluyendo los datos de las variables explicativas para los periodos en los que queremos predecir el valor de la variable dependiente. Si adicionalmente, también se quiere evaluar la capacidad predictiva del modelo, deben introducirse los valores que toma la variable dependiente para las observaciones fuera de la muestra utilizada en la estimación.

A partir de los datos de la Tabla 3.1, en el Workfile pulsamos *Proc* → *Structure/Resize Current Page* como se muestra en la Figura 6.1. En la casilla *End date* de la ventana que se abre a continuación, reemplazamos 2018Q4 por 2019Q1 y damos OK (véase Figura 6.2).

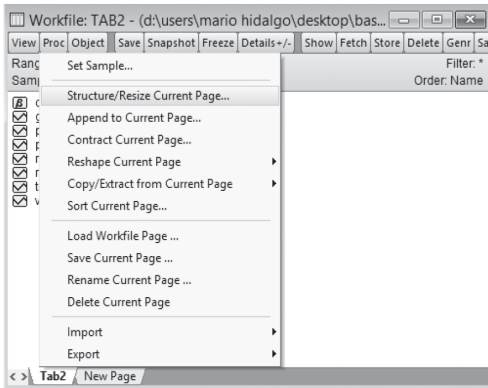


Figura 6.1

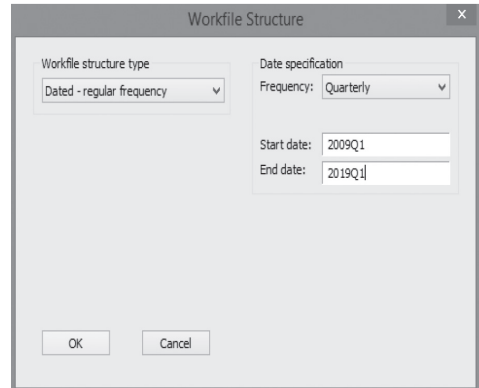


Figura 6.2

Con este procedimiento hemos ampliado el tamaño de la muestra (*Sample*) para el primer trimestre de 2019 y damos *Yes* en continuar como se indica en la Figura 6.3. A partir de esta rutina, se genera una inserción de una nueva observación en la muestra (2019Q1), sin ningún valor hasta el momento (espacio vacío).

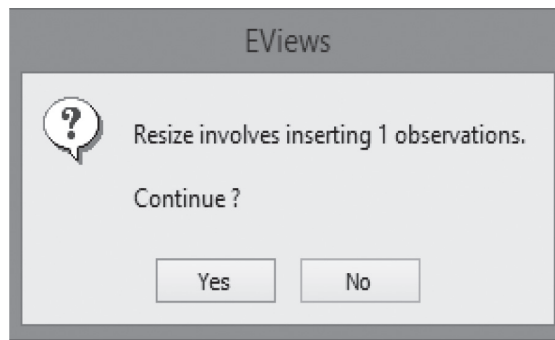


Figura 6.3

La siguiente tarea es introducir los valores correspondientes a las variables PE, PC, GP y R en el primer trimestre de 2019 para la predicción de las ventas en el primer trimestre de 2019. Para ello seleccionamos como grupo las cuatro variables pulsando *Open* → *as Group* en el menú emergente obtenido al hacer clic con el botón derecho del ratón sobre las variables seleccionadas simultáneamente, como se muestra en la Figura 6.4. En seguida, se visualizan las variables seleccionadas con sus datos respectivos (véase Figura 6.5).

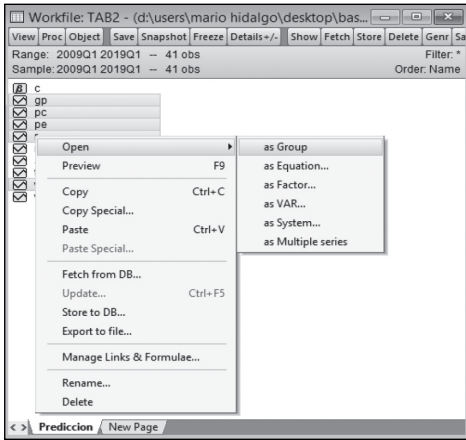


Figura 6.4

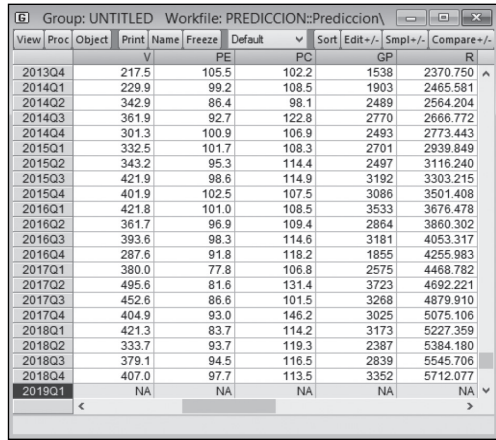


Figura 6.5

A continuación en la pantalla *Group* se pulsa en *Edit* y se introducen los siguientes valores: PE = 97.7; PC = 113.5; GP = 3016.8 y R = 5712.1, teniendo cuidado de escribir punto (no coma) para los decimales (véase Figura 6.6). La tarea siguiente es elegir *Proc* → *Forecast* en la ventana del objeto ecuación como se indica en la Figura 6.7.

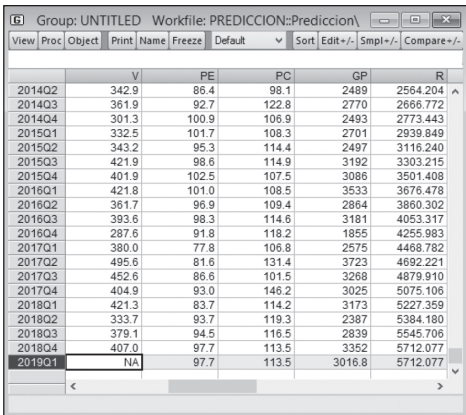


Figura 6.6

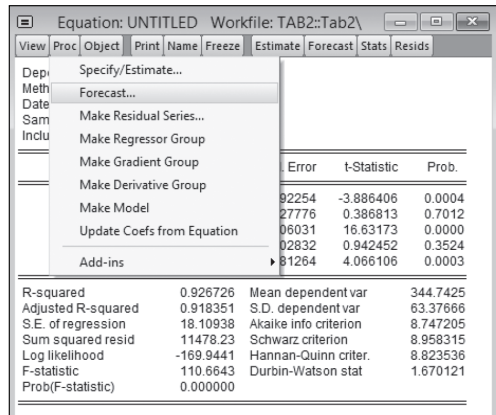


Figura 6.7

En la casilla *S.E. (optional)* escribimos *sef* (el vector *vf* contiene las predicciones y *sef* sus errores estándar) como se muestra en la Figura 6.7 y damos OK.

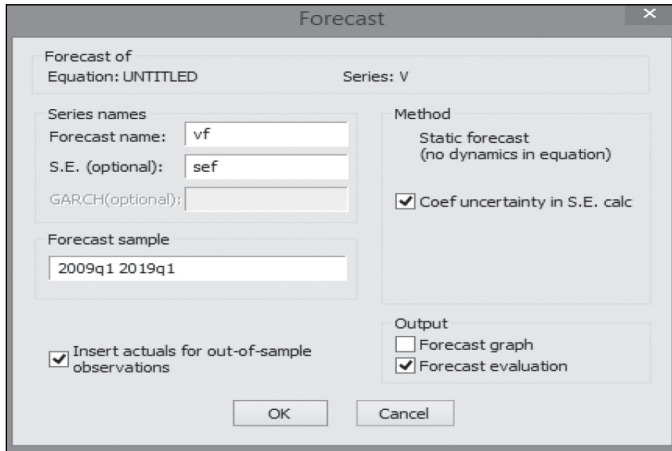


Figura 6.7

En la pantalla del *Workfile* aparecen los vectores *vf* y *sef* como se muestra en la Figura 6.8. Si damos clic en *vf* se abre una ventana que contiene los valores de la variable dependiente, cuyo valor para 2019Q1 es $388,5220 \approx 388,53$ que corresponde a la predicción de las ventas de la editorial, como se indica en la Figura 6.9.

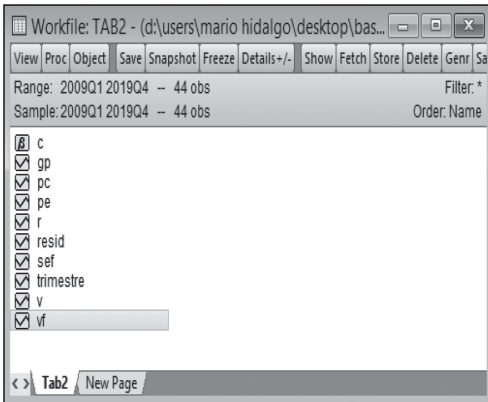


Figura 6.8

Year	Value
2013Q4	214.9259
2014Q1	264.6412
2014Q2	346.8517
2014Q3	366.3900
2014Q4	321.1872
2015Q1	341.1416
2015Q2	334.1577
2015Q3	398.1281
2015Q4	379.6256
2016Q1	427.9252
2016Q2	369.2704
2016Q3	399.5629
2016Q4	279.9924
2017Q1	378.1172
2017Q2	489.7118
2017Q3	431.2180
2017Q4	400.7888
2018Q1	429.7748
2018Q2	332.8701
2018Q3	376.7531
2018Q4	422.1509
2019Q1	388.5290

Figura 6.9

Si hacemos clic sobre el vector *sef* como se muestra en la Figura 6.10 se obtiene los errores estándar de las predicciones, cuyo valor correspondiente al primer trimestre de 2019 es $19,72963 \approx 19,73$ (véase Figura 6.11), el cual se utilizará para calcular enseguida de manera manual, el intervalo de confianza para la predicción individual.

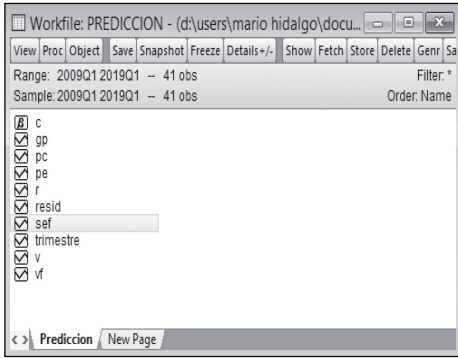


Figura 6.10

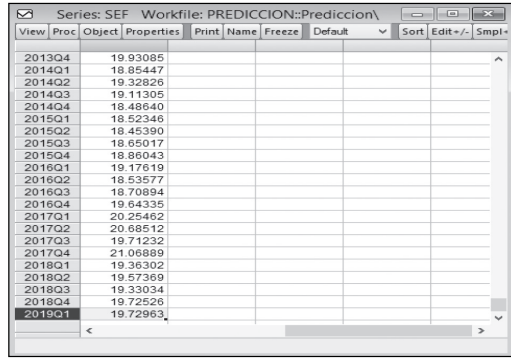


Figura 6.11

6.1 Intervalo de confianza de la predicción.

Como $t_{0,025} = 2,030^{24}$ para $T-K-1 = 40 - 4 - 1$ grados de libertad, un intervalo de confianza de la predicción para una respuesta individual cuando $PE = 97,7$; $PC = 113,5$; $GP = 3.016,8$ y $R = 5.712,1$ vendrá dado por:

$$[388,53 \pm (2,030)(19,73)] = [348,48 \quad 428,58]$$

O expresado de otra manera,

$$\Pr[348,48 \leq Y_p \leq 428,58] = 95\%$$

Don Pablo a regañadientes pide que Emilio le explique el resultado obtenido. Emilio, ante la petición insistente de su padre, exclama: ¡Mira papá!, con un nivel de confianza de 95%, la predicción indica que las ventas de la editorial en el primer trimestre de 2019 estarán en el intervalo comprendido entre \$ 348,48 y \$ 428,58 millones.

6.2 Evaluación de la capacidad predictiva del modelo.

Cuando se le ordena a EViews realizar predicciones también permite calcular diversos estadísticos para analizar la capacidad predictiva del modelo, para ello es necesario que el archivo de datos recoja los valores observados de

²⁴ Este valor se extrae de la tabla estadística t de Student con $gl = 35$ y una probabilidad en la cola derecha = 0,025.

la variable dependiente para el periodo de pronóstico. Si no lo hiciéramos así y no expandiéramos el rango, la predicción recogería los valores estimados de la variable dependiente en el periodo muestral. EViews calcula los siguientes estadísticos: la raíz del error cuadrático medio (*Root Mean Squared Error*), el error absoluto medio (*Mean Absolute Error*), el error absoluto medio del porcentaje de error (*Mean Abs. Percent Error*) y el coeficiente de desigualdad de Theil (*Theil Inequality Coefficient*). Este último coeficiente varía entre 0 y 1, indicando una mejor capacidad predictiva del modelo cuando más se acerque a cero. Todos los estadísticos descritos hasta ahora indican una mejor capacidad predictiva del modelo cuanto más cercanos a cero estén, lo que permite comparar un modelo con otros alternativos y elegir el mejor.

Para evaluar la capacidad predictiva del modelo, sobre la ventana del objeto ecuación seleccionamos *Proc* → *Forecast* y damos clic. En el cuadro de diálogo que se abre, en la casilla *Forecast name* digitamos *vf*, en la casilla *S.E. (optional)* escribimos *sef* y en *output* señalamos la opción *Forecast evaluation* y pulsamos OK (véase Figura 6.12). De este modo, obtenemos el conjunto de estadísticos de la predicción arriba mencionados disponibles en la Tabla 6.1.

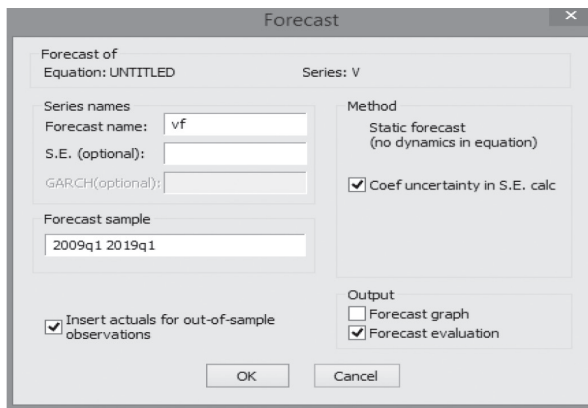


Figura 6.12

Tabla 6.1
Evaluación del pronóstico.

Forecast: VF	
Actual: V	
Forecast sample: 2009Q1 2019Q1	
Included observations: 41	
<hr/>	
Root Mean Squared Error	16,93977
Mean Absolute Error	12,99165
Mean Absolute Percentage Error	4,035701
Theil inequality Coefficient	0,024188
Bias Proportion	0,000000
Variance Proportion	0,019022
Covariance Proportion	0,980978
Theil U2 Coefficient	0,29759
Symmetric MAPE	4,034163
<hr/>	

Aunque en el caso que nos ocupa, no existe un patrón de comparación de nuestro modelo con otros, podríamos sacar algunas conclusiones, por ejemplo, el coeficiente de desigualdad de Theil es bastante cercano a cero (0,024188), por consiguiente, este modelo tiene una buena capacidad predictiva. Para la obtención del gráfico de la predicción, en la ventana del objeto ecuación seleccionamos: *Proc* → *Forecast* y damos clic; en *Output* señalamos la opción *Forecast graph* (Figura 6.13) y damos OK. De esta manera, se obtiene el gráfico de la predicción (Figura 6.14), donde el valor predicho (VF) se mueve en medio de la banda de ± 2 S.E. (± 2 desviaciones típicas, que corresponden al intervalo de confianza de la predicción arriba calculado).

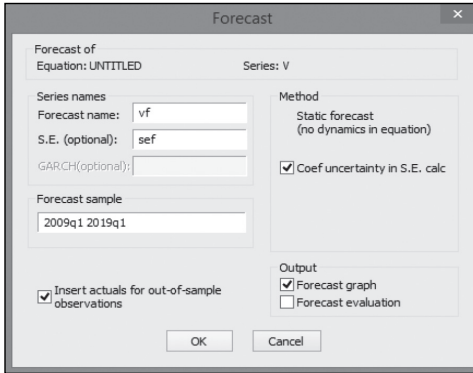


Figura 6.13

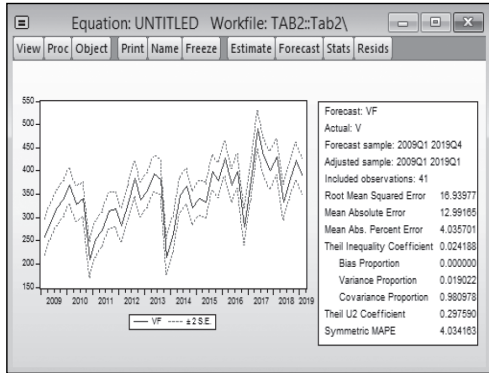


Figura 6.14

6.3 Verificación de la capacidad predictiva del modelo.

La capacidad predictiva del modelo en cuestión puede probarse a través del siguiente procedimiento: En el archivo de trabajo *Tab2* eliminamos la información de la fila correspondiente al trimestre 2018Q4, y posteriormente, cargamos este archivo en EViews con otro nombre. Mediante *Open an existing EViews workfile*, buscamos el archivo en la carpeta correspondiente (*all files *.**) y procedemos a cargarlo y guardarlo en una carpeta de trabajo. Con el puntero del ratón y la tecla CTRL oprimida, seleccionamos las variables del modelo en siguiente orden: *v pe pc gp r* y damos *Open → as Equation*, tal como se muestra en la Figura 6.15. En la ventana de diálogo: *Equation specification* ordenamos las variables, así: *v c pe pc gp r* y damos Aceptar, como se indica en la Figura 6.16.

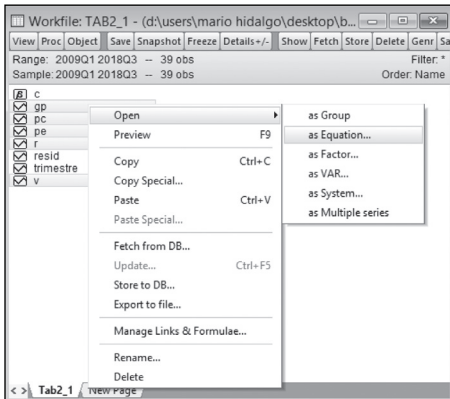


Figura 6.15

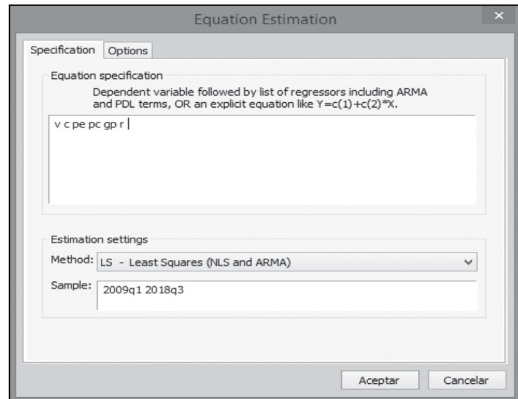


Figura 6.16

En la Tabla 6.2 se muestra ordenadas en columnas las variables en el orden seleccionado, el parámetro o coeficiente de la pendiente de cada una de las variables, incluido el término independiente (*coefficient*); la desviación típica (*Std. Error*), el estadístico t (*t-Statistic*) y los p-valores (*prob*). Todas las variables, excepto PC y R son significativas individualmente al 5% (Prob < 0,05), pero también, las variables en conjunto son significativas al 5%, de acuerdo al estadístico F (*F-statistic*) con una probabilidad menor de 0,05 [Prob (F-statistic) < 0,05].

Tabla 6.2***Estimación del modelo de ventas de la editorial.***

Variable Dependiente: V

Método: MCO

Muestra: 2009Q 2018Q3

Observaciones incluidas: 39

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	234,7545	60,62660	3,872137	0,0005
PE	-1,809687	0,505757	-3,578173	0,0011
PC	0,083393	0,331778	0,251353	0,8031
GP	0,100635	0,006054	16,62337	0,0000
R	0,003655	0,003031	1,205816	0,2362
R-squared	0,926666	Mean dependent var		343,1462
Adjusted R-squared	0,919038	S.D. dependent var		63,38522
S.E. of regression	18,14653	Akaike info criterion		8,754045
Sum squared resid	11.196,09	Schwarz criterion		8,967322
Log likelihood	-165,7039	Hannan-Quinn criterion		8,830567
F-statistic	107,4077	Durbin-Watson stat		1,680743
Prob(F-statistic)	0,000000			

Por consiguiente, el modelo estimado con 39 observaciones (2009Q1... 2018Q3) es el siguiente:

$$\widehat{Ventas} = 234,75 - 1,81PE + 0,083PC + 0,101GP + 0,004R$$

Posteriormente, en la ventana de *Workfile* seleccionamos *Proc* → *Structure/Resize Current Page* y damos clic; enseguida se abre una ventana de diálogo, y específicamente, en la casilla *End date* reemplazamos 2018Q3 por 2018Q4 y damos OK, como se indica en la Figura 6.17. Enseguida, damos Yes en continuar (véase Figura 6.18). Lo que pretendemos aquí es predecir el valor de la variable dependiente “ventas de la editorial” (V) en el cuarto trimestre de 2018 (2018Q4). Como de antemano ya conocemos el valor de las ventas para este trimestre ($V = 407$ millones), así como también el valor que toman las variables explicativas PE, PC, GP y R, lo queremos es predecir el valor de las ventas (valor predicho) y compararlo con el valor observado (valor real), y de este modo, determinar la precisión del pronóstico del modelo.

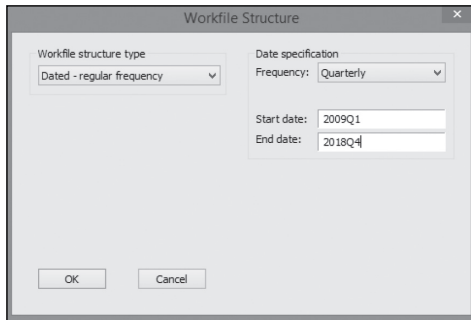


Figura 6.17

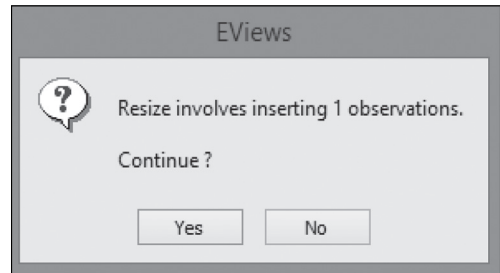


Figura 6.18

Posteriormente, seleccionamos *Open* → *as Group* y damos clic. En la ventana de las variables, más específicamente, en el trimestre 2018Q4 pulsamos *Edit* e insertamos los datos originales de las variables $PE = 97.7$; $PC = 113.5$; $GP = 3.352$ y $R = 5.712,1$ como se muestra en la Figura 6.19. Sobre la ventana del objeto ecuación seleccionamos *Proc* → *Forecast* y damos clic. En la casilla *S.E. (optional)* de la ventana que se abre a continuación, digitamos *sef* que corresponde a los errores estándar de la predicción y damos OK (véase Figura 6.20).

	PE	PC	GP	R
2014Q1	99.2	108.5	1903	2465.581
2014Q2	86.4	98.1	2489	2564.204
2014Q3	92.7	122.8	2770	2666.772
2014Q4	100.9	106.9	2493	2773.443
2015Q1	101.7	108.3	2701	2939.849
2015Q2	95.3	114.4	2497	3116.240
2015Q3	98.6	114.9	3192	3303.215
2015Q4	102.5	107.5	3086	3501.408
2016Q1	101.0	108.5	3533	3676.478
2016Q2	98.9	109.4	2864	3860.302
2016Q3	98.3	114.6	3181	4053.317
2016Q4	91.8	118.2	1855	4255.983
2017Q1	77.8	106.8	2575	4468.782
2017Q2	81.6	131.4	3723	4692.221
2017Q3	86.6	101.5	3268	4879.910
2017Q4	93.0	146.2	3025	5075.106
2018Q1	83.7	114.2	3173	5227.359
2018Q2	93.7	119.3	2387	5384.180
2018Q3	94.5	116.5	2839	5545.706
2018Q4	97.7	113.5	3352	5712.100

Figura 6.19

Figura 6.20

De este modo obtenemos el gráfico y los estadísticos de evaluación de la predicción, como se muestra en la Figura 6.21.

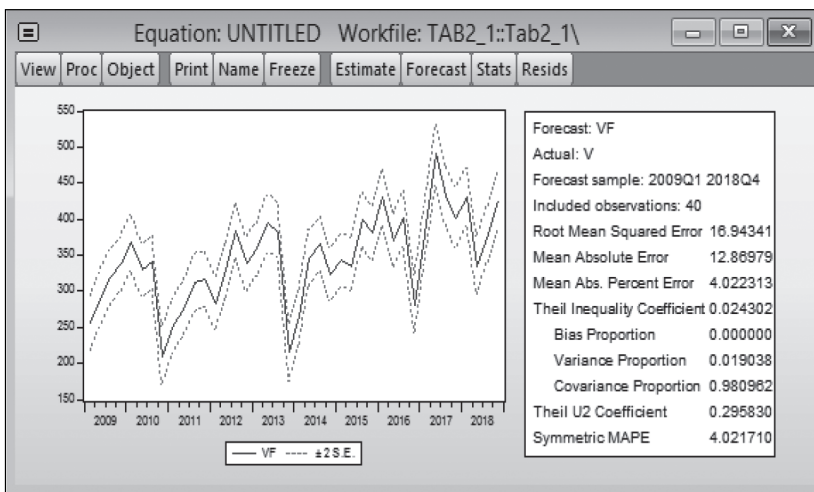


Figura 6.21

En la ventana Workfile damos clic en el vector vf (véase Figura 6.22) y obtenemos los valores de la predicción de la variable dependiente “Ventas” (véase Figura 6.23). Nos interesa uno en especial, el correspondiente al cuarto trimestre de 2018 (2018Q4), cuyo valor es \$391,89 millones. Ahora, comparamos el valor predicho \$391,89 millones con el valor real de las ventas que para este mismo trimestre es de \$407,0 millones. En conclusión, el valor predicho (391,89 está ligeramente por debajo del valor real (407,0).

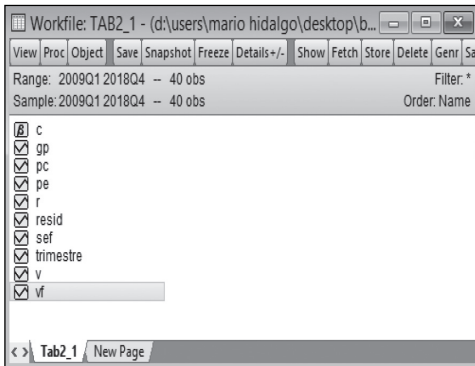


Figura 6.22

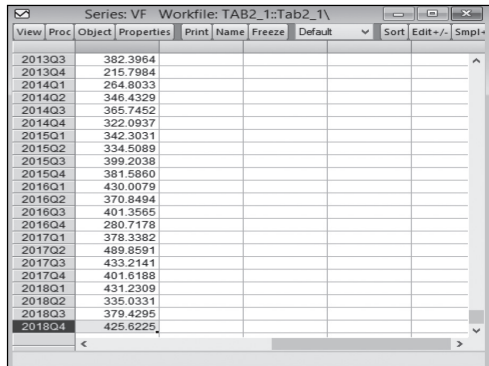


Figura 6.23

Por lo tanto, se obtiene el siguiente resultado:

$$IC(Y_p)_{0,95} = [351,05 \pm (20,12)(2,032)]$$

$$IC(Y_p)_{0,95} = [351,05 \quad 432,73]$$

Los valores $[351,05 \quad 432,73]$ establecen los límites entre los que se espera que esté el volumen de ventas de la editorial en el caso de que las variables explicativas del modelo tomen los siguientes valores: $PE = 97,7$; $PC = 113,5$; $GP = 3.016,8$ y $R = 5.712,1$, al tiempo que el valor central $391,89$ ($351,05 + 432,73/2 = 391,89$), constituye la mejor aproximación puntual al valor que se espera obtener.

Don Pablo por ser un hombre extremadamente pragmático, siente preocupación por el tiempo y esfuerzo dedicado hasta ahora a la construcción del modelo (3.6). Él sabe suficientemente, que el desarrollo de este capítulo tiene como propósito pronosticar el valor de las ventas de la editorial en el primer trimestre de 2019, lo cual permitirá programar con antelación la logística comercial y preparar el presupuesto de la editorial. Emilio ha dado respuesta a la preocupación de su padre, a partir de la predicción de las ventas en el periodo en cuestión. Antes que su padre levante la voz y critique su accionar, él decide anunciar la cifra de ventas esperadas.

Enseguida, toma la palabra y en tono firme exclama: ¡Padre!, el método de predicción utilizado parece indicar con un 95% de confianza, que las ventas

de la editorial en el primer trimestre de 2019 serán de \$391,89 millones, algo inferiores a las observadas en el cuarto trimestre de 2018 (\$407,0 millones). No te preocupes por esta cifra, solo recuerda la teoría de las fluctuaciones de los ciclos económicos: la predicción es sólo una aproximación a un comportamiento futuro, no podemos tomarla como una verdad revelada.

Ante la intervención de Emilio, Don Pablo pregunta: ¿y para qué nos sirve proyectar las ventas? Emilio contesta de manera veloz y contundente: Simplemente, para anticiparnos al futuro mediante acciones en el presente. Papá, ¿acaso no te parece una razón suficiente?

RETO 5:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

**FACTORES EXPLICATIVOS
CUALITATIVOS Y
VARIABLE FICTICIAS**

Prosiguiendo con Alonso et al, (2005), llamamos variables ficticias a las variables, que tomando valores cuantitativos, tratan de representar las diferentes posibles situaciones o casos que se producen en los factores cualitativos de interés. (...) Los valores numéricos que pueden atribuirse son, en principio, completamente arbitrarios y no tienen más objeto que establecer un código que permita distinguir numéricamente cada caso de todos los demás. Por lo tanto, una variable ficticia se define sin más que atribuir un número diferente a cada uno de los casos posibles en el factor que consideremos (p. 170).

Siguiendo con la discusión que mantienen Don Pablo y su hijo Emilio sobre la importancia de las variables incluidas en el modelo para explicar las ventas de la editorial, este último confiesa que se les ha pasado por alto el fenómeno del incremento que experimentan sus ventas en el periodo vacacional. Dicho fenómeno puede asociarse a que en época veraniega se dispone de más tiempo libre para la lectura.

Si dicho fenómeno vacacional es sistemático y significativo y no ha sido recogido en el modelo propuesto, Don Pablo se pregunta: ¿en qué quedan sus discusiones sobre la importancia de cada una de las variables incluidas en el modelo hasta ahora propuesto?.

Además, Don Pablo acaba de recibir, a través de sus clientes, los resultados de una pequeña encuesta, en la que se pidió información a los lectores, sobre el número de horas a la semana dedicado a la lectura, además información sobre el sexo, edad, educación y número de aparatos de televisión, video y computadores en la unidad familiar. Don Pablo desea que Emilio utilice esta información en su próxima campaña publicitaria.

Podríamos considerar si se produce o no un efecto estacional en las ventas de la editorial. El fenómeno estacional de que las ventas se incrementan sistemáticamente en verano por el efecto vacacional puede ser recogido definiendo dos variables ficticias:

$$D_{1t} = \begin{cases} 1 & \text{si } t \in \text{vacación} \Rightarrow \text{si } t \in \text{tercer trimestre} \\ 0 & \text{si } t \in \text{resto del año} \Rightarrow \text{si } t \in \text{trimestres 1, 2 y 4} \end{cases}$$

y su complementaria:

$$D_t^* = \begin{cases} 1 & \text{si } t \in \text{resto del año} \\ 0 & \text{si } t \in \text{vacación} \end{cases}$$

Como solo hay dos casos en el fenómeno estacional que interesa, basta con una sola variable para representarlo en un modelo. Aunque el modelo más sencillo posible, que sólo tiene en cuenta el posible efecto estacional, no es razonable para explicar las ventas de la editorial, vamos a proponerlo aquí, únicamente como primera aproximación a esta hipótesis. Utilizando la información de la Tabla 7.1, el modelo sería:

Tabla 7.1
Horas de lectura semanales.

Trimestre	Ventas (V)	Efecto estacional (L)	Trimestre	Ventas (V)	Efecto estacional (L)
2009:1	275,5	0	2014:1	229,9	0
2009:2	285,6	0	2014:2	342,9	0
2009:3	336,8	1	2014:3	361,9	1
2009:4	333,4	0	2014:4	301,3	0
2010:1	357,0	0	2015:1	332,5	0
2010:2	325,2	0	2015:2	343,2	0
2010:3	362,2	1	2015:3	421,9	1
2010:4	232,0	0	2015:4	401,9	0
2011:1	252,2	0	2016:1	421,8	0
2011:2	322,1	0	2016:2	361,7	0
2011:3	297,3	1	2016:3	393,6	1
2011:4	298,9	0	2016:4	287,6	0
2012:1	246,8	0	2017:1	380,0	0
2012:2	322,4	0	2017:2	495,6	0
2012:3	383,5	1	2017:3	452,6	1
2012:4	321,8	0	2017:4	404,9	0
2013:1	351,5	0	2018:1	421,3	0
2013:2	381,1	0	2018:2	333,7	0
2013:3	412,5	1	2018:3	379,1	1
2013:4	217,5	0	2018:4	407,0	0

Fuente: Alonso et al., (2005).

$$V_t = \beta_0 + \beta_1 L_t + \mu t \quad (7.1)$$

Dónde: V_t = Ventas de la editorial y L_t = Efecto estacional

Mediante MCO y siguiendo el procedimiento aprendido en capítulos anteriores para la estimación de la ecuación de regresión lineal simple, obtenemos los resultados que se muestran en la Tabla 7.2.

Tabla 7.2***Estimación del modelo de ventas contra el efecto estacional.***

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	332,9433	11,07950	30,05040	0,0000
L	47,19667	22,15899	2,129910	0,0397
R-squared	0,106650	Mean dependent var		344,7425
Adjusted R-squared	0,083141	S.D. dependent var		63,37666
S.E. of regression	60,68491	Akaike info criterion		11,09797
Sum squared resid	139.941,0	Schwarz criterion		11,18242
Log likelihood	-219,9595	Hannan-Quinn criterion		11,12851
F-statistic	4,536517	Durbin-Watson stat		0,887397
Prob(F-statistic)	0,039715			

Que, una vez estimado proporciona los siguientes resultados:

$$\widehat{Ventas} = 332,943 + 47,197L_t \quad (7.2)$$

$$(11,08) \quad (22,16)$$

$$R^2 = 10,67\%; \quad \bar{R}^2 = 8,31\%$$

Por lo que la t muestral o t -ratio para el contraste de significación $H_0: \beta_1 = 0$, es:

$$|t| = \frac{47,20}{22,16} = 2,13$$

De otro lado, en las tablas de la distribución t de Student para un nivel de significación de 5% y con $40-1-1 = 38$ grados de libertad encontramos un valor crítico igual a 2,024, por tanto:

$$|t| = 2,13 > 2,024$$

Por lo tanto, al contrastar las hipótesis nula y alternativa formuladas de la siguiente manera:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : No existe un diferencial significativo en las ventas de la editorial cuando es verano con respecto al resto del año.		\mathcal{H}_a : Existe un diferencial significativo en las ventas de la editorial cuando es verano con respecto al resto del año.
$\mathcal{H}_0: \beta_1 = 0$		$\mathcal{H}_a: \beta_1 \neq 0$

Entonces, se rechaza la hipótesis nula al nivel de significancia de 5%. En otras palabras, existe un diferencial significativo en las ventas de la editorial cuando es verano con respecto al resto del año.

En relación con los datos de la encuesta obtenida, Don Pablo desea saber si las mujeres leen más que los hombres y si el hecho de haber cursado estudios universitarios tiene un efecto diferencial importante en el número de horas a la semana que se dedica a la lectura.

Si nos proponemos ahora, realizar el análisis de los datos de la encuesta con la información disponible, sea L_i el número de horas semanales que el individuo i dedica a la lectura, si el sexo del individuo i y la educación E_i del individuo i . Las clases o categorías que se definen a estos factores cualitativos son:

Sexo: H = hombre; M = mujer

Educación: P = estudios primarios; S = estudios secundarios (bachillerato completo) y U = estudios universitarios (pregrado).

Ahora, definimos las siguientes variables ficticias:

Sexo:

$$S_i = \begin{cases} 1 & \text{si el sexo de } i \text{ es hombre} \\ 0 & \text{si el sexo de } i \text{ es mujer} \end{cases}$$

Educación:

$$P_i = \begin{cases} 1 & \text{si el nivel máximo de estudios es primaria} \\ 0 & \text{en caso contrario} \end{cases}$$

$$S_i = \begin{cases} 1 & \text{si el nivel máximo de estudios es bachillerato} \\ 0 & \text{en caso contrario} \end{cases}$$

$$U_i = \begin{cases} 1 & \text{si el nivel máximo de estudios es universitarios} \\ 0 & \text{en caso contrario} \end{cases}$$

Eliminando las categorías H = hombre y P = estudios primarios, se especifica el modelo, así:

$$Y_i = \alpha_0 + \alpha_1 M_i + \alpha_2 S_i + \alpha_3 U_i + \mu_i \quad i = 1, 2, \dots, T \quad (7.3)$$

Para interpretar los coeficientes asociados a las variables ficticias del modelo buscamos, como se explicó en apartados anteriores, el valor medio de la variable dependiente: L = número de horas de lectura semanales. De donde, si restamos horizontalmente, observamos que el efecto diferencial debido al nivel educativo es común para ambos sexos, α_2 para el nivel de bachillerato y α_3 para el universitario y $\alpha_3 - \alpha_2$ para la diferencia entre el universitario y los estudios secundarios. Si estimamos este modelo a partir de los datos que proporciona la encuesta aplicada a los lectores, se obtienen los resultados que se muestran en la Tabla 7.3.

Tabla 7.3***Estimación del modelo con variables cualitativas.***

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9,074951	0,680715	13,33150	0,0000
M	1,957958	0,716643	2,732124	0,0069
S	-0,659454	0,871554	-0,756642	0,4502
U	4,169960	0,864468	4,823731	0,0000
R-squared	0,182394	Mean dependent var		11,20812
Adjusted R-squared	0,169685	S.D. dependent var		5,509948
S.E. of regression	5,020755	Akaike info criterion		6,085133
Sum squared resid	4.865,140	Schwarz criterion		6,151797
Log likelihood	-595,3856	Hannan-Quinn criterion		6,112119
F-statistic	14,35163	Durbin-Watson stat		2,082362
Prob(F-statistic)	0,000000			

$$\widehat{lectura} = 9,075 + 1,958M_i - 0,659S_i + 4,170U_i \quad (7.4)$$

$$(0,680) \quad (0,717) \quad (0,872) \quad (0,864)$$

$$R^2 = 18,24\%; \quad \bar{R}^2 = 16,97\%$$

Los resultados indican que tanto el sexo como el nivel de estudios tienen influencia en el número de horas de lectura semanales. No obstante, es preciso llevar a cabo los contrastes adecuados para poder afirmar lo anterior con mayor seguridad. Por ello, consideremos en primer lugar los contrastes individuales:

En cuanto al sexo:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_{01} : La condición de ser mujer no influye en el tiempo dedicado a la lectura.	frente	\mathcal{H}_{a1} : La condición de ser mujer influye en el tiempo dedicado a la lectura.
$\mathcal{H}_{01}: \alpha_1 = 0$		$\mathcal{H}_{a1}: \alpha_1 \neq 0$

En cuanto al nivel de estudios:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_{02} : El nivel de estudios secundarios no influye en el tiempo dedicado a la lectura.	frente	\mathcal{H}_{a2} : El nivel de estudios secundarios influye en el tiempo dedicado a la lectura.
$\mathcal{H}_{02}: \alpha_2 = 0$		$\mathcal{H}_{a2}: \alpha_2 \neq 0$

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_{03} : El nivel de estudios universitarios no influye en el tiempo dedicado a la lectura.	frente	\mathcal{H}_{a3} : El nivel de estudios universitarios influye en el tiempo dedicado a la lectura.
$\mathcal{H}_{03}: \alpha_3 = 0$		$\mathcal{H}_{a3}: \alpha_3 \neq 0$

Puesto que los estadísticos t toman los siguientes valores:

$$t_M = \frac{1,958}{0,717} = 2,73; \quad t_S = \frac{0,659}{0,872} = 0,76 \quad \text{y} \quad t_U = \frac{4,170}{0,864} = 4,83$$

Por consiguiente, se toman las siguientes decisiones:

$|t_M| = 2,73 > 1,97 \rightarrow$ Se rechaza la hipótesis nula, por consiguiente, la condición de ser mujer influye en el tiempo dedicado a la lectura.

$|t_S| = 0,76 < 1,97 \rightarrow$ No se rechaza la hipótesis nula, por ende, el nivel de estudios secundarios no influye en el tiempo dedicado a la lectura.

$|t_U| = 4,83 > 1,97 \rightarrow$ Se rechaza la hipótesis nula, por lo tanto, el nivel de estudios universitarios influye en el tiempo dedicado a la lectura.

En conclusión, dado que en la distribución encontramos el valor $t_{0,025}$ con 193 grados de libertad ($T-K-1 = 197-3-1 = 193$) que es igual a 1,97, rechazamos la primera hipótesis nula, es decir, en cuanto al sexo, se observa mayor dedicación a la lectura por parte de las mujeres. No se rechaza la segunda hipótesis nula, es decir, el nivel de estudios secundarios no presenta diferencias respecto al nivel de estudios primarios en lo que se refiere a las horas de lectura (el nivel de estudios primarios y secundarios no influyen en

la lectura); y se rechaza la tercera hipótesis nula, por consiguiente, en cuanto a la formación educativa, los universitarios leen más que los que no lo son.

Como a Don Pablo le interesa conocer un modelo más completo que incorpore, además de las variables cuantitativas ya estudiadas (precio de la editorial, precio de la competencia, gasto en publicidad y renta personal), el efecto estacional, ya estudiado en forma aislada, procederemos a estimar dicho modelo. En la Tabla 7.4 se muestra los resultados de la aplicación de MCO.

Tabla 7.4

Estimación modelo de ventas con variables cuantitativas y cualitativas.

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	230,7539	57,89491	3,989183	0,0003
PE	-1,796590	0,476892	-3,767286	0,0006
PC	0,178973	0,316206	0,566001	0,5751
GP	0,096767	0,006070	15,94293	0,0000
R	0,003237	0,002738	1,182183	0,2453
L	13,19365	6,702739	1,968396	0,0572
R-squared	0,934222	Mean dependent var		344,7425
Adjusted R-squared	0,924548	S.D. dependent var		63,37666
S.E. of regression	17,40859	Akaike info criterion		8,689286
Sum squared resid	10.304,00	Schwarz criterion		8,942617
Log likelihood	-167,7857	Hannan-Quinn criterion		8,780882
F-statistic	96,57753	Durbin-Watson stat		1,682572
Prob(F-statistic)	0,000000			

Por lo tanto, el modelo sería:

$$\widehat{Ventas} = 230,75 - 1,797PE + 0,179PC + 0,097GP + 0,003R + 13,194L \quad (7.5)$$

$$(57,84) \quad (0,477) \quad (0,316) \quad (0,006) \quad (0,003) \quad (6,703)$$

$$R^2 = 93,42\%; \quad \bar{R}^2 = 92,45\%$$

¿Es el efecto vacacional significativo?, ¿tiene razón Emilio cuando afirma que en verano de cada año (tercer trimestre: julio a agosto) se incrementan sistemáticamente las ventas de la editorial?

A partir de la información disponible en la Tabla 7.5 daremos respuesta al interrogante de Emilio. La hipótesis nula que debemos contrastar es:

Hipótesis nula		Hipótesis alternativa
\mathcal{H}_0 : El efecto vacacional no es significativo en las ventas de la editorial.	frente	\mathcal{H}_a : El efecto vacacional es significativo en las ventas de la editorial.
$\mathcal{H}_0: \alpha_4 = 0$		$\mathcal{H}_a: \alpha_4 \neq 0$

La distribución apropiada es la t de Student, y el valor del estadístico viene dado por:

$$t = \frac{13,194}{6,703} = 1,968$$

Tabla 7.5***Variables cuantitativas y cualitativas (efecto estacional).***

Trimestre	Ventas	PE	PC	GP	R	L	Trimestre	Ventas	PE	PC	GP	R	L
2009:1	275,5	98,6	105,5	1.868,0	895,0	0	2014:1	229,9	99,2	108,5	1.903,0	2.465,6	0
2009:2	285,6	96,6	104,4	2.157,0	957,7	0	2014:2	342,9	86,4	98,1	2.489,0	2.564,2	0
2009:3	336,8	102,0	115,6	2.541,0	1.024,7	1	2014:3	361,9	92,7	122,8	2.770,0	2.666,8	1
2009:4	333,4	91,9	102,7	2.561,0	1.075,9	0	2014:4	301,3	100,9	106,9	2.493,0	2.773,4	0
2010:1	357,0	105,6	106,1	3.103,0	1.129,7	0	2015:1	332,5	101,7	108,3	2.701,0	2.939,8	0
2010:2	325,2	102,8	100,2	2.661,0	1.186,2	0	2015:2	343,2	95,3	114,4	2.497,0	3.116,2	0
2010:3	362,2	102,1	97,2	2.757,0	1.245,5	1	2015:3	421,9	98,6	114,9	3.192,0	3.303,2	1
2010:4	232,0	99,1	93,3	1.403,0	1.307,8	0	2015:4	401,9	102,5	107,5	3.086,0	3.501,4	0
2011:1	252,2	100,3	94,6	1.856,0	1.386,3	0	2016:1	421,8	101,0	108,5	3.533,0	3.676,5	0
2011:2	322,1	104,8	109,4	2.123,0	1.455,6	0	2016:2	361,7	96,9	109,4	2.864,0	3.860,3	0
2011:3	297,3	85,6	94,7	2.181,0	1.528,3	1	2016:3	393,6	98,3	114,6	3.181,0	4.053,3	1
2011:4	298,9	103,0	111,4	2.520,0	1.604,8	0	2016:4	287,6	91,8	118,2	1.855,0	4.256,0	0
2012:1	246,8	100,4	105,0	2.134,0	1.685,0	0	2017:1	380,0	77,8	106,8	2.575,0	4.468,8	0
2012:2	322,4	93,8	118,3	2.473,0	1.786,1	0	2017:2	495,6	81,6	131,4	3.723,0	4.692,2	0
2012:3	383,5	99,7	102,1	3.125,0	1.875,4	1	2017:3	452,6	86,6	101,5	3.268,0	4.879,9	1
2012:4	321,8	104,6	110,6	2.753,0	1.969,2	0	2017:4	404,9	93,0	146,2	3.025,0	5.075,1	0
2013:1	351,5	100,0	97,3	2.869,0	2.067,6	0	2018:1	421,3	83,7	114,2	3.173,0	5.227,4	0
2013:2	381,1	103,7	91,8	3.301,0	2.171,0	0	2018:2	333,7	93,7	119,3	2.387,0	5.384,2	0
2013:3	412,5	96,7	96,8	3.043,0	2.279,6	1	2018:3	379,1	94,5	116,5	2.839,0	5.545,7	1
2013:4	217,5	105,5	102,2	1.538,0	2.370,8	0	2018:4	407,0	97,7	113,5	3.352,0	5.712,1	0

PE: Precio promedio de la editorial; PC: Precio de la competencia; GP: Gasto en publicidad; R: Renta personal; L: Efecto estacional (verano)

Fuente: Elaboración propia a partir de Alonso et al. (2005).

De otra parte, en las tablas de la distribución t de Student, para un nivel de significación de 5% ($t_{0,025}$) para 34 grados de libertad ($T-K-1 = 40 - 5 - 1 = 34$), encontramos un valor crítico igual a 2,032, por consiguiente, $|t| = 1,968 < 2,032 \rightarrow$ No se rechaza la hipótesis nula, por tanto, el efecto vacacional no es significativo para explicar las ventas de la editorial, en el verano de cada año no se incrementan sistemáticamente las ventas de la editorial.

En conclusión, no se rechaza la hipótesis nula $H_0: \alpha_4 = 0$ al nivel de significación del 5%, es decir, no hay evidencia de que sistemáticamente las ventas de la editorial se incrementen en verano, es decir, no existe un efecto vacacional. Las conclusiones hasta ahora extraídas por Don Pablo y Emilio son válidas, al parecer no se ha omitido ningún factor relevante en el modelo.

Emilio quien controla los gastos en publicidad, cree haber notado que una unidad monetaria más gastada en publicidad en periodo vacacional, no tiene el mismo efecto medio en las ventas que la misma unidad monetaria gastada en cualquier otra época del año. Si dicho diferencial fuera relevante, debería ser incluido en el modelo con el que se analizan las ventas de la editorial. Para realizar este análisis, primero creamos la variable $L_t GP_t$, la cual resulta de la multiplicación de las variables efecto estacional (L_t) por el gasto en publicidad efectuado por la editorial en el tercer trimestre de cada año (época de verano) denotado por GP_t . Esto tiene como propósito determinar, si en verdad, existe un efecto diferencial del gasto en publicidad en el periodo vacacional, como sospecha Emilio.

A partir de la información de la Tabla 7.6 se re-estima el modelo por MCO, cuyo resultado se muestra en la Tabla 7.7.

Tabla 7.6***Variables cuantitativas y cualitativas (efecto estacional y gasto en publicidad en periodo vacacional).***

Trimestre	Ventas	PE	PC	GP	R	L	LGP	Trimestre	Ventas	PE	PC	GP	R	L	LGP
2009:1	275,5	98,6	105,5	1.868,0	895,0	0	0,0	2014:1	229,9	99,2	108,5	1.903,0	2.465,6	0	0,0
2009:2	285,6	96,6	104,4	2.157,0	957,7	0	0,0	2014:2	342,9	86,4	98,1	2.489,0	2.564,2	0	0,0
2009:3	336,8	102,0	115,6	2.541,0	1.024,7	1	2.541,0	2014:3	361,9	92,7	122,8	2.770,0	2.666,8	1	2.770,0
2009:4	333,4	91,9	102,7	2.561,0	1.075,9	0	0,0	2014:4	301,3	100,9	106,9	2.493,0	2.773,4	0	0,0
2010:1	357,0	105,6	106,1	3.103,0	1.129,7	0	0,0	2015:1	332,5	101,7	108,3	2.701,0	2.939,8	0	0,0
2010:2	325,2	102,8	100,2	2.661,0	1.186,2	0	0,0	2015:2	343,2	95,3	114,4	2.497,0	3.116,2	0	0,0
2010:3	362,2	102,1	97,2	2.757,0	1.245,5	1	2.757,0	2015:3	421,9	98,6	114,9	3.192,0	3.303,2	1	3.192,0
2010:4	232,0	99,1	93,3	1.403,0	1.307,8	0	0,0	2015:4	401,9	102,5	107,5	3.086,0	3.501,4	0	0,0
2011:1	252,2	100,3	94,6	1.856,0	1.386,3	0	0,0	2016:1	421,8	101,0	108,5	3.533,0	3.676,5	0	0,0
2011:2	322,1	104,8	109,4	2.123,0	1.455,6	0	0,0	2016:2	361,7	96,9	109,4	2.864,0	3.860,3	0	0,0
2011:3	297,3	85,6	94,7	2.181,0	1.528,3	1	2.181,0	2016:3	393,6	98,3	114,6	3.181,0	4.053,3	1	3.181,0
2011:4	298,9	103,0	111,4	2.520,0	1.604,8	0	0,0	2016:4	287,6	91,8	118,2	1.855,0	4.256,0	0	0,0
2012:1	246,8	100,4	105,0	2.134,0	1.685,0	0	0,0	2017:1	380,0	77,8	106,8	2.575,0	4.468,8	0	0,0
2012:2	322,4	93,8	118,3	2.473,0	1.786,1	0	0,0	2017:2	495,6	81,6	131,4	3.723,0	4.692,2	0	0,0
2012:3	383,5	99,7	102,1	3.125,0	1.875,4	1	3.125,0	2017:3	452,6	86,6	101,5	3.268,0	4.879,9	1	3.268,0
2012:4	321,8	104,6	110,6	2.753,0	1.969,2	0	0,0	2017:4	404,9	93,0	146,2	3.025,0	5.075,1	0	0,0
2013:1	351,5	100,0	97,3	2.869,0	2.067,6	0	0,0	2018:1	421,3	83,7	114,2	3.173,0	5.227,4	0	0,0
2013:2	381,1	103,7	91,8	3.301,0	2.171,0	0	0,0	2018:2	333,7	93,7	119,3	2.387,0	5.384,2	0	0,0
2013:3	412,5	96,7	96,8	3.043,0	2.279,6	1	3.043,0	2018:3	379,1	94,5	116,5	2.839,0	5.545,7	1	2.839,0
2013:4	217,5	105,5	102,2	1.538,0	2.370,8	0	0,0	2018:4	407,0	97,7	113,5	3.352,0	5.712,1	0	0,0

PE: Precio promedio de la editorial; PC: Precio de la competencia; GP: Gasto en publicidad; R: Renta personal; L: Efecto estacional (verano); LGP: Gasto en publicidad efectuado en verano.

Fuente: Elaboración propia a partir de Alonso et al. (2005).

Tabla 7.7***Estimación modelo de ventas con efecto estacional y gasto en publicidad en periodo vacacional.***

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	249,5446	58,24393	4,284475	0,0001
PE	-1,978222	0,484418	-4,083712	0,0003
PC	0,240958	0,313563	0,768455	0,4477
GP	0,094968	0,006087	15,60148	0,0000
R	0,002048	0,002807	0,729632	0,4708
L	-65,46654	53,43266	-1,225216	0,2292
LGP	0,027389	0,018463	1,483454	0,1474
R-squared	0,938334	Mean dependent var		344,7425
Adjusted R-squared	0,927122	S.D. dependent var		63,37666
S.E. of regression	17,10912	Akaike info criterion		8,674729
Sum squared resid	9.659,830	Schwarz criterion		8,970283
Log likelihood	-166,4946	Hannan-Quinn criterion		8,781592
F-statistic	83,69007	Durbin-Watson stat		1,621307
Prob(F-statistic)	0,000000			

Con la incorporación de esta nueva variable en el modelo, se obtiene:

$$\widehat{Ventas} = 249,54 - 1,978PE + 0,241PC + 0,095GP + 0,002R - 65,467L + 0,027LGP \quad (7.6)$$

$$(58,24) \quad (0,484) \quad (0,314) \quad (0,006) \quad (0,002) \quad (53,433) \quad (0,018)$$

$$R^2 = 93,83\%; \quad \bar{R}^2 = 92,71\% \quad \Sigma \hat{\mu}^2 = 9.659,830$$

¿Qué podemos concluir sobre el fenómeno vacacional? El efecto de las vacaciones en la lectura se recoge en los coeficientes α_5 que acompaña a la variable L_t y α_6 que acompaña a la variable L_tGP_t . La hipótesis nula de interés para el contraste de significación individual de cada uno de estos coeficientes, es:

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : El efecto vacacional no es significativo en las ventas de la editorial		\mathcal{H}_a : El efecto vacacional es significativo en las ventas de la editorial.
$\mathcal{H}_0: \alpha_5 = 0$		$\mathcal{H}_0: \alpha_5 \neq 0$

Hipótesis nula	frente	Hipótesis alternativa
\mathcal{H}_0 : Una unidad monetaria más gastada en publicidad no tiene el mismo efecto en el periodo vacacional que en el resto del año.		\mathcal{H}_a : Una unidad monetaria más gastada en publicidad tiene el mismo efecto en el periodo vacacional que en el resto del año.
$\mathcal{H}_0: \alpha_6 = 0$		$\mathcal{H}_0: \alpha_6 \neq 0$

Los valores del estadístico t vienen dados por:

$$|t| = \frac{65,467}{53,433} = 1,23 \quad \text{y} \quad |t| = \frac{0,027}{0,018} = 1,50$$

La distribución apropiada es en ambos casos una t y en las tablas de la distribución de t de Student para un nivel de significación de 5% y para 40 grados de libertad ($T-K-1 = 40-6-1 = 33$), encontramos un valor crítico igual a 2,035, por lo tanto:

$$|t| = 1,23 < 2,035 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t| = 1,50 < 2,035 \rightarrow \text{No se rechaza la hipótesis nula}$$

Puesto que ninguno de los estadísticos supera el valor crítico, la decisión es no rechazar ninguna de las dos hipótesis nulas al nivel de significación de 5%; el efecto estacional sobre las ventas de la editorial parece ser inexistente tanto en la ordenada como en la pendiente de los gastos en publicidad.

Un contraste conjunto de ambos coeficientes $\mathcal{H}_0: \alpha_5 = \alpha_6 = 0$ se lleva a cabo con el estadístico F a través de las diferencias de sumas de cuadrados de residuos. El modelo no restringido es el que acabamos de estimar y el modelo restringido [modelo (3.6) estimado en el capítulo 3] es:

$$\widehat{Ventas} = 243,205 - 1,913PE + 0,127PC + 0,100GP + 0,003R$$

$$(59,81) \quad (0,49) \quad (0,33) \quad (0,006) \quad (0,003)$$

$$R^2 = 92,67\%; \quad \bar{R}^2 = 91,84\%; \quad \Sigma\mu^2 = 11.478,23$$

El valor muestral de F es:

$$\mathcal{F} = \frac{(11.478,23 - 9.659,83)/2}{9.659,83/(40-7)} = 3,11$$

Al nivel de significación de 5%, el valor crítico que se obtiene de la distribución de Snedecor $(2, 33) = 3,28$, por lo tanto, $3,11 < 3,28$; por lo que se no se rechaza la hipótesis de que conjuntamente no existe efecto diferencial en la ordenada y en la pendiente de los gastos de publicidad. Como conclusión, no es conveniente mantener las variables L_t y LGP_t conjuntamente en el modelo, si queremos estimar el efecto estacional con alguna precisión.

De otro lado, hemos comprobado que la variable de nivel L_t , no resulta significativa cuando se encuentra sola en el modelo. Podemos proceder a eliminarla y re-estimar el modelo, cuyo resultado obtenido se muestra en la Tabla 7.8.

Tabla 7.8***Estimación modelo de ventas con gasto en publicidad en periodo vacacional.***

Variable Dependiente: V

Método: MCO

Muestra: 20009Q 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	233,2044	57,11279	4,083226	0,0003
PE	-1,820566	0,470444	-3,869888	0,0005
PC	0,194089	0,313506	0,619093	0,5400
GP	0,096176	0,006051	15,89433	0,0000
R	0,003065	0,002702	1,134507	0,2645
LGP	0,004940	0,002293	2,154637	0,0384
R-squared	0,935529	Mean dependent var		344,7425
Adjusted R-squared	0,926048	S.D. dependent var		63,37666
S.E. of regression	17,23476	Akaike info criterion		8,669214
Sum squared resid	10.099,250	Schwarz criterion		8,922546
Log likelihood	-167,3843	Hannan-Quinn criterion		8,760811
F-statistic	98,67342	Durbin-Watson stat		1,677145
Prob(F-statistic)	0,000000			

Por lo tanto, el modelo estimado es el siguiente:

$$\widehat{Ventas}_t = 233,204 - 1,821PE_t + 0,194PC_t + 0,096GP_t + 0,003R_t + 0,0049LGP_t \quad (7.7)$$

$$(57,11) \quad (0,47) \quad (0,31) \quad (0,006) \quad (0,003) \quad (0,002)$$

$$R^2 = 93,55\%; \quad \bar{R}^2 = 92,60\%; \quad \Sigma\mu^2 = 10.099,25$$

Esto significa que, individualmente, la variable de interacción LGP_t resulta significativa, ya que el estadístico para el contraste de significación toma el valor $\frac{0,0049}{0,002} = 2,45$ el cual es mayor que el valor crítico de la tabla t de Student con 34 grados de libertad ($T-K-1 = 40-5-1 = 34$) que es 2,032 ($2,45 > 2,032$), por ende, se rechaza la hipótesis nula. La conclusión razonable parece ser la de incluir en el modelo el efecto estacional a través de la variable LGP_t , evitando así sesgar las estimaciones y obteniendo al mismo tiempo una estimación precisa de los coeficientes.

Retomando ahora el análisis de los datos de la encuesta con toda la información disponible, sea de nuevo, L_i = el número de horas semanales que el individuo i dedica a la lectura y sean: ED_i = edad del cliente, AP_i = número de aparatos de TV, video y computadores disponibles en la unidad familiar (variables cuantitativas) y S_i = sexo y E_i = educación. Las clases o categorías que definen a estos factores cualitativos son: H = hombre y M = mujer para el sexo; P = estudios primarios, S = secundarios y U = universitarios para el factor educación. Definimos para estos factores las mismas variables ficticias anteriores, con lo que, si eliminamos las categorías Hombre (H) y estudios primarios (P), el modelo resultante es:

$$L_i = \alpha_0 + \alpha_1 M_i + \alpha_2 B_i + \alpha_3 U_i + \alpha_4 ED_i + \alpha_5 AP_i + \mu_i \quad (7.8)$$

Los resultados de la estimación del modelo (7.8) se muestran en la Tabla 7.9. A partir de los datos que proporciona la encuesta aplicada a los lectores, se obtiene:

$$\widehat{lectura} = 10,728 - 1,943M - 0,412S + 4,226U + 0,061ED - 0,702AP \quad (7.9)$$

$$(1,94) \quad (0,69) \quad (0,85) \quad (0,84) \quad (0,04) \quad (0,22)$$

$$R^2 = 24,07\%; \quad \bar{R}^2 = 22,08\%; \quad \Sigma\mu^2 = 4.518,32; \quad n = 197$$

Tabla 7.9***Estimación modelo utilizando distintas variables cualitativas.***

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10,72833	1,942207	5,523782	0,0000
M	-1,943075	0,694250	-2,798810	0,0057
S	-0,412040	0,847045	-0,486444	0,6272
U	4,226305	0,840076	1,698781	0,0000
ED	0,061684	0,036311	1,698781	0,0910
AP	-0,701642	0,216369	-3,242806	0,0014
R-squared	0,240678	Mean dependent var		11,20812
Adjusted R-squared	0,220801	S.D. dependent var		5,509948
S.E. of regression	4,863756	Akaike info criterion		6,031483
Sum squared resid	4.518,320	Schwarz criterion		6,161479
Log likelihood	-588,1010	Hannan-Quinn criterion		6,071962
F-statistic	12,10805	Durbin-Watson stat		2,019664
Prob(F-statistic)	0,000000			

Los estadísticos individuales son:

$$|t_M| = \frac{0,372}{1,12} = 0,33; |t_S| = \frac{0,5765}{1,16} = 0,50; |t_U| = \frac{6,796}{1,16} = 5,86; |t_{MS}| = \frac{0,041}{0,04} = 1,025; |t_{MU}| = \frac{0,626}{0,21} = 2,98$$

$$|t_{ED}| = \frac{0,426}{1,65} = 0,26 \text{ y } |t_{AP}| = \frac{5,350}{1,66} = 3,22$$

La distribución apropiada es una t y en las tablas de la distribución t de Student para un nivel de significación de 5% y para 191 grados de libertad ($T-K-1 = 197-5-1 = 191$), encontramos un valor crítico igual a 1,972, por consiguiente:

$|t_M| = 2,80 > 1,972 \rightarrow$ Se rechaza la hipótesis nula

$|t_S| = 0,49 < 1,972 \rightarrow$ No se rechaza la hipótesis nula

$|t_U| = 5,03 > 1,972 \rightarrow$ Se rechaza la hipótesis nula

$|t_{ED}| = 1,70 < 1,972 \rightarrow$ No se rechaza la hipótesis nula

$|t_{AP}| = 3,24 > 1,972 \rightarrow$ Se rechaza la hipótesis nula

Así pues, concluimos que en el modelo existe un efecto diferencial derivado del sexo, el número medio estimado de horas semanales de lectura de las mujeres es mayor (para cualquier nivel educativo) que el de los hombres y el número medio estimado de horas de lectura de personas con estudios universitarios es mayor que el de personas con estudios primarios (para ambos sexos).

Resulta también útil estimar el modelo restringido (véase Tabla 7.10) en el que se supone que ni la edad (ED), ni tampoco el número de aparatos de TV, video y computadores personales disponibles en la unidad familiar son factores relevantes para explicar el número de horas de lectura semanales.

Tabla 7.10

Estimación modelo lectura contra edad de los clientes y aparatos en el hogar.

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10,17098	2,062923	4,930374	0,0000
ED	0,078900	0,039872	1,978827	0,0493
AP	-0,680419	0,237494	-2,864992	0,0046
R-squared	0,064609	Mean dependent var		11,20812
Adjusted R-squared	0,054966	S.D. dependent var		5,509948
S.E. of regression	5,356379	Akaike info criterion		6,209565
Sum squared resid	4.466,015	Schwarz criterion		6,259563
Log likelihood	-608,6421	Hannan-Quinn criterion		6,229804
F-statistic	6,699923	Durbin-Watson stat		2,084109
Prob(F-statistic)	0,001536			

$$\widehat{lectura} = 10,171 + 0,079ED - 0,680AP \quad (7.10)$$

$$(2,063) \quad (0,040) \quad (0,237)$$

$$R^2 = 0,065; \quad \bar{R}^2 = 0,055; \quad \Sigma\mu^2 = 5.566,02; \quad n = 197$$

Un contraste conjunto $H_0: \alpha_1 = \alpha_2 = 0$ frente $H_\alpha: \alpha_1 \neq \alpha_2 \neq 0$ se realiza a través de diferencias de cuadrados de residuos, del siguiente modo:

Hipótesis nula		Hipótesis alterna
H_0 : La edad y el número de aparatos de entretenimiento disponibles en el hogar no son factores relevantes para explicar el número de horas de lectura semanales.	frente	H_α : La edad y el número de aparatos de entretenimiento disponibles en el hogar son factores relevantes para explicar el número de horas de lectura semanales.
$H_0: \alpha_1 = \alpha_2 = 0$		$H_\alpha: \alpha_1 \neq \alpha_2 \neq 0$

$$F = \frac{(5.566,02 - 4.518,32)/2}{4.518,32/(197-6)} = 14,68$$

Como $F_{0,05}(3, 191) = 2,652 \rightarrow 14,68 > 2,652$, por consiguiente, se rechaza la hipótesis nula. En conclusión, la edad y el número de aparatos de entretenimiento disponibles en el hogar (TV, video y PC) no son factores relevantes para explicar el número de horas de lectura semanales.

La aplicación del modelo que se introduce en esta sección para estudiar los datos procedentes de la encuesta de hábitos de lectura supone el siguiente modelo:

$$L_i = \alpha_0 + \alpha_1 M_i + \alpha_2 S_i + \alpha_3 U_i + \alpha_4 MS + \alpha_5 MU + \alpha_6 ED_i + \alpha_7 AP_i + \mu_i \quad (7.11)$$

Este modelo incluye las dos interacciones entre las variables ficticias para recoger la posible interrelación entre dichos factores: sexo y educación. Para interpretar los coeficientes obtenemos la expresión correspondiente a la esperanza matemática de L en cada caso, $E(L_i | i \text{ es } H, S) = \alpha_0 + \alpha_1$.

Ahora procederemos a especificar el modelo (7.12), en el cual se incorpora una combinación de variables cualitativas como el sexo y el nivel de estudios y factores individuales como la edad y el número de aparatos de entrenamiento disponibles en el hogar.

$$L_i = \alpha_0 + \alpha_1 M_i + \alpha_2 S_i + \alpha_3 U_i + \alpha_4 MS + \alpha_5 MU + \alpha_6 Ed_i + \alpha_7 Ap_i + \mu_i \quad (7.12)$$

Enseguida, procedemos a estimar en EViews el modelo (7.12), cuyos resultados se muestran en la Tabla 7.11.

Al estimar el modelo se obtiene:

$$\widehat{Lectura} = 10,542 - 0,372M - 0,576S + 6,796U + 0,041MS - 0,626MU + 0,426ED - 5,350AP \quad (7.13)$$

(1,89) (1,12) (1,16) (1,16) (0,04) (0,21) (1,65) (1,66)

$R^2 = 29,41\%$; $\bar{R}^2 = 26,79\%$; $\Sigma\hat{\mu}^2 = 4.200,63$; $n = 197$

Los estadísticos individuales son:

$$|t_M| = \frac{0,372}{1,12} = 0,33; \quad |t_S| = \frac{0,5765}{1,16} = 0,50; \quad |t_U| = \frac{6,796}{1,16} = 5,86; \quad |t_{MS}| = \frac{0,041}{0,04} = 1,025; \quad |t_{MU}| = \frac{0,626}{0,21} = 2,98$$

$$|t_{ED}| = \frac{0,426}{1,65} = 0,26 \text{ y } |t_{AP}| = \frac{5,350}{1,66} = 3,22$$

La distribución apropiada es una t y en las tablas de la distribución t de Student para un nivel de significación de 5% y para 189 grados de libertad ($T-K-1 = 197-7-1 = 189$), encontramos un valor crítico igual a 1,9726, por consiguiente:

$$|t_M| = 0,33 < 1,9726 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_S| = 0,50 < 1,9726 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_U| = 5,86 > 1,9726 \rightarrow \text{Se rechaza la hipótesis nula}$$

$$|t_{MS}| = 1,025 < 1,9726 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_{MU}| = 2,98 > 1,9726 \rightarrow \text{Se rechaza la hipótesis nula}$$

$$|t_{ED}| = 0,26 < 1,9726 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_{AP}| = 3,22 > 1,9726 \rightarrow \text{Se rechaza la hipótesis nula}$$

Por lo anterior, concluimos que en el modelo existe un efecto diferencial derivado del sexo, así pues, el número medio estimado de horas semanales de lectura de las mujeres es mayor (para cualquier nivel educativo) que el de los hombres y el número medio estimado de horas de lectura de personas con estudios universitarios es mayor que el de personas con estudios primarios (para ambos sexos).

En el modelo (7.13) las variables: estudios universitarios (U), mujeres con estudios universitarios (MU) y disponibilidad de aparatos de entretenimiento en la unidad familiar (AP) no tienen significación estadística al 5%, por ende, estos factores no son relevantes a la hora de explicar el número de horas de lectura semanales.

Otro camino para llegar a esta misma conclusión, consiste en observar la columna de probabilidad (*Prob*) de la Tabla 7.11. En esta se aprecia que los parámetros estimados son significativamente distintos de cero al 95% (p-valores menores que 0,05), salvo los relativos a las variables M, S, MS y ED. Al parecer las mujeres con estudios universitarios (MU) y la existencia de aparatos de entretenimiento (AP) son factores relevantes en la explicación del número de horas de lectura semanales.

Tabla 7.11

Estimación modelo utilizando distintas variables cualitativas.

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10,54179	1,889576	5,578918	0,0000
M	-0,372134	1,117308	-0,333064	0,7395
S	-0,575851	1,161704	-0,495695	0,6207
U	6,795608	1,155432	5,881444	0,0000
MS	0,424972	1,652828	0,257118	0,7974
MU	-5,350037	1,663128	-3,216852	0,0015
ED	0,041327	0,035956	1,149386	0,2518
AP	-0,625799	0,210953	-2,966538	0,0034
R-squared	0,294067	Mean dependent var		11,20812
Adjusted R-squared	0,267921	S.D. dependent var		5,509948
S.E. of regression	4,714399	Akaike info criterion		5,978882
Sum squared resid	4.200,631	Schwarz criterion		6,112210
Log likelihood	-580,9198	Hannan-Quinn criterion		6,032854
F-statistic	11,24726	Durbin-Watson stat		2,014014
Prob(F-statistic)	0,000000			

El lector notará, que al eliminar del modelo las variables irrelevantes M, S, MS y ED, el valor del \bar{R}^2 se incrementó ligeramente de 26,79% a 27,72%, mientras que el R^2 disminuyó de 29,41% a 28,83%, Como se mencionó anteriormente, el coeficiente de determinación se eleva a medida que se incorporan nuevas variables en el modelo, mientras que el coeficiente de determinación ajustado disminuye; por consiguiente, el coeficiente de determinación ajustado es un mejor indicador del ajuste de un modelo. La Tabla 7.12 muestra los resultados de la estimación del modelo ajustado (7.14).

Tabla 7.12

Estimación modelo lectura contra variables cualitativas relevantes.

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12,09013	0,843601	14,33158	0,0000
U	7,301693	0,900343	8,109900	0,0000
MU	-5,908649	1,188862	-4,969919	0,0000
AP	-0,654975	0,207506	-3,156417	0,0019
R-squared	0,288282	Mean dependent var		11,20812
Adjusted R-squared	0,277219	S.D. dependent var		5,509948
S.E. of regression	4,684367	Akaike info criterion		5,946434
Sum squared resid	4.235,056	Schwarz criterion		6,013098
Log likelihood	-581,7238	Hannan-Quinn criterion		5,973420
F-statistic	26,05824	Durbin-Watson stat		2,032895
Prob(F-statistic)	0,000000			

Por consiguiente, se obtiene:

$$\widehat{Lectura} = 12,090 + 7,302U - 5,909MU - 0,655AP \quad (7.14)$$

$$(0,84) \quad (0,90) \quad (1,19) \quad (0,21)$$

$$R^2 = 28,83\%; \quad \bar{R}^2 = 27,72\%; \quad \Sigma \hat{\mu}^2 = 4.235,06; \quad n = 197$$

En síntesis, el número de horas de lectura semanales depende de manera directa del nivel de estudios universitarios (las personas a medida que alcanzan un mayor nivel de estudios tienden a leer mucho más) y, en forma inversa, del número de aparatos de entretenimiento disponibles en la unidad familiar (a mayor número de aparatos de TV, video y computadores disponibles en el hogar, el número de horas de lectura disminuye).

Ahora procederemos a realizar tres contrastes adicionales:

1. No existe relación entre los factores cualitativos: $\mathcal{H}_0: \alpha_4 = \alpha_5 = 0$
2. No existe efecto diferencial debido al sexo en el nivel de estudios universitarios: $\mathcal{H}_0: \alpha_3 = 0$.
3. Ni el sexo ni la educación ayudan a explicar el número de horas semanales dedicadas a la lectura: $\mathcal{H}_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$

Llevaremos a cabo el primer contraste, es decir, si el efecto del sexo y el nivel educativo no explican el número de horas semanales dedicadas a la lectura, así: $\mathcal{H}_0: \alpha_4 = \alpha_5 = 0$. Esto puede realizarse por la vía de la diferencia de la suma de cuadrados de residuos, para lo cual, estimamos el siguiente modelo restringido:

$$L_i = \alpha_0 + \alpha_4 MB + \alpha_5 MU + \mu_i \quad (7.15)$$

Al estimar el modelo, se obtiene:

Tabla 7.13***Estimación modelo lectura contra educación de las clientes mujeres.***

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11,60432	0,465773	24,91408	0,0000
MS	-2,018110	1,121064	-1,800174	0,0734
MU	-0,673282	1,121064	-0,600574	0,5488
R-squared	0,016861	Mean dependent var		11,20812
Adjusted R-squared	0,006726	S.D. dependent var		5,509948
S.E. of regression	5,491387	Akaike info criterion		6,259350
Sum squared resid	5.850,134	Schwarz criterion		6,309348
Log likelihood	-613,5460	Hannan-Quinn criterion		6,279590
F-statistic	1,663604	Durbin-Watson stat		2,116531
Prob(F-statistic)	0,192146			

$$\widehat{Lectura} = 11,604 - 2,018MS - 0,673MU \quad (7.16)$$

$$(0,47) \quad (1,12) \quad (1,12)$$

$$R^2 = 1,70\% \quad \bar{R}^2 = 0,70\%; \quad \Sigma \hat{\mu}^2 = 5.850,13; \quad n = 197$$

El estadístico para este caso toma el valor de:

$$\mathcal{F} = \frac{(5.850,13 - 4.200,63)/2}{4.200,63/(197-7-1)} = 37,10$$

En las tablas de la distribución \mathcal{F} se tiene que el nivel de significación del 5%, el valor crítico es $\mathcal{F}_{0,05}(2, \infty)$ es igual a 3,04 \rightarrow como $37,10 > 3,04$, por lo tanto, se rechaza la hipótesis nula de que MB y MU sean conjuntamente iguales a cero. Esto permite concluir que existe relación entre estos factores cualitativos.

Con respecto al segundo contraste: $\mathcal{H}_0: \alpha_5 = 0$. Estimamos el siguiente modelo restringido:

$$L_i = \alpha_0 + \alpha_5 MU + \mu_i \quad (7.17)$$

Al estimar el modelo se obtiene los resultados que muestran la Tabla 7.14:

Tabla 7.14

Estimación modelo lectura contra mujeres con educación universitaria.

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11,25595	0,426097	26,41642	0,0000
MU	-0,324918	1,110561	-0,292571	0,7702
R-squared	0,000439	Mean dependent var		11,20812
Adjusted R-squared	-0,004687	S.D. dependent var		5,509948
S.E. of regression	5,522846	Akaike info criterion		6,265764
Sum squared resid	5.947,856	Schwarz criterion		6,299096
Log likelihood	-615,1778	Hannan-Quinn criterion		6,279257
F-statistic	0,085598	Durbin-Watson stat		2,106672
Prob(F-statistic)	0,770161			

$$\widehat{Lectura} = 11,256 - 0,325MU \quad (7.18)$$

$$(0,43) \quad (1,11)$$

$$R^2 = 0,0439\%; \quad \bar{R}^2 = -0,04687\%; \quad \Sigma \hat{\mu}^2 = 5.947,86; \quad n = 197$$

El estadístico F para este caso toma el valor de:

$$\mathcal{F} = \frac{(5.947,86 - 4.200,63)/2}{4.200,63/(197-7-1)} = 39,30$$

En las tablas de la distribución se tiene que el nivel de significación del 5%, el valor crítico es $F_{0,05}(1, \infty)$ es igual a 3,84 \rightarrow como $39,30 > 3,84$, por lo tanto, se rechaza la hipótesis nula de que MU sea igual a cero, por consiguiente, existe relación entre la lectura y las mujeres con estudios universitarios.

Con respecto al tercer contraste: $\mathcal{H}_0: \alpha_5 = 0$. Estimamos el siguiente modelo restringido:

$$L_i = \alpha_0 + \alpha_1 M_i + \alpha_2 B_i + \alpha_3 U_i + \alpha_4 MS + \alpha_5 MU + \mu_i \quad (7.19)$$

Los resultados de la estimación se muestran en la Tabla 7.15.

Tabla 7.15

Estimación modelo final de lectura.

Variable Dependiente: L

Método: MCO

Muestra: 1 197

Observaciones incluidas: 197

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10,08571	0,815092	12,37371	0,0000
M	-0,138346	1,129735	-0,122459	0,9027
S	-0,773214	1,179422	-0,655587	0,5129
U	7,090756	1,161160	6,106617	0,0000
MS	0,412053	1,674752	0,246038	0,8059
MU	-6,107090	1,661941	-3,674673	0,0003
R-squared	0,253613	Mean dependent var		11,20812
Adjusted R-squared	0,234074	S.D. dependent var		5,509948
S.E. of regression	4,822151	Akaike info criterion		6,014301
Sum squared resid	4.441,350	Schwarz criterion		6,114297
Log likelihood	-586,4086	Hannan-Quinn criterion		6,054780
F-statistic	12,97989	Durbin-Watson stat		2,044171
Prob(F-statistic)	0,000000			

Al estimar el modelo, se obtiene:

$$\widehat{Lectura} = 10,086 - 0,138M - 0,773S + 7,091U + 0,412MS - 6,107MU \quad (7.20)$$

$$(0,82) \quad (1,13) \quad (1,18) \quad (1,16) \quad (1,67) \quad (1,66)$$

$$R^2 = 25,36\%; \quad \bar{R}^2 = 23,41\%; \quad \Sigma \hat{\mu}^2 = 4.441,35; \quad n = 197$$

Los estadísticos individuales son:

$$|t_M| = \frac{0,138}{1,13} = 0,12; \quad |t_S| = \frac{0,773}{1,18} = 0,66; \quad |t_U| = \frac{7,091}{1,16} = 6,11; \quad |t_{MS}| = \frac{0,412}{1,67} = 0,25; \quad |t_{MU}| = \frac{6,107}{1,66} = 3,67$$

La distribución apropiada es una t y en las tablas de la distribución t de Student para un nivel de significación de 5% y para 189 grados de libertad ($T-K-1 = 197-5-1 = 191$), encontramos un valor crítico igual a 1,97246, por consiguiente:

$$|t_M| = 0,12 < 1,97246 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_S| = 0,66 < 1,97246 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_U| = 6,11 > 1,97246 \rightarrow \text{Se rechaza la hipótesis nula}$$

$$|t_{MS}| = 0,25 < 1,97246 \rightarrow \text{No se rechaza la hipótesis nula}$$

$$|t_{MU}| = 3,67 > 1,97246 \rightarrow \text{Se rechaza la hipótesis nula}$$

El modelo (7.20) las variables: estudios universitarios (U) y mujeres con estudios universitarios (MU) tienen significación estadística al 5%, por ende, estos factores son relevantes a la hora de explicar el número de horas de lectura semanales.

También llegamos a esta misma conclusión, al observar la columna de probabilidad (*Prob.*) en la Tabla 7.15. En esta se aprecia que los parámetros estimados son significativamente distintos de cero al 95% (p-valores menores que 0,05), salvo los relativos a las variables M, S y MS.

Si se realiza un contraste de significatividad conjunta, el estadístico para este caso toma el valor de:

$$\mathcal{F} = \frac{(4.441,35 - 4.200,63)/2}{4.200,63/(197-7-1)} = 5,41$$

En las tablas de la distribución se tiene que el nivel de significación del 5%, el valor crítico es $\mathcal{F}_{0,05}(5, \infty)$ es igual a 2,21 \rightarrow como $5,41 > 2,21$, por consiguiente, se rechaza la hipótesis nula de que M, S, U, MS y MU sean, conjuntamente iguales a cero, por tanto, existe relación entre el sexo y el nivel educativo.

Don Pablo al observar los modelos especificados y estimados por Emilio usando en su gran mayoría, variables de tipo cualitativo como el efecto vacacional, el sexo y la escolaridad y, otras de carácter cuantitativo como la edad y el número de aparatos de entretenimiento disponibles en el hogar; a pesar del gran esfuerzo intelectual realizado por su hijo, cree no haber entendido de manera clara los factores que efectivamente explican la lectura medida en número de horas semanales. Se recuerda que Emilio, con todas las objeciones posibles de su padre, es el líder y defensor del gasto en publicidad, el cual ha demostrado ser hasta el momento, un factor determinante de las ventas de la editorial.

Emilio, por su parte, convencido de la terquedad de su padre, considera resaltar de forma clara, los hallazgos del trabajo efectuado en este capítulo. Mirando a su padre, en tono enérgico, le responde: Si estuviste atento en lo que hice, debiste comprender que la campaña publicitaria que emprenda la editorial debe ir especialmente dirigida a las *mujeres con estudios universitarios* ya que, sistemáticamente, el número de horas dedicado a la lectura en este grupo es mayor que en el de los hombres. En otras palabras, el grupo de interés comercial de la editorial son las mujeres con estudios universitarios, quienes han demostrado ser mejores lectoras que los hombres. Esta evidencia empírica, nos indica que es necesario fortalecer la estrategia publicitaria dirigida a este tipo de clientes. ¡Mira papá!, los ahorros en el gasto en publicidad que tú planteas desde que tengo uso de razón, solo provocará pérdida de beneficios para la editorial. ¡Espero que te haya quedado claro el mensaje!

RETO 6:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

MULTICOLINEALIDAD

Supuesto 8: *No debe existir multicolinealidad entre las variables explicativas incluidas en el modelo de regresión.*

La multicolinealidad es una condición que ocurre cuando algunas variables explicativas incluidas en el modelo están correlacionadas con otras variables explicativas. Según Gujarati (2010:323), si la multicolinealidad es perfecta, los coeficientes de regresión de las variables explicativas son indeterminados, y sus errores estándar, infinitos. Si la multicolinealidad es menos que perfecta, los coeficientes de regresión, aunque sean determinados, poseen grandes errores estándar (en relación con los coeficientes mismos), lo cual significa que los coeficientes no pueden ser estimados con gran precisión o exactitud". Lo anterior parece indicar, que la multicolinealidad pone en riesgo y en tela de juicio, el efecto marginal que cada variable explicativa ejerce sobre la variable dependiente, es decir, las variables explicativas no son tan independientes a la hora de explicar la variable dependiente.

Una cuestión importante que debe analizarse al estudiar los resultados de un modelo de regresión es el grado de relación lineal existente entre las observaciones de las variables explicativas, en nuestro caso, la relación existente entre el precio de la editorial (PE), el precio de la competencia (PC), el gasto en publicidad (GP) y la renta personal (R). A este respecto, pueden existir tres posibles situaciones: multicolinealidad perfecta, ortogonalidad y multicolinealidad imperfecta.

Carrascal et al, (2004), afirman que la multicolinealidad imperfecta consiste en la existencia de una relación lineal fuerte entre las variables explicativas

del modelo especificado. La multicolinealidad imperfecta es un problema muy común en la práctica econométrica y no conlleva el incumplimiento de ninguna de las hipótesis en las que se fundamenta el modelo de regresión lineal clásico. Por lo tanto, los estimadores mínimos cuadrados están determinados y son los mejores estimadores que pueden obtenerse, cumpliendo todas las propiedades deseadas para un estimador. Sin embargo, no son buenos estimadores debido a las consecuencias que se derivan de este problema, las cuales se sintetizan en:

- Varianzas grandes de los estimadores. Cuanto más fuerte sea la relación lineal entre la variable X y el resto de las variables explicativas del modelo, es decir, cuanto mayor sea el grado de multicolinealidad, mayor será la varianza del estimador de su parámetro.
- Inestabilidad de los estimadores ante pequeñas variaciones muestrales: el hecho de que la varianza de los estimadores sea elevada implica que los estimadores son inestables.
- Dificultad para interpretar los coeficientes, y por lo tanto, sus estimaciones. Los coeficientes de regresión β se interpretan como el cambio que se produce en la variable dependiente Y al variar la variable explicativa X en una unidad, permaneciendo el resto de las variables explicativas constantes. Cuando existe multicolinealidad imperfecta es imposible suponer que el resto de las variables permanecen constantes al existir altas correlaciones lineales entre las mismas. Por este motivo, los parámetros pierden su significado, y por lo tanto, sus estimaciones.

La multicolinealidad imperfecta hace imposible determinar el efecto aislado de cada variable explicativa sobre la variable dependiente. Como la multicolinealidad imperfecta no supone el incumplimiento de ninguna hipótesis clásica, EViews no incluye ningún menú específico para su detección, pero a través de una serie de indicios y procedimientos se puede detectar su presencia.

Ahora, detectemos la presencia o no de multicolinealidad imperfecta en modelo de ventas de la editorial. Situados en la ventana del objeto ecuación

seleccionamos *Proc* → *Make Regresor Group* y damos clic (véase Figura 8.1), inmediatamente se abre la ventana de un objeto grupo que contiene las variables incluidas en el modelo, incluida la variable dependiente, como se muestra en la Figura 8.2.

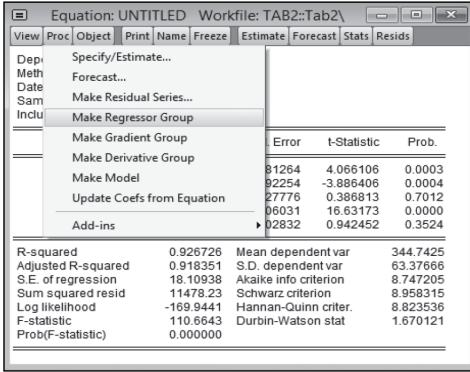


Figura 8.1

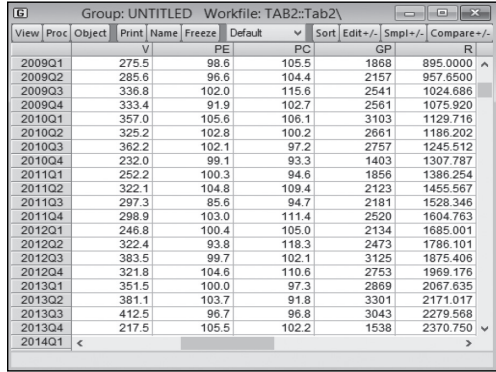


Figura 8.2

Sobre el objeto grupo calculamos la matriz de correlaciones, para ello seleccionamos *View* → *Covariance Analysis* y damos clic como se muestra en la Figura 8.3 y en el cuadro de diálogo que se abre marcamos la opción *Correlation* y damos OK (Figura 8.4).

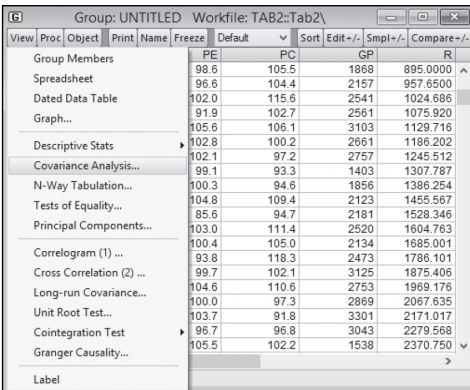


Figura 8.3

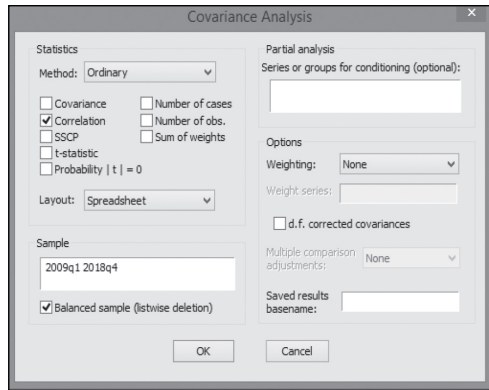


Figura 8.4

De este modo, obtenemos la matriz de correlación que aparece en la Tabla 8.1. Analizando el coeficiente de correlación lineal se observa que en ninguna de las relaciones de las variables explicativas, este coeficiente tiene una

magnitud apreciable ($r \approx 1$), por lo general, se encuentra por debajo de 0,6 que se interpreta como una intensidad entre moderada y débil, por lo tanto, entre las variables explicativas del modelo de ventas de la editorial no se aprecia indicios de existencia de multicolinealidad imperfecta.

Tabla 8.1

Matriz de correlación lineal.

Variables explicativas	PE	PC	GP	R
PE	1,000000	-0,258176	-0,159481	-0,498638
PC	-0,258176	1,000000	0,285806	0,556504
GP	-0,159484	0,285806	1,000000	0,464269
R	-0,498638	0,556504	0,464269	1,000000

Según Alonso et al, (2005:161), la multicolinealidad puede también detectarse por medio de los coeficientes de determinación obtenidos en las regresiones auxiliares de cada variable explicativa sobre las restantes. Si algunos de los coeficientes de determinación de las regresiones auxiliares son grandes, lo cual indica una correlación alta entre las variables explicativas, entonces posiblemente nos enfrentamos a un problema de multicolinealidad en los datos.

Lo anterior nos exige estimar los siguientes modelos auxiliares, siguiendo el mismo procedimiento aprendido hasta el momento.

$$\widehat{PE} = \hat{\beta}_0 + \hat{\beta}_1 PC_t + \hat{\beta}_2 GP_t + \hat{\beta}_3 R_t + \hat{\mu}_t \quad (8.1)$$

$$\widehat{PC} = \hat{\beta}_0 + \hat{\beta}_1 PE_t + \hat{\beta}_2 GP_t + \hat{\beta}_3 R_t + \hat{\mu}_t \quad (8.2)$$

$$\widehat{GP} = \hat{\beta}_0 + \hat{\beta}_1 PE_t + \hat{\beta}_2 PC_t + \hat{\beta}_3 R_t + \hat{\mu}_t \quad (8.3)$$

$$\widehat{R} = \hat{\beta}_0 + \hat{\beta}_1 PE_t + \hat{\beta}_2 PC_t + \hat{\beta}_3 GP_t + \hat{\mu}_t \quad (8.4)$$

Con el apoyo de EViews realizamos las estimaciones respectivas. Únicamente ilustraremos la estimación del modelo (8.1), puesto que el proceso se repite de la misma manera para las demás regresiones auxiliares.

En el ventana de *Workfile* seleccionamos las variables de interés en el siguiente orden de intervención: pe, pc, gp y r (Figura 8.5), y posteriormente, abrimos como ecuación (*as Equation*) y damos clic. En la ventana *Equation specification* ordenamos las variables en el siguiente orden (primero la variable dependiente y posteriormente, el intercepto seguido de las variables explicativas): pe c pc gp r y damos Aceptar, como se indica en la Figura 8.6.

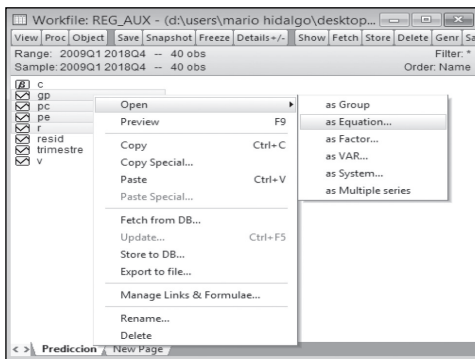


Figura 8.5

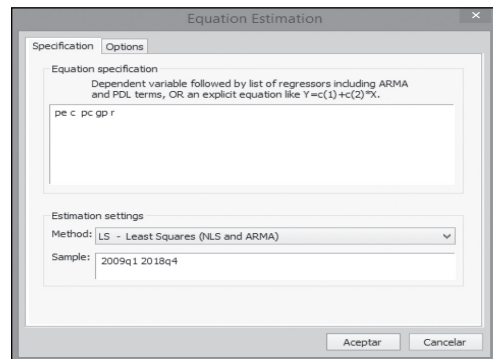


Figura 8.6

En las Tablas 8.2, 8.3, 8.4 y 8.5 se muestran los resultados de las regresiones auxiliares.

Tabla 8.2***Regresiones auxiliares: Precio medio de la editorial.***

Variable Dependiente: PE

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	99,13258	11,71045	8,465307	0,0000
PC	0,015611	0,110948	0,140708	0,8889
GP	0,001138	0,002033	0,559873	0,5790
R	-0,002545	0,000860	-2,958811	0,0054
R-squared	0,255661	Mean dependent var		96,76750
Adjusted R-squared	0,193633	S.D. dependent var		6,828049
S.E. of regression	6,131447	Akaike info criterion		6,559378
Sum squared resid	1.353,407	Schwarz criterion		6,728266
Log likelihood	-127,1876	Hannan-Quinn criterion		6,620443
F-statistic	4,121690	Durbin-Watson stat		1,619692
Prob(F-statistic)	0,013027			

Tabla 8.3***Regresiones auxiliares: Precio medio de la competencia.***

Variable Dependiente: PC

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	92,40861	26,22531	3,523642	0,0012
PE	0,035209	0,250231	0,140708	0,8889
GP	0,000643	0,003065	0,209753	0,8350
R	0,003958	0,001280	3,092051	0,0038
R-squared	0,311035	Mean dependent var		108,5175
Adjusted R-squared	0,253621	S.D. dependent var		10,65847
S.E. of regression	9,208198	Akaike info criterion		7,372706
Sum squared resid	3.052,473	Schwarz criterion		7,541593
Log likelihood	-143,4541	Hannan-Quinn criterion		7,433769
F-statistic	5,417426	Durbin-Watson stat		2,486258
Prob(F-statistic)	0,003519			

Tabla 8.4***Regresiones auxiliares: gasto en publicidad.***

Variable Dependiente: GP

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.206,531	1.640,679	0,735385	0,4669
PE	7,58349	13,54487	0,559873	0,5790
PC	1,898841	9,052752	0,209753	0,8350
R	0,180061	0,072288	2,490885	0,0175
R-squared	0,223398	Mean dependent var		2646,825
Adjusted R-squared	0,158681	S.D. dependent var		545,6214
S.E. of regression	500,4627	Akaike info criterion		15,36358
Sum squared resid	9.016.666	Schwarz criterion		15,53247
Log likelihood	-303,2717	Hannan-Quinn criterion		15,42465
F-statistic	3,451928	Durbin-Watson stat		1,478912
Prob(F-statistic)	0,026455			

Tabla 8.5***Regresiones auxiliares: Renta personal.***

Variable Dependiente: R

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2303,603	3.498,775	0,658403	0,5145
PE	-76,87078	25,98030	-2,958811	0,0054
PC	53,01542	17,14571	3,092051	0,0038
GP	0,816447	0,327774	2,490885	0,0175
R-squared	0,526328	Mean dependent var		2779,105
Adjusted R-squared	0,486856	S.D. dependent var		1487,668
S.E. of regression	1.065,677	Akaike info criterion		16,87525
Sum squared resid	40.884.057	Schwarz criterion		17,04414
Log likelihood	-333,5050	Hannan-Quinn criterion		16,93631
F-statistic	13,33400	Durbin-Watson stat		0,788488
Prob(F-statistic)	0,000005			

Ahora, visualizamos los resultados de la estimación de las cuatro regresiones auxiliares, así:

$$\widehat{PE} = 99,13 + 0,02PC_t + 0,0011GP_t - 0,003R_t \quad \bar{R}^2 = 19,36\% \quad (8.5)$$

$$\widehat{PC} = 92,41 + 0,035PE_t + 0,000643GP_t + 0,0040R_t \quad \bar{R}^2 = 25,36\% \quad (8.6)$$

$$\widehat{GP} = 1206,53 + 7,58PE_t + 1,90PC_t + 0,18R_t \quad \bar{R}^2 = 15,87\% \quad (8.7)$$

$$\widehat{R} = 2.303,60 - 76,87PE_t + 53,015PC_t + 0,82GP_t \quad \bar{R}^2 = 48,68\% \quad (8.8)$$

Como puede observarse, la correlación entre las variables explicativas del modelo de ventas de la editorial es débil (menor de 0,5). Aunque podría suceder que existiera colinealidad entre los regresores a pesar de que los coeficientes de correlación simple sean reducidos, las regresiones auxiliares entre las variables explicativas no muestran en este caso, evidencia de relación lineal entre ellas (no hay multicolinealidad). Por ende, podemos concluir que el precio de la competencia y la renta personal tienen poca capacidad explicativa, sin que pueda atribuirse su poca significación a problemas de datos. Por otro lado, la capacidad explicativa del gasto en publicidad queda perfectamente evidenciada.

Por lo anterior, el modelo inicial propuesto por Don Pablo queda desechado; la propuesta de utilizar únicamente el precio como variable explicativa proporciona un modelo con error de especificación por omisión de variables relevantes, lo que impide que podamos confiar en los resultados numéricos obtenidos en su estimación: los coeficientes estimados serán sesgados y los contrastes que podamos realizar con esos resultados serán inválidos.

No cabe duda a este respecto, que el modelo propuesto por Emilio es más adecuado. Quizás Emilio debería aceptar que, a pesar de su creencia, los datos no avalan que el precio de la competencia sea una variable relevante. No obstante, como está encariñado con esta idea, decide mantener la variable en el modelo diciendo que al fin y al cabo no sesga los resultados aunque resulten menos precisos de lo que podrían ser.

En vista de todo lo que va oyendo, y para hacerse una idea que le permita discutir luego los resultados con su hijo, Don Pablo desea conocer el grado de correlación existente entre las variables incluidas en el modelo, así que Emilio le presenta los siguientes resultados:

Don Pablo: Si a mí sólo me interesa conocer el efecto del precio, ¿por qué voy a tener que meterme en un modelo más complicado? me basta con el que yo he propuesto en primer lugar.

Emilio: No puede ser, lo que tú propones es un disparate. Las consecuencias de no incluir otras variables relevantes son muy graves.

Don Pablo: ¿En qué sentido son muy graves? No entiendo el porqué de esa gravedad si el coeficiente estimado asociado a los precios no cambia tanto entre los dos modelos.

La respuesta a estas preguntas tiene que ver, naturalmente, con el hecho de que si no se incluyen variables relevantes en el Capítulo 5, provocará sesgos en la estimación de los coeficientes y de la varianza estimada de los residuos, invalidando la inferencia que se pueda realizar.

Dado que con los resultados precedentes se comprueba que la variable gasto en publicidad (GP), tiene indudable capacidad explicativa sobre las ventas, su exclusión del modelo como pretende Don Pablo provocaría sesgo en la estimación de los coeficientes (los elementos de la matriz $X'X$ muestra que este regresor no es ortogonal a la variable P , precio), en la estimación de la varianza e invalidaría la inferencia. La conversación entre padre e hijo continúa así:

Emilio: ¿Comprendes ahora por qué los resultados del modelo en que se incluye solamente la variable precio no son válidos? Ni la estimación ni los estadísticos de contraste son fiables.

Don Pablo: Comprendo, pero...entonces también el modelo propuesto por ti tiene problemas, ya que la variable 'Precio de la Competencia' no es relevante.

Efectivamente, la inclusión de esta variable no resulta necesaria de acuerdo con los resultados obtenidos hasta el momento. Con objeto de comprobar si su falta de significación podría ser debido a la existencia de una relación lineal estrecha con los otros regresores, se llevaron a cabo las regresiones auxiliares.

A partir de estos resultados resulta evidente que no existe problema de colinealidad entre las variables explicativas del modelo, lo que implica que el precio de la competencia no es significativo, sin que pueda encontrarse una razón para ello en la presencia de otros regresores en el modelo. Mantener la variable precio de la competencia (PC) en el modelo supone incurrir en un error de especificación por inclusión de variables irrelevantes; aunque sus consecuencias no son importantes en términos de propiedades básicas del estimador obtenido, que continua siendo insesgado y consistente, si supone un costo mantenerla al obtener menor precisión en la estimación que si la quitáramos.

RETO 7:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

HETEROSCEDASTICIDAD

Supuesto 3: *El valor medio de la perturbación μ_i es igual a cero: Dado el valor de X_i , la media o el valor esperado del término de perturbación aleatoria μ_i es cero. Simbólicamente, tenemos que: $E(\mu_i|X_i) = 0$.*

Supuesto 4: *Homoscedasticidad o varianza constante de μ_i : La varianza del término de error, o de perturbación, es la misma sin importar el valor de X . Simbólicamente, tenemos que: $Var(\mu_i) = \sigma^2$*

La palabra heteroscedasticidad se puede desglosar en dos partes, heteros (diferente) y cedasticidad (dispersión). De tal manera que, si unimos estas dos palabras obtenemos que la heteroscedasticidad que se equivale a que las perturbaciones (residuos) del modelo tienen diferente dispersión, es decir, que no son constantes a lo largo de toda la muestra. Por lo tanto, la heteroscedasticidad es la existencia de una varianza no constante en las perturbaciones aleatorias de un modelo econométrico.

Las principales consecuencias que se derivan del incumplimiento de la hipótesis de homoscedasticidad en los resultados de la estimación mínimo cuadrática son:

- Un error en el cálculo del estimador de la matriz de varianzas y covarianzas de los estimadores mínimo cuadráticos.
- Una pérdida de eficiencia en el estimador mínimo cuadrático. Al ser la matriz de las varianzas de las perturbaciones no escalar, el estimador MCO no es eficiente, es decir, es solo Estimador Lineal Insesgado (ELI) y no el Mejor Estimador Lineal Insesgado (MELI).

La hipótesis de homoscedasticidad establece que la varianza de las perturbaciones es igual a una constante: $\text{Var}(\mu_t) = \sigma^2$. Esta hipótesis, junto con el resto de hipótesis básicas, implica también una distribución homoscedástica para la variable dependiente, esto es, $Y_t \sim (X_t' \beta, \sigma^2)$ para todo t .

9.1 Especificación y causas de la heteroscedasticidad.

Se denomina heteroscedasticidad al caso en que, en el modelo de regresión, las varianzas de las perturbaciones son lineales para distintas observaciones. Formalmente, esto se indica escribiendo que $E(\mu_t^2) = \sigma_t^2$, para $t = 1, 2, 3, \dots, T$.

Don Pablo ya ha comprobado que los resultados obtenidos hasta el momento sobre el modelo propuesto para las ventas de su editorial, dependen de manera crucial de las hipótesis básicas, varianzas iguales u homoscedasticidad, no se satisface para las ventas de la editorial. Por ello, quiere que Emilio realice los análisis oportunos para decidir: 1) cuáles son las consecuencias que se derivan para los resultados encontrados hasta el momento sino se satisface dicha hipótesis, 2) cómo decidir si el problema es relevante para su modelo, y 3) qué hacer en el caso de que, efectivamente, se encuentre de que existe un problema de heteroscedasticidad.

Emilio, siempre dispuesto a convencer a su padre de que él tiene razón y de que se pueden tomar decisiones correctas si se tienen en cuenta los problemas específicos de los diferentes componentes del modelo de ventas de la editorial, se pone manos a la obra. Por lo que se refiere a los datos de nuestra editorial, y dada la Figura 5.5 del capítulo 5 en el que los residuos parecen crecer con el precio promedio de los libros, resultaría necesario llevar a cabo un contraste sobre la existencia o no de heteroscedasticidad. Por otro lado, en el capítulo 7 hemos encontrado que el modelo correcto debería incluir, además de las variables inicialmente consideradas, la variable LGP_t que recoge el comportamiento estacional. Lo razonable, por lo tanto, sería contrastar si este último modelo (7.8), presenta el mismo tipo de comportamiento en sus residuos. Examinados estos residuos en un gráfico similar al que se acaba de citar, encontramos la misma estructura de comportamiento que posee el modelo (4.1).

9.2 Detección de la heteroscedasticidad mediante métodos gráficos.

Para la detección de la heteroscedasticidad, abrimos el archivo de trabajo *Tab2*, nos situamos en la ventana del Workfile, haciendo clic en cualquier parte de la misma y pulsamos el botón *Genr* de su barra de herramientas. En el cuadro de diálogo resultante escribimos `abse=@abs(resid)` y pulsamos OK, como se muestra en la Figura 9.2. El resultado es un nuevo objeto serie llamado “**abse**” que recoge el valor absoluto de los errores de la ecuación que acabamos de estimar. Recuérdese que el objeto serie *resid* incluye los errores de la ecuación que se mantenga por defecto (*eq default*), en este caso, Eq01. La serie que recoge el cuadrado de los errores (CUADRAE) se genera repitiendo los mismos pasos y escribiendo en el cuadro de diálogo `cuadrae=resid^2`, como se indica en la Figura 9.3.

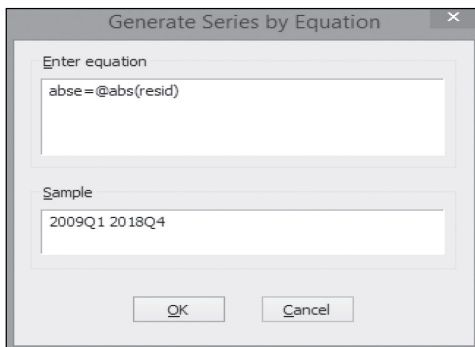


Figura 9.2

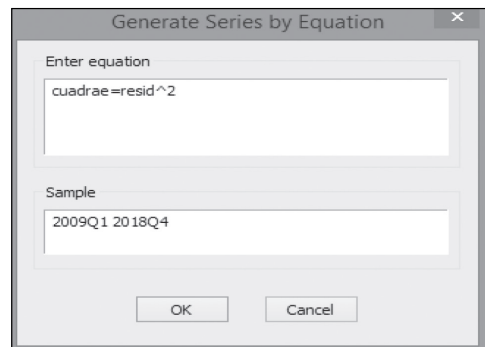


Figura 9.3

Seleccionamos la variable precio de la editorial (PE) y ABSE oprimiendo la tecla CTRL más el botón izquierdo del ratón y las abrimos como grupo mediante *Open* → *as Group* (Figura 9.4) y hacemos clic, obteniendo una ventana con las series PE y ABSE (véase Figura 9.5).

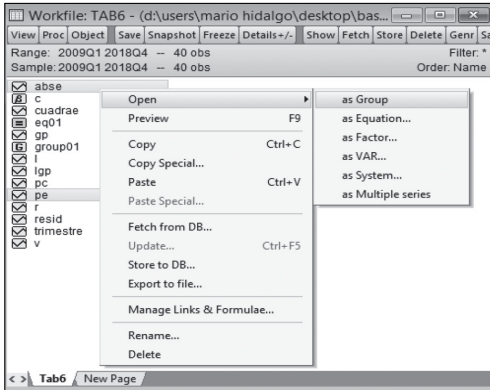


Figura 9.4

PE	ABSE	
2009Q1	98.8	62.76929
2009Q2	96.6	59.73418
2009Q3	102.0	10.54103
2009Q4	91.9	28.53669
2010Q1	105.6	43.45784
2010Q2	102.8	1.766987
2010Q3	102.1	36.29427
2010Q4	99.1	104.5031
2011Q1	100.3	80.06413
2011Q2	104.8	5.731882
2011Q3	85.6	86.89110
2011Q4	103.0	23.82652
2012Q1	100.4	85.11089
2012Q2	93.8	32.82504
2012Q3	99.7	49.11640
2012Q4	104.6	4.725392
2013Q1	100.0	18.17614
2013Q2	103.7	60.84619
2013Q3	96.7	67.51906
2013Q4	105.5	96.39541
2014Q1	99.2	106.2498
2014Q2	86.4	38.46515
2014Q3

Figura 9.5

Sobre la ventana del objeto grupo realizamos el gráfico mediante la opción *View* → *Graph* y damos clic (véase Figura 9.6). En la solapa *Specific* de la ventana que se abre, elegimos la opción *Scatter* y damos OK (Figura 9.7).

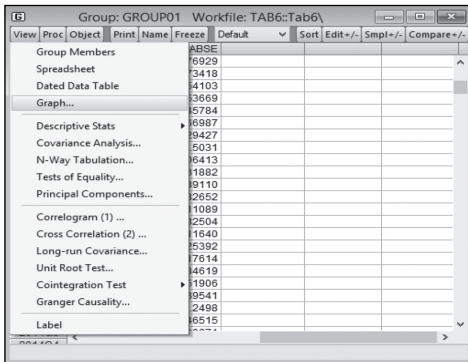


Figura 9.6

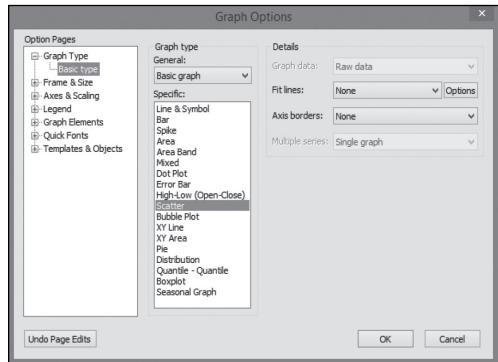


Figura 9.7

De este modo se obtiene el gráfico de los residuos de la regresión ventas contra precio de la editorial (véase Figura 9.8). En este gráfico se aprecia que la dispersión del valor absoluto de errores crece a medida que aumenta el precio promedio, lo cual puede ser un indicio de heteroscedasticidad. El gráfico de los errores al cuadrado se obtiene a través del mismo procedimiento: En el *Workfile* del archivo *Tab2* seleccionamos con el puntero del ratón PE y CUADRAE, abrimos como grupo mediante *Open* → *as Group* y damos clic. Posteriormente, seleccionamos *View* → *Graph* damos clic y elegimos *Scatter* y damos OK, obteniendo el gráfico correspondiente, como se muestra en la

Figura 9.9. Observamos que este gráfico es similar al anterior, mostrando una dispersión creciente del cuadrado de los errores a medida que aumenta el precio promedio de la editorial.

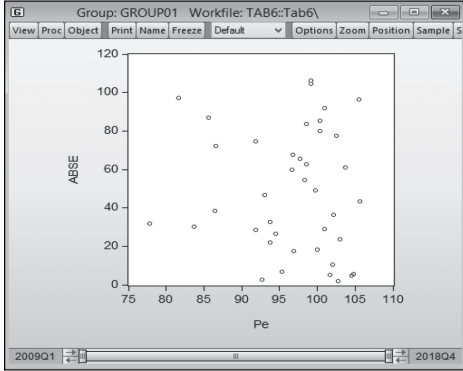


Figura 9.8

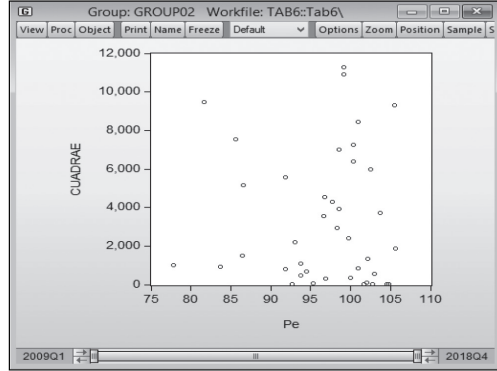


Figura 9.9

A continuación, utilizando el procedimiento aprendido hasta ahora, vamos a realizar los diagramas de dispersión de los errores contra las variables precio de la competencia (PC), gasto en publicidad (GP) y renta personal (R). Para el primer caso estimamos la ecuación de la regresión ventas contra precio de la competencia a la que la llamaremos Eq02. Nos situamos en la ventana del Workfile, haciendo clic en cualquier parte de la misma y pulsamos el botón Genr de su barra de herramientas. En el cuadro de diálogo resultante escribimos $abse=@abs(resid)$ y pulsamos OK como se muestra en la Figura 9.10. Pulsando la tecla CTRL más el botón izquierdo del ratón seleccionamos las variables precio de la competencia (PC) y ABSE y las abrimos como grupo mediante *Open* → *as Group* y hacemos clic, obteniendo una ventana con las series PC y ABSE (Figura 9.11).

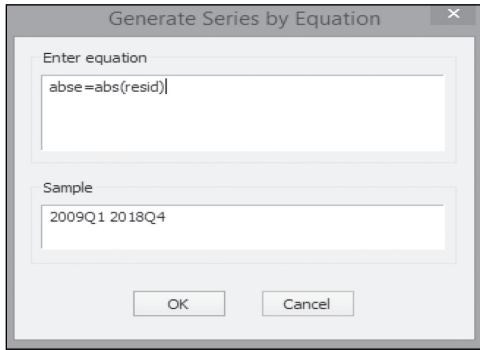


Figura 9.10

View	Proc	Object	Print	Name	Freeze	Default	Sort	Edit+/-	Smp+/-	Compare+/-
				Pc						
				ABSE						
				2009Q1	105.5	62.85126				
				2009Q2	104.4	50.42140				
				2009Q3	115.6	22.94364				
				2009Q4	102.7	0.879298				
				2010Q1	105.1	17.37790				
				2010Q2	100.2	1.925559				
				2010Q3	97.2	41.42861				
				2010Q4	93.3	80.51096				
				2011Q1	94.6	63.06444				
				2011Q2	109.4	24.51169				
				2011Q3	94.7	18.17624				
				2011Q4	111.4	51.94780				
				2012Q1	105.0	90.49223				
				2012Q2	118.3	43.06240				
				2012Q3	102.1	52.35013				
				2012Q4	110.6	27.35335				
				2013Q1	97.3	30.51681				
				2013Q2	91.8	71.76612				
				2013Q3	96.8	92.57584				
				2013Q4	102.2	113.8617				
				2014Q1	108.5	114.8054				
				2014Q2	98.1	20.22236				
				2014Q3				

Figura 9.11

Posteriormente, seleccionamos *View* → *Graph* damos clic y elegimos *Scatter* y damos OK, obteniendo el gráfico correspondiente, como se indica en la Figura 9.12 También obtenemos la serie del cuadrado de los residuos para la realización del respectivo gráfico (véase Figura 9.13). En las figuras mencionadas no se observa una dispersión creciente de los residuos a medida que aumenta el precio de la competencia, por lo tanto, no hay indicios de heteroscedasticidad.

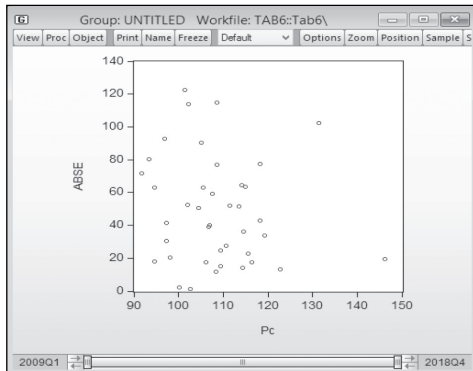


Figura 9.12

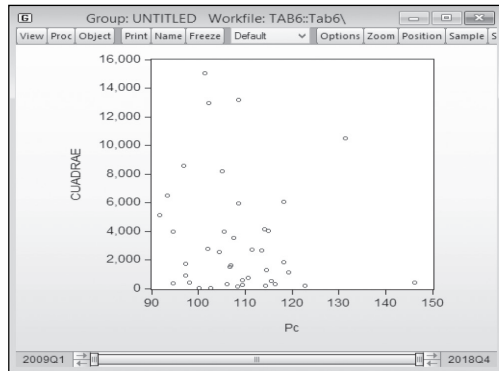


Figura 9.13

Enseguida, seleccionamos *View* → *Graph* damos clic y elegimos *Scatter* y pulsamos OK, obteniendo el gráfico que se muestra en la Figura 9.14. También obtenemos la serie del cuadrado de los residuos para la realización del respectivo gráfico (véase Figura 9.15). En estas figuras tampoco se observa una dispersión creciente de los residuos a medida que aumenta el gasto en publicidad, por consiguiente, no hay indicios de heteroscedasticidad.

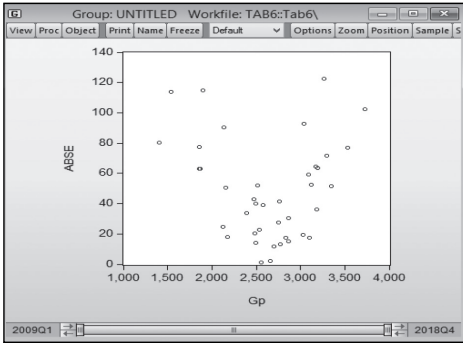


Figura 9.14

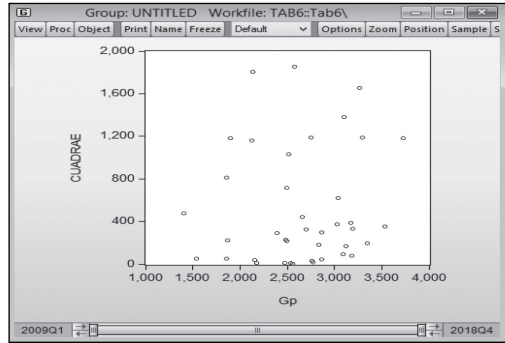


Figura 9.15

Por último, seleccionamos *View* → *Graph* damos clic y elegimos *Scatter* y pulsamos OK, obteniendo de este modo el gráfico que se muestra en la Figura 9.16. También obtenemos la serie del cuadrado de los residuos para la realización del respectivo gráfico (véase Figura 9.17). En estas figuras tampoco se observa una dispersión creciente de los residuos a medida que aumenta la renta, por consiguiente, no hay indicios de heteroscedasticidad.

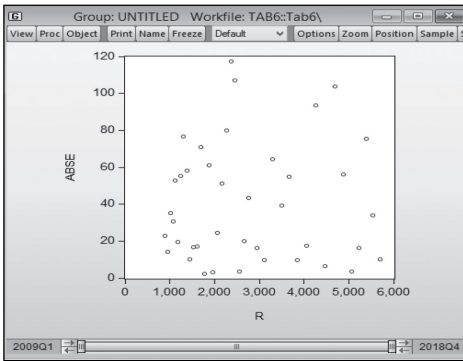


Figura 9.16

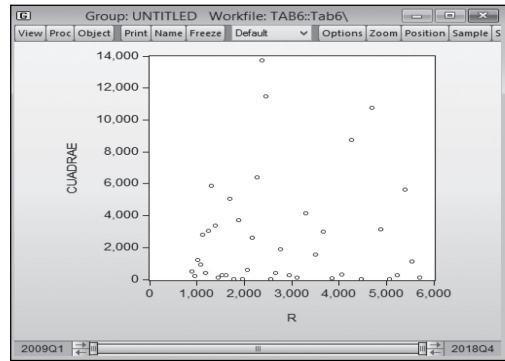


Figura 9.17

También pueden representarse los residuos contra los valores predichos para analizar la posible heteroscedasticidad. Para ello se estima la ecuación de regresión múltiple con los datos de la Tabla 3.1, seleccionando las variables en el siguiente orden: V, PE, PC, GP y R. Sobre la pantalla del objeto ecuación se calculan los valores predichos eligiendo *Proc* → *Forecast* como se indica en la Figura 9.18. En la pantalla *Forecast* que aparece a continuación rellenamos las opciones como se muestra en la Figura 9.19 (este vector contiene las predicciones de la variable dependiente Ventas) y damos OK.

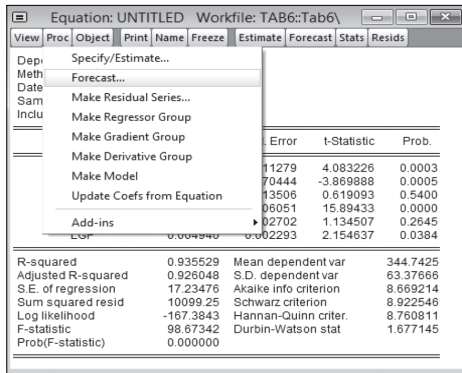


Figura 9.18

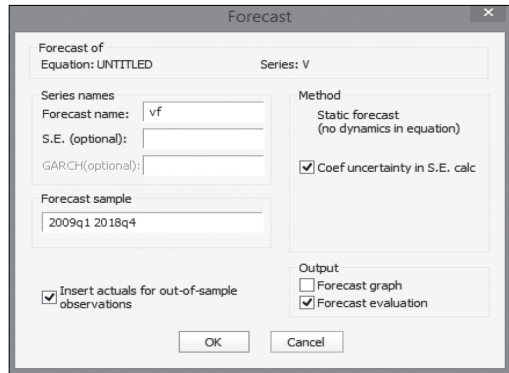


Figura 9.19

En la ventana del objeto ecuación seleccionamos *Quick* → *Graph* damos clic y en el cuadro de diálogo de *Series List* escribimos primero el vector de ventas proyectada (*vf*) contra los residuos y damos OK, como se muestra en las Figuras 9.20 y 9.21.

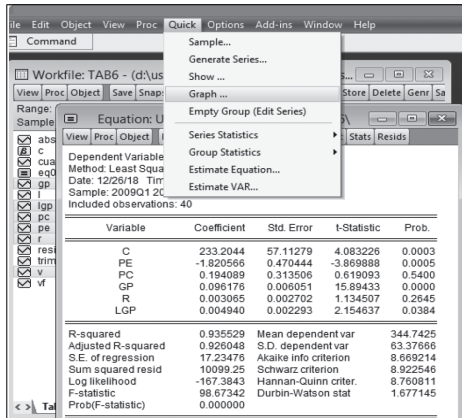


Figura 9.20

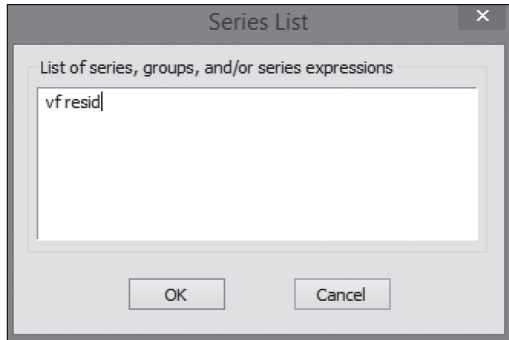


Figura 9.21

Luego, en la pantalla de *Graph Options* elegimos *Scatter* y damos OK (Figura 9.22). A continuación aparece el gráfico de las ventas proyectadas contra los residuos como se indica en la Figura 9.23.

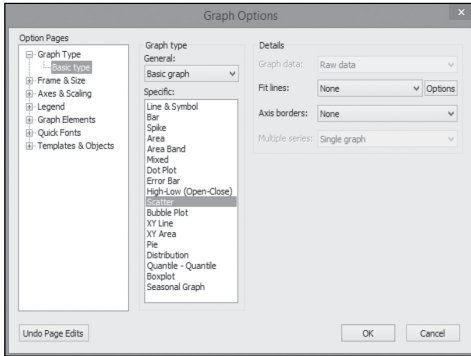


Figura 9.22

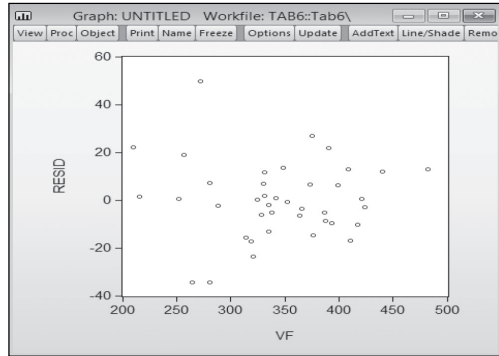


Figura 9.23

Repetimos el mismo procedimiento para obtener el gráfico de los residuos contra cada una de las variables explicativas del modelo. Para ello, en la ventana del objeto ecuación seleccionamos *Quick* → *Graph* damos clic y en el cuadro de diálogo de *Series List* editamos la variable precio de la editorial (PE) contra los residuos y pulsamos OK, como se muestra en la Figura 9.24. El gráfico de los residuos se visualiza en la Figura 9.25.

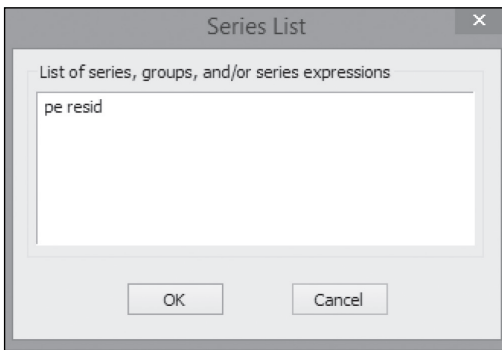


Figura 9.24

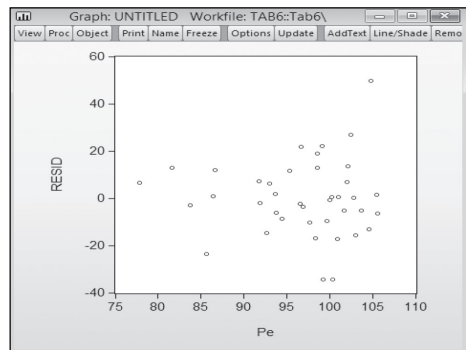


Figura 9.25

Luego obtenemos el gráfico de los residuos contra el precio de la competencia. Para ello, en la ventana del objeto ecuación seleccionamos *Quick* → *Graph* damos clic y en el cuadro de diálogo de *Series List* editamos la variable precio de la competencia (PC) contra los residuos y pulsamos OK, como se muestra en la Figura 9.26. El gráfico de los residuos se visualiza en la Figura 9.27.

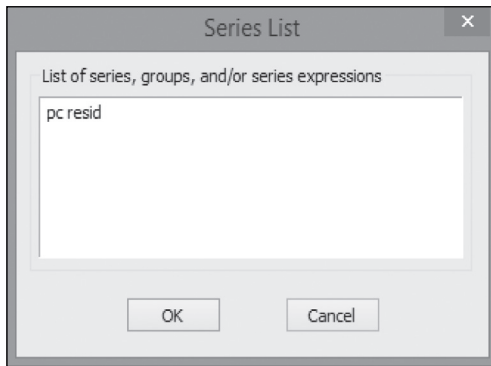


Figura 9.26

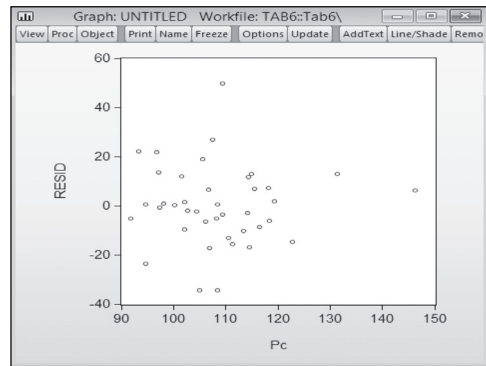


Figura 9.27

Posteriormente, obtenemos el gráfico de los residuos contra el gasto en publicidad. Para ello, en la ventana del objeto ecuación seleccionamos *Quick* → *Graph* damos clic y en el cuadro de diálogo de *Series List* editamos la variable gasto en publicidad (GP) contra los residuos y pulsamos OK, como se muestra en la Figura 9.28. El gráfico de los residuos se muestra en la Figura 9.29.

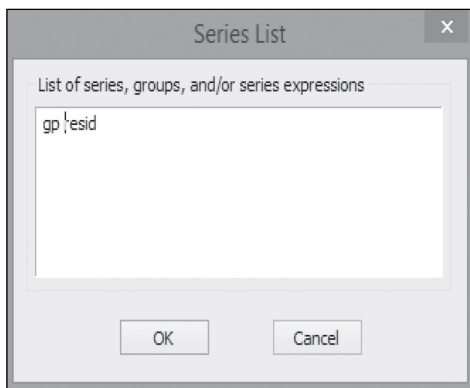


Figura 9.28

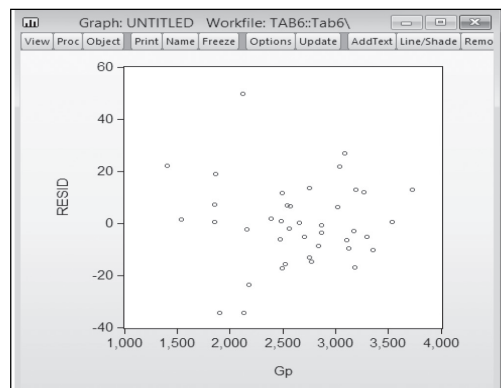


Figura 9.29

Por último, obtenemos el gráfico de los residuos contra la renta personal. Para ello, en la ventana del objeto ecuación seleccionamos *Quick* → *Graph* damos clic y en el cuadro de dialogo de *Series List* editamos la variable renta personal (R) contra los residuos y pulsamos OK, como se muestra en la Figura 9.30. El gráfico de los residuos se indica en la Figura 9.31.

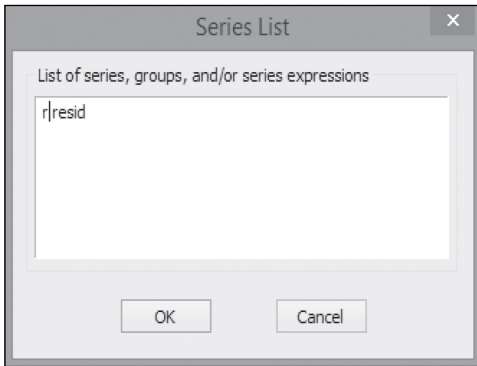


Figura 9.30

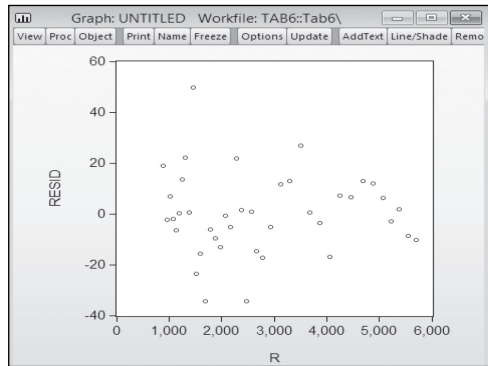


Figura 9.31

Al parecer la variable PE es la candidata a provocar los problemas de heteroscedasticidad, pero para estar más seguros utilizaremos un contraste formal, por ejemplo, el contraste de Glesjer. Para ello, realizamos una regresión múltiple de ABS(RESID) contra PE rellenando la pantalla *Equation Estimation*. Seleccionamos *Quick* → *Estimate equation* como se indica en la Figura 9.32. En el cuadro de diálogo que se abre escribimos *ABS(RESID) c pe* como se muestra en la Figura 9.33 y damos Aceptar.

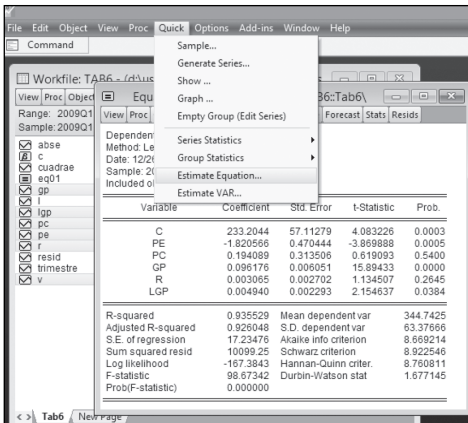


Figura 9.32

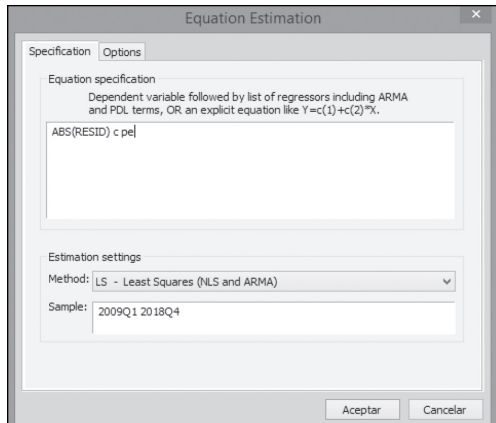


Figura 9.33

A continuación se obtienen los resultados de la Tabla 9.1.

Tabla 9.1***Contraste de Glesjer.***

Variable Dependiente: ABS(RESID)

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-15,75959	24,69627	-0,638136	0,5272
PE	0,283754	0,254595	1,114529	0,2721
R-squared	0,031654	Mean dependent var		11,69856
Adjusted R-squared	0,006171	S.D. dependent var		10,88989
S.E. of regression	10,85624	Akaike info criterion		7,656063
Sum squared resid	4.478,598	Schwarz criterion		7,740507
Log likelihood	-151,1213	Hannan-Quinn criterion		7,686595
F-statistic	1,242175	Durbin-Watson stat		2,182983
Prob(F-statistic)	0,272052			

Los coeficientes no resultan significativos al 94% (p-valores mayores que 0,06) y el modelo global tampoco es significativo (p-valor de la F mayor que 0,05). Por consiguiente, no se rechaza la nulidad de cualquier coeficiente incluida la constante. Según el contraste de Glesjer, la hipótesis de homoscedasticidad es equivalente a la hipótesis de pendiente nula en este modelo, por ende, no hay presencia de heteroscedasticidad provocada por la variable: precio de la editorial (PE).

9.3 Detección de la heteroscedasticidad mediante contrastes.

9.3.1 Contraste de White.

Para llevar a cabo el contraste de White, estimamos la regresión por MCO de los residuos al cuadrado contra el precio de la editorial (PE_i); para ello seleccionamos CUADRAE²⁵ y PE a través de *Open* \rightarrow *as Equation*, cuyo

²⁵ El lector recordará que CUADRAE representa los residuos al cuadrado de la regresión ventas (V) contra el precio de la editorial (PE), los cuales fueron obtenidos a través del procedimiento aprendido en la detección de la heteroscedasticidad por métodos gráficos.

resultado aparece en la Tabla 9.2. Luego, nos situamos en la ventana del objeto ecuación y seleccionamos *View* → *Residual Diagnostics* → *Heteroskedasticity Test* y damos clic. En la solapa *Test type* de la ventana que se abre elegimos White y damos OK como se muestra en la Figura 9.34.

Tabla 9.2

Detección de heteroscedasticidad (contraste de White).

Variable Dependiente: CUADRAE

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	358,6718	7.991,771	0,044880	0,9644
PE	23,23331	82,38761	0,282000	0,7795
R-squared	0,002088	Mean dependent var		2.606,901
Adjusted R-squared	-0,024172	S.D. dependent var		3471,397
S.E. of regression	3.513,103	Akaike info criterion		19,21509
Sum squared resid	4,69E+08	Schwarz criterion		19,29954
Log likelihood	-382,3019	Hannan-Quinn criterion		19,24563
F-statistic	0,079524	Durbin-Watson stat		1,557103
Prob(F-statistic)	0,779474			

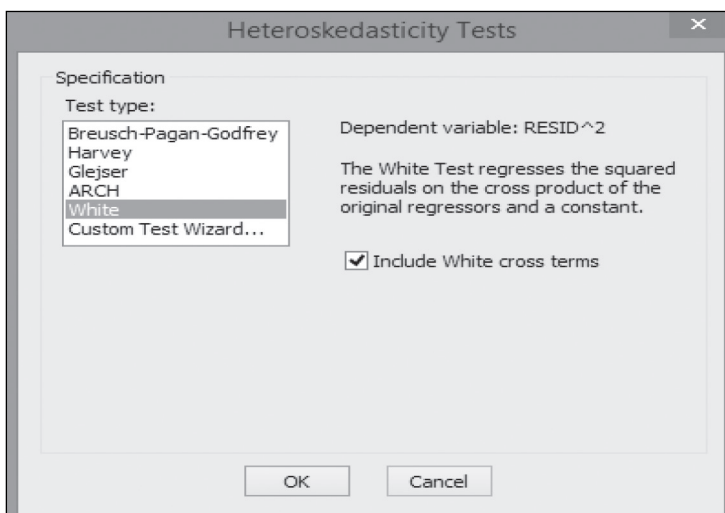


Figura 9.34

En la salida del Test de heteroscedasticidad de White (véase Tabla 9.3) se muestra el estadístico Obs*R-squared junto con el estadístico F (*F-statistic*), el cual contrasta la significación conjunta de las variables que aparecen como explicativas en la regresión auxiliar. También aparecen los resultados de la regresión auxiliar, lo que permite, entre otras cosas, identificar cuál de las dos opciones ha sido elegida. Los valores obtenidos no permiten rechazar la hipótesis de homoscedasticidad, puesto que la probabilidad de rechazar esta hipótesis siendo cierta es mayor de 5% (0,05), por consiguiente, no existe heteroscedasticidad en el modelo estimado.

F- statistic \rightarrow Prob F(2, 37) = 0,4059 > 0,05 \rightarrow Por lo tanto, se acepta la hipótesis de homoscedasticidad.

Obs*R-squared \rightarrow Prob Chi-Squared(2) = 0,3862 > 0,05 \rightarrow Por consiguiente, se acepta la hipótesis de homoscedasticidad.

Tabla 9.3

Prueba de heteroscedasticidad de White.

Heteroskedasticity Test White				
F- statistic	0,924004	Prob. F(2, 37)		0,4059
Obs*R-squared	1,902809	Prob. Chi-Square(2)		0,3862
Scaled explained SS	3,531070	Prob. Chi-Square(2)		0,1711
Test Equation:				
Variable Dependiente: RESID^2				
Método: MCO				
Muestra: 2009Q1 2018Q4				
Observaciones incluidas: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8,60E+08	6,24E+08	1,377936	0,1765
PE^2	97.408,20	71.826,22	1,356165	0,1833
PE	18.238.706	13.421.475	-1,358920	0,1824
R-squared	0,047570	Mean dependent var		11.724.796
Adjusted R-squared	-0,003912	S.D. dependent var		24.079.639
S.E. of regression	24.126.699	Akaike info criterion		36,90757
Sum squared resid	2,15E+16	Schwarz criterion		37,03424
Log likelihood	-735,1515	Hannan-Quinn criterion		36,95337
F-statistic	0,924004	Durbin-Watson stat		1,494602
Prob(F-statistic)	0,405891			

Con el propósito de verificar la no existencia de heteroscedasticidad en el resto de las variables explicativas del modelo, realizaremos el contraste de White a cada una de ellas, siguiendo el mismo procedimiento aprendido. Si aplicamos el test de White a la variable precio de la competencia (PC), los valores obtenidos no permiten rechazar la hipótesis de homoscedasticidad, puesto que la probabilidad de rechazar esta hipótesis siendo cierta es mayor de 5% (0,05), por consiguiente, no existe heteroscedasticidad (véase Tabla 9.4).

F- statistic \rightarrow Prob F(2, 37) = 0,9555 > 0,05 \rightarrow Por ende, se acepta la hipótesis de homoscedasticidad.

Obs*R-squared \rightarrow Prob Chi-Squared(2) = 0,9521 > 0,05 \rightarrow Por consiguiente, se acepta la hipótesis de homoscedasticidad.

Tabla 9.4

Prueba de heteroscedasticidad de White.

Heteroskedasticity Test White				
F- statistic	0,045541	Prob. F(2, 37)		0,9555
Obs*R-squared	0,098226	Prob. Chi-Square(2)		0,9521
Scaled explained SS	0,148573	Prob. Chi-Square(2)		0,9284
Test Equation:				
Variable Dependiente: RESID^2				
Método: MCO				
Muestra: 2009Q1 2018Q4				
Observaciones incluidas: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	50.127.184	3,03E+08	0,165417	0,8695
PC^2	1.642,915	23.115,88	0,071073	0,9437
PC	502.632,5	5.308.263	-0,094689	0,9251
R-squared	0,002456	Mean dependent var		1.511.776
Adjusted R-squared	-0,051466	S.D. dependent var		28.019.643
S.E. of regression	28.731.622	Akaike info criterion		37,25693
Sum squared resid	3,05E+16	Schwarz criterion		37,37360
Log likelihood	-742,1387	Hannan-Quinn criterion		37,30273
F-statistic	0,045541	Durbin-Watson stat		1,186276
Prob(F-statistic)	0,955534			

Continuamos con los residuos al cuadrado de la regresión Ventas (V_t) contra el gasto en publicidad (GP_t), obteniendo los resultados que se muestran en la Tabla 9.5. Los valores obtenidos no permiten rechazar la hipótesis de homoscedasticidad, puesto que la probabilidad de rechazar esta hipótesis siendo cierta es mayor de 5% (0,05), por consiguiente, no existe heteroscedasticidad.

F- statistic \rightarrow Prob F(2, 37) = 0,8520 > 0,05 \rightarrow Se acepta la hipótesis de homoscedasticidad.

Obs*R-squared \rightarrow Prob Chi-Squared(2) = 0,8417 > 0,05 \rightarrow Se acepta la hipótesis de homoscedasticidad.

Tabla 9.5

Prueba de heteroscedasticidad de White

Heteroskedasticity Test White				
F- statistic	0,160834	Prob. F(2, 37)		0,852
Obs*R-squared	0,344751	Prob. Chi-Square(2)		0,8417
Scaled explained SS	0,300948	Prob. Chi-Square(2)		0,8603
Test Equation:				
Variable Dependiente: RESID^2				
Método: MCO				
Muestra: 2009Q1 2018Q4				
Observaciones incluidas: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	347.651	1.229.161	-0,282836	0,7789
GP^2	0,106073	0,189026	-0,561158	0,5781
GP	533,2822	976,0095	0,546390	0,5881
R-squared	0,008619	Mean dependent var		289.947,7
Adjusted R-squared	-0,044969	S.D. dependent var		408.415,2
S.E. of regression	417.497,3	Akaike info criterion		28,79398
Sum squared resid	6,45E+12	Schwarz criterion		28,92065
Log likelihood	-572,8796	Hannan-Quinn criterion		28,83978
F-statistic	0,160834	Durbin-Watson stat		1,986516
Prob(F-statistic)	0,852026			

Por último, realizaremos el test de heteroscedasticidad de White de los residuos al cuadrado de la regresión Ventas (V_t) contra la renta personal (R_t). Los valores obtenidos no permiten rechazar la hipótesis de homoscedasticidad, puesto que la probabilidad de rechazar esta hipótesis siendo cierta es mayor de 5% (0,05), por consiguiente, no existe heteroscedasticidad (véase Tabla 9.6).

F- statistic \rightarrow Prob F(2, 37) = 0,3660 > 0,05 \rightarrow Por consiguiente, se acepta la hipótesis de homoscedasticidad.

Obs*R-squared \rightarrow Prob Chi-Squared(2) = 0,3473 > 0,05 \rightarrow Por tanto, se acepta la hipótesis de homoscedasticidad.

Tabla 9.6

Prueba de heteroscedasticidad de White

Heteroskedasticity Test White					
F- statistic	1,032999	Prob. F(2, 37)			0,366
Obs*R-squared	2,115393	Prob. Chi-Square(2)			0,3473
Scaled explained SS	3,966036	Prob. Chi-Square(2)			0,1377
Test Equation:					
Variable Dependiente: RESID^2					
Método: MCO					
Muestra: 2009Q1 2018Q4					
Observaciones incluidas: 40					
Variable	Coefficient	Std. Error	t-Statistic		Prob.
C	13.829.160	18.171.361	-0,761042		0,4515
R^2	2,772561	2,089531	-1,326882		0,1927
R	19.044,12	13.522,55	1,408323		0,1674
R-squared	0,052885	Mean dependent var			11.700.073
Adjusted R-squared	0,001689	S.D. dependent var			24.152.409
S.E. of regression	24.131.999	Akaike info criterion			36,90801
Sum squared resid	2,10E+16	Schwarz criterion			37,03468
Log likelihood	-735,1603	Hannan-Quinn criterion			36,95381
F-statistic	1,032999	Durbin-Watson stat			1,330793
Prob(F-statistic)	0,365975				

9.3.2 Contraste de Breusch-Pagan-Godfrey.

A través del test de Breusch-Pagan se analiza si la varianza estimada de los residuos del modelo depende de los valores de las variables explicativas. Con las restricciones del método de mínimos cuadrados ordinarios la media de la varianza estimada de los residuos debe ser cero, de modo que dada la suposición de que la varianza de los residuos no depende de las variables explicativas, la estimación de la varianza se puede obtener a partir de la media de los residuos al cuadrado. Si la suposición no fuera correcta, podría ocurrir que la varianza estuviera relacionada linealmente con las variables explicativas, por lo tanto, los estimadores del modelo no serían eficientes ni para explicar un fenómeno económico ni tampoco para predecir, que en últimas, es el propósito de la econometría.

Para llevar a cabo el contraste de Breusch-Pagan-Godfrey en el modelo de la editorial, se lleva a cabo la regresión con todas las variables explicativas del modelo (si la heteroscedasticidad depende sólo del precio de la editorial, aparecerá este hecho en la regresión auxiliar). Así pues, se realiza la regresión de $\hat{\mu}_t^2$ sobre una constante y las variables PE_t , PC_t , GP_t , R_t y LGP_t , mediante el siguiente modelo:

$$\hat{\mu}_t^2 = \beta_0 + \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t + \beta_5 LGP_t \quad (9.1)$$

Al estimar el modelo, se obtiene:

$$\hat{\mu}_t^2 = 233,204 - 1,821PE_t + 0,194PC_t + 0,096GP_t + 0,003R_t + 0,0049LGP_t \quad (9.2)$$

(57,11) (0,47) (0,31) (0,006) (0,003) (0,002)

$$R^2 = 93,55\%; \quad \bar{R}^2 = 92,60\%; \quad \Sigma \hat{\mu}_t^2 = 10.099,25$$

El estadístico toma entonces un valor igual a $40 \times 0,126 = 5,04$, el cual es menor que el valor crítico de 9,488 del 5% de significación encontrado en las tablas de la distribución chi-cuadrado, $\chi^2(4)$. Por lo tanto, no se puede rechazar al nivel de significación de 5%, la hipótesis nula de heteroscedasticidad, en otras palabras, se acepta la hipótesis de homoscedasticidad. Podemos

concluir que en nuestro modelo no existe ningún problema importante de heteroscedasticidad en los datos utilizados.

La aplicación en EViews del contraste de heteroscedasticidad de Breusch-Pagan-Godfrey (considerado como una regresión de los residuos al cuadrado en la regresión original) se realiza del siguiente modo: en la ventana del objeto ecuación del modelo estimado z , seleccionamos *View* → *Residual Diagnostics* → *Heteroskedasticity Test* y damos clic como se muestra en la Figura 9.35. En la solapa de especificación (*Specification*) elegimos el contraste Breusch-Pagan-Godfrey y damos OK, como se indica en la Figura 9.36.

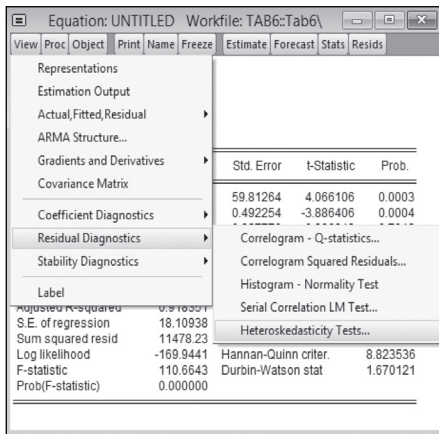


Figura 9.35

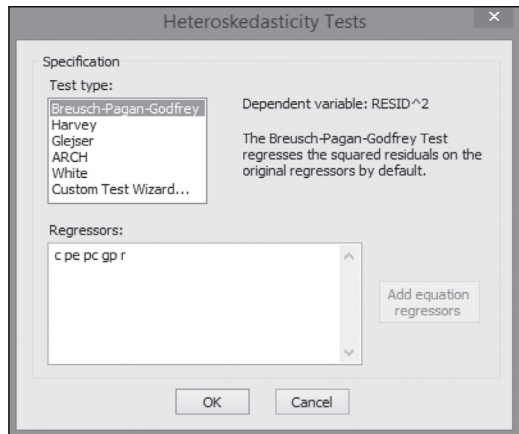


Figura 9.36

En la Tabla 9.7 se visualizan los estadísticos de contraste como el estadístico F (F-statistic) de contraste conjunto, el R cuadrado de las observaciones y el Scaled explained SS junto con sus probabilidades. Como se puede ver los tres estadísticos tiene una probabilidad superior al 5% (F-statistic → 0,3397 > 0,05; Obs*R-squared → 0,3165 > 0,05 y Scaled explained SS → 0,3297 > 0,05); por tanto, no rechazamos la hipótesis nula de homoscedasticidad. En conclusión, este contraste indica la no presencia de heteroscedasticidad en las perturbaciones.

Tabla 9.7***Prueba de heteroscedasticidad de Breusch-Pagan-Godfrey.***

Heteroskedasticity Test Breusch-Pagan-Godfrey				
F- statistic	1,172470	Prob. F(2, 37)		0,3397
Obs*R-squared	4,726525	Prob. Chi-Square(2)		0,3165
Scaled explained SS	4,610258	Prob. Chi-Square(2)		0,3297
Test Equation:				
Variable Dependiente: RESID^2				
Método: MCO				
Muestra: 2009Q1 2018Q4				
Observaciones incluidas: 40				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	762,3239	1.518,774	-0,501934	0,6189
PE	15,52572	12,49941	1,242116	0,2225
PC	0,238445	8,322960	-0,028649	0,9773
GP	0,143936	0,153137	-0,153137	0,3537
R	0,016644	0,071916	-0,231436	0,8183
R-squared	0,118163	Mean dependent var		286,9558
Adjusted R-squared	0,017382	S.D. dependent var		463,8860
S.E. of regression	459,8367	Akaike info criterion		15,21609
Sum squared resid	7.400.744	Schwarz criterion		15,42720
Log likelihood	-299,3218	Hannan-Quinn criterion		15,29242
F-statistic	1,172470	Durbin-Watson stat		2,399420
Prob(F-statistic)	0,339726			

Hubo un momento en el cual Don Pablo estaba preocupado por el resultado del análisis gráfico de los residuos, pero Emilio sabiamente le recuerda que dicho análisis es válido como orientación hacia posibles problemas de especificación que puede tener el modelo, pero solo los contrastes de hipótesis son válidos para tomar decisiones.

Dados los resultados, no del todo concluyentes, de los contrastes de heteroscedasticidad realizados anteriormente, en el modelo de ventas de la editorial parece razonable utilizar el estimador original de los coeficientes obtenido por MCO. No obstante, y con el objeto de garantizar que los

contrastes realizados tienen, al menos, validez asintótica, Emilo ha reestimado las varianzas de los coeficientes utilizando el estimador robusto de White. Para ello, abrimos en Eviews el archivo de la Tabla 7.6, posteriormente, estimamos la regresión de las ventas de la editorial (V) contra el precio promedio de la editorial (PE), el precio de la competencia (PC), el gasto en publicidad (GP), la renta personal (R) y la variable que combina horas semanales de lectura y gasto en publicidad (LGP). Enseguida, en la ventana del objeto ecuación pulsamos *Estimate* en la barra de herramientas y en el cuadro de diálogo que se abre (*Estimation Equation*), pulsamos en la pestaña *Options* y en *Covariance method* elegimos la opción *Huber-White* y damos Aceptar, como se indica en la Figura 9.37. Este procedimiento nos permite obtener los errores estándar consistentes (*White-Hinkley (HC) heteroskedasticity consistent standard errors and covariance*) a través del estimador robusto de White como se observa en la Tabla 9.8.

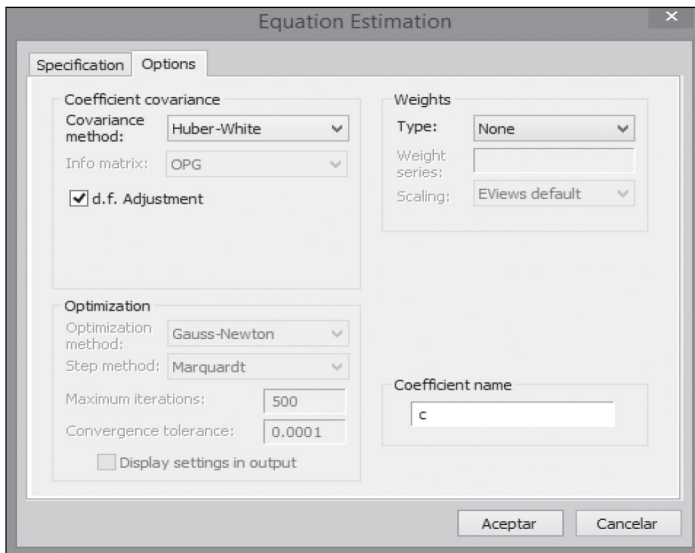


Figura 9.37

Tabla 9.8***Estimador robusto de White.***

Variable Dependiente: V

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	233,2044	52,26232	4,462190	0,0001
PE	-1,820566	0,424088	-4,292895	0,0001
PC	0,194089	0,234499	0,827675	0,4136
GP	0,096176	0,005399	17,81487	0,0000
R	0,003065	0,001810	1,693496	0,0995
LGP	0,004940	0,001876	2,632886	0,0126
R-squared	0,935529	Mean dependent var		344,7425
Adjusted R-squared	0,926048	S.D. dependent var		63,37666
S.E. of regression	7,23476	Akaike info criterion		8,669215
Sum squared resid	10,099,25	Schwarz criterion		8,922546
Log likelihood	-167,3843	Hannan-Quinn criterion		8,760811
F-statistic	98,67342	Durbin-Watson stat		1,677145
Prob(F-statistic)	0,00000	Wald F-statistic		81,29737
Prob(Wald F-statistic)	0,00000			

Si comparamos los coeficientes estimados de las Tablas 9.8 y 9.9 concluimos que son exactamente iguales, pero los valores de la columna de los errores estándar (*Std. Error*) se han modificado un poco. El estimador robusto de *White-Hinkley* hace que las desviaciones típicas de cada uno de los coeficientes se reduzcan, y por ende, la varianza de los residuos estimados. Cuando no se presupone la existencia de heteroscedasticidad, se comprueba que las varianzas estimadas de los coeficientes difieren muy poco (por ejemplo, el error estándar del coeficiente del precio de la editorial se reduce de 0,470444 a 0,424088) permaneciendo constantes los coeficientes de determinación normal y ajustado y la suma de los cuadrados de los residuos. Por lo tanto, puede concluirse que no parece que el modelo de ventas de la editorial presente problemas de heteroscedasticidad importantes.

Tabla 9.9***Estimador robusto de White-Hinkley.***

Variable Dependiente: V

Método: MCO

Muestra: 2009Q1 2018Q4

Observaciones incluidas: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	233,2044	57,11279	4,083226	0,0003
PE	-1,820566	0,470444	-3,869888	0,0005
PC	0,194089	0,313506	0,619093	0,5400
GP	0,096176	0,006051	15,89433	0,0000
R	0,003065	0,002702	1,134507	0,2645
LGP	0,004940	0,002293	2,154637	0,0384
R-squared	0,935529	Mean dependent var		344,7425
Adjusted R-squared	0,926048	S.D. dependent var		63,37666
S.E. of regression	17,23476	Akaike info criterion		8,669214
Sum squared resid	10.099,25	Schwarz criterion		8,922546
Log likelihood	-167,3843	Hannan-Quinn criterion		8,760811
F-statistic	98,67342	Durbin-Watson stat		1,677145
Prob(F-statistic)	0,00000			

Aplicando el estimador robusto de White-Hinkley se obtiene el siguiente resultado:

$$\widehat{Ventas}_t = 233,204 - 1,821PE_t + 0,194PC_t + 0,096GP_t + 0,003R_t + 0,0049LGP_t \quad (9.3)$$

$$(52,26) \quad (0,42) \quad (0,23) \quad (0,005) \quad (0,002) \quad (0,002)$$

$$R^2 = 93,55\%; \quad \bar{R}^2 = 92,60\%; \quad \Sigma \hat{\mu}^2 = 10.099,25$$

RETO 8:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

AUTOCORRELACIÓN

Supuesto 2: Las variable explicativas X_i y el término de error son independientes, esto es, $cov(X_i, \mu_i) = 0$.

Supuesto 5: No hay autocorrelación entre las perturbaciones: Dados dos valores cualesquiera de X , X_i y X_j ($i \neq j$), la correlación entre dos μ_i y μ_j cualesquiera ($i \neq j$) es cero. En pocas palabras, estas observaciones se muestrean de manera independiente. Simbólicamente, $cov(\mu_i, \mu_j | X_i, X_j) = 0$.

La autocorrelación es un caso particular del modelo de regresión generalizado que se produce cuando las perturbaciones del modelo presentan correlaciones entre ellas. La autocorrelación supone que la matriz de varianzas y covarianzas de las perturbaciones presentan valores distintos de cero en los elementos que están fuera de la diagonal principal (Gujarati, 2004 Griffiths y Judge, 1993).

10.1 Detección de la autocorrelación mediante métodos gráficos.

Carrascal et al, (2004), señalan que existen diferentes procedimientos para detectar correlaciones entre las perturbaciones. Dado que estas no son observables, las variables que se utilizan con los residuos mínimo cuadráticos. A partir de ellos, se construyen representaciones gráficas y/o se realizan contrastes específicos. Los gráficos de los residuos mínimo cuadráticos no son un instrumento definitivo para detectar autocorrelación en el modelo, pero sí un indicador muy útil de la presencia de problemas de especificación en el

mismo (entre ellos, el de autocorrelación) cuando reflejan comportamiento sistemáticos. Las representaciones más ilustrativas que inducen a pensar en la presencia de correlaciones entre las perturbaciones son:

- El gráfico de los residuos frente al tiempo, o si trabajamos con datos transversales, frente a alguna variable previamente ordenada.
- El gráfico de los residuos frente a sí mismos retardados un periodo.
- Los correlogramas de las funciones de autocorrelación simple y parcial de los residuos.

La forma de realizar en Eviews los gráficos anteriores parte de la estimación inicial de un modelo por mínimos cuadrados ordinarios, lo que origina al mismo tiempo el objeto serie resid con los valores de los residuos de dicha estimación. Para ello vamos a partir de la estimación del modelo completo (7.7).

$$\widehat{Ventas} = 233,204 - 1,821PE_t + 0,194PC_t + 0,096GP_t + 0,003R_t + 0,0049LGP_t$$

$$(57,11) \quad (0,47) \quad (0,31) \quad (0,006) \quad (0,003) \quad (0,002)$$

$$R^2 = 93,55\%; \quad \bar{R}^2 = 92,60\%; \quad \Sigma \hat{u}^2 = 10.099,25$$

Abrimos el archivo de trabajo *Tab6*, seleccionamos las variables V, PE, PC, GP, R y LGP pulsamos la tecla CTRL y el botón izquierdo del ratón abrimos como ecuación (*Open* → *as Equation*) y estimamos el modelo por MCO como se muestra en la Tabla 7.6. Desde el objeto ecuación visualizamos la serie de los residuos frente al tiempo mediante *View* → *Actual, Fitted, Residual* → *Actual, Fitted, Residual Table* y damos clic, como se indica en la Figura 10.1. En la Figura 10.2 se incluyen los valores observados y estimados de la variable dependiente (ventas) y sus residuos.

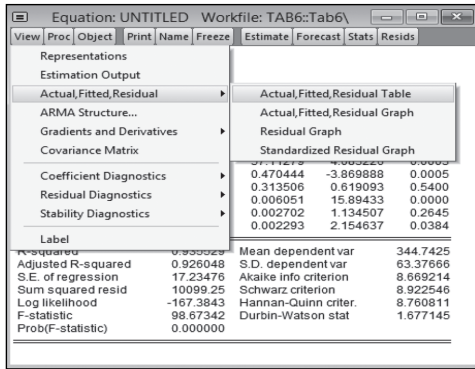


Figura 10.1

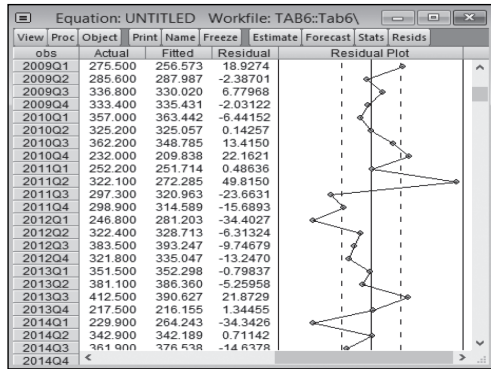


Figura 10.2

Mediante *View* → *Actual, Fitted, Residual* → *Actual, Fitted, Residual Graph* se visualizan en forma gráfica los valores observados, estimados y los residuos (véase Figura 10.3). Se aprecia que los residuos (Figura 10.4) no se comportan de forma totalmente aleatoria (en algunos periodos los residuos se salen de la banda que oscila entre +20 y -20), aunque tampoco se observa una senda de signos continuamente alternantes o grupos de residuos de un signo seguidos de otros de signo contrario. Esto genera indicios de existencia autocorrelación.

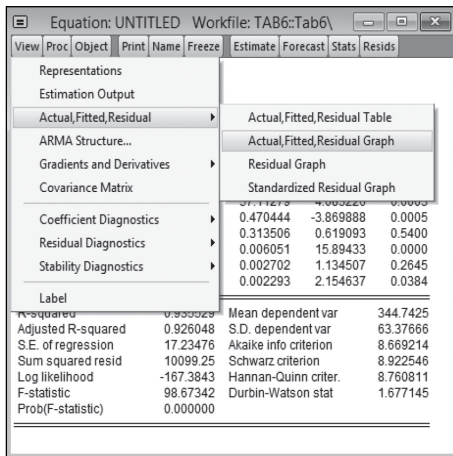


Figura 10.3

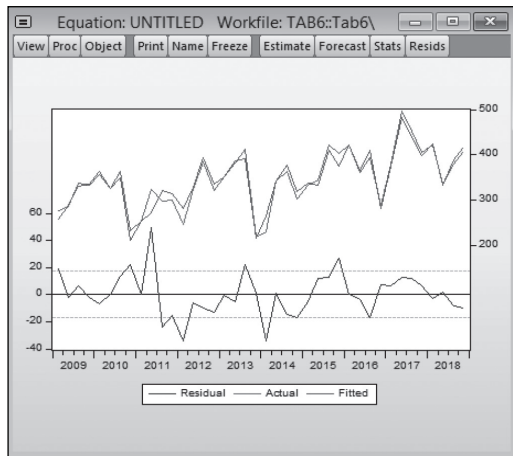


Figura 10.4

El gráfico de los residuos frente a los residuos retardados puede visualizarse si creamos un objeto grupo con dichas series. Para ello, una posibilidad es pulsar el botón *Show* en la barra de herramientas del Workfile y teclear en el

cuadro de diálogo que aparece *resid resid(-1)* y damos OK (véase Figura 10.5), lo cual genera un objeto grupo con dichas series (Figura 10.5). Posteriormente, sobre el objeto series seleccionamos la opción *View* → *Graph*.

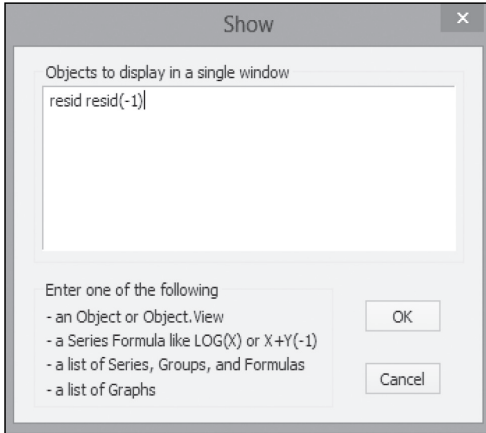


Figura 10.5

View	Proc	Object	Print	Name	Freeze	Default	Sort	Edit +/-	Smj
		RESID		RESID(-1)					
		2009Q1	18.92745		NA				
		2009Q2	-2.387005		18.92745				
		2009Q3	6.779676		-2.387005				
		2009Q4	-2.031223		6.779676				
		2010Q1	-6.441516		-2.031223				
		2010Q2	0.142573		-6.441516				
		2010Q3	13.41503		0.142573				
		2010Q4	22.16212		13.41503				
		2011Q1	0.486358		22.16212				
		2011Q2	-49.81501		0.486358				
		2011Q3	-23.66308		-49.81501				
		2011Q4	-15.68926		-23.66308				
		2012Q1	-34.40265		-15.68926				
		2012Q2	-6.313237		-34.40265				
		2012Q3	-9.746793		-6.313237				
		2012Q4	-13.24699		-9.746793				
		2013Q1	-0.798373		-13.24699				
		2013Q2	-5.259584		-0.798373				
		2013Q3	2.187292		-5.259584				
		2013Q4	1.344553		2.187292				
		2014Q1	-34.34259		1.344553				
		2014Q2	0.711419		-34.34259				
		2014Q3							

Figura 10.6

Posteriormente, sobre el objeto series seleccionamos la opción *View* → *Graph* y damos clic (Figura 10.7), en la venta de *Graph Options* que se abre elegimos *Scatter* y damos OK, como se indica en la Figura 10.8.

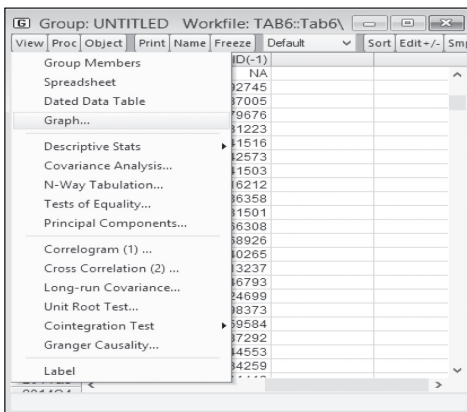


Figura 10.7

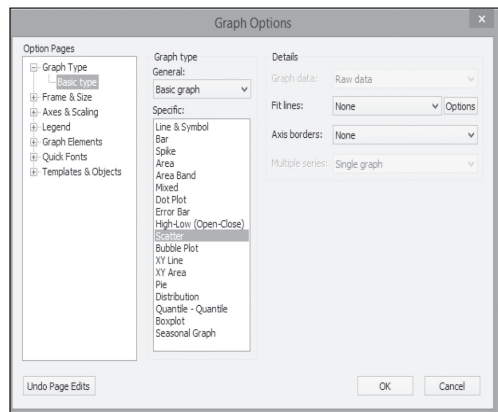


Figura 10.8

En el gráfico obtenido (Figura 10.9) puede observarse que la mayoría de los puntos se encuentra ubicados en el primer y tercer cuadrante, lo que nos llevaría a pensar en la posible existencia de autocorrelación según un

esquema AR(1) con coeficiente positivo. Finalmente, los correlogramas de las funciones de autocorrelación simple y parcial podemos crearlos desde la ventana del objeto ecuación seleccionando *View* → *Residual Diagnostics* → *Correlogram – Q-statistics*, como se muestra en la Figura 10.10 y enseguida, en la ventana que se abre damos OK como se muestra en la Figura 10.11.

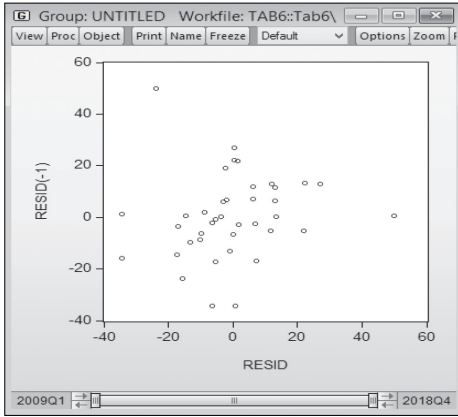


Figura 10.9

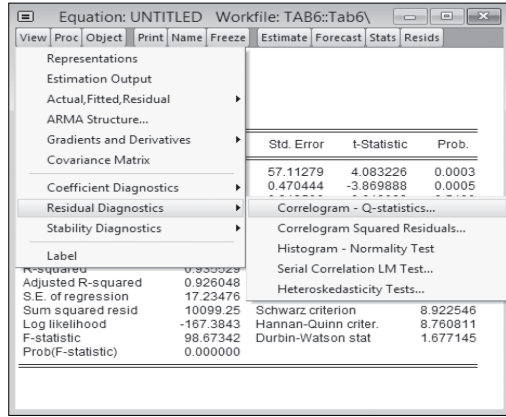


Figura 10.10

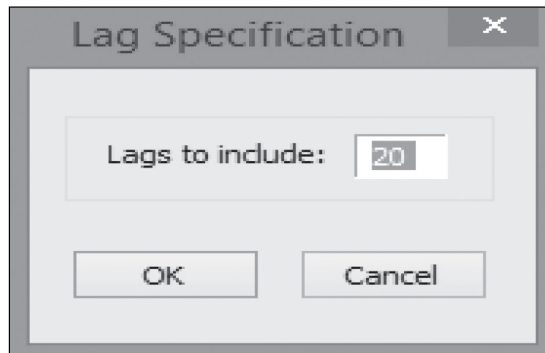


Figura 10.11

Los correlogramas de la serie original de los residuos para 20 retardos se visualizan en la Figura 10.12. Los retardos de autocorrelación (*Autocorrelation*) y correlación simple (*Partial Correlation*) se encuentran dentro de las bandas respectivas. Este comportamiento gráfico genera indicios de no existencia de autocorrelación en las perturbaciones.

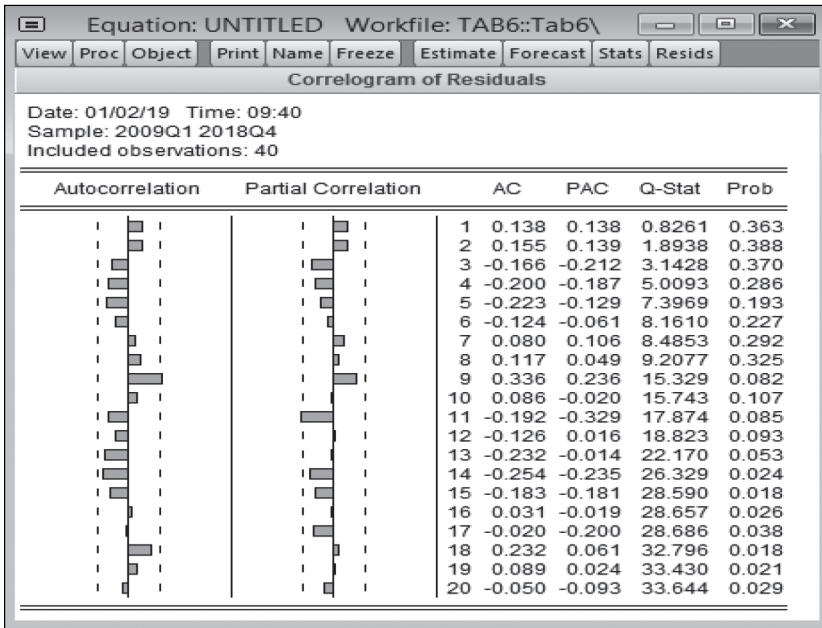


Figura 10.12

10.2 Detección de la autocorrelación mediante contrastes.

10.2.1 El estadístico de Durbin-Watson.

La hipótesis nula del contraste de Durbin-Watson plantea ausencia de autocorrelación, mientras que la hipótesis alternativa considera la existencia de correlación serial mediante un esquema autorregresivo del orden 1, AR(1). Es un contraste de una cola, según se establezca que el valor del coeficiente de autocorrelación de primer orden sea positivo o negativo.

$$H_0: \rho = 0$$

$$H_a: \rho < 0 \text{ ó } \rho > 0$$

El estadístico de contraste se define como:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Y puede expresarse en términos de coeficiente de autocorrelación muestral, $\hat{\rho}$, a partir de $d = 2(1-\hat{\rho})$

Según lo anterior, podemos observar que el estadístico d de Durbin-Watson toma valores entre 0 y 4, de forma que si no hay correlación serial ($\hat{\rho} \approx 0$), su valor estará alrededor de 2, mientras que valores cercanos a 0 indicarán la presencia de autocorrelación positiva ($\hat{\rho} \approx 1$) y valores cercanos a 4 mostrarán autocorrelación negativa ($\hat{\rho} \approx -1$).

El contraste se realiza de la siguiente forma:

- Si en la hipótesis alternativa se contrasta autocorrelación positiva de primer orden ($d < 2$) y el estadístico d es inferior a d_L , entonces se rechaza la hipótesis nula de ausencia de autocorrelación. Si por el contrario, si el estadístico supera el valor de d_U , no se puede rechazar la hipótesis nula de incorrelación. Finalmente, valores de d entre d_L y d_U , no permite extraer ninguna conclusión.
- Si en la hipótesis alternativa se contrasta autocorrelación negativa ($d < 2$) el procedimiento es como el anterior, pero considerando el estadístico como $4-d$.

En nuestro modelo el valor del estadístico d de Durbin-Watson es igual a 1,677 como aparece en la estimación de la ecuación (Figura 10.13) con el nombre *Durbin-Watson stat*. Considerando que $\alpha = 0,05$; $T = 40$ y $K' = 5$, en las tablas estadísticas utilizando GRETL se obtiene que $d_L = 1,2305$ y $d_U = 1,7859$ (véase Figura 10.14), por consiguiente, $1,2305 > 1,677 < 1,7859$, lo que no permite extraer ninguna conclusión, aunque $d = 1,677 \approx 2$, se podría considerar que el problema de autocorrelación no es tan grande.

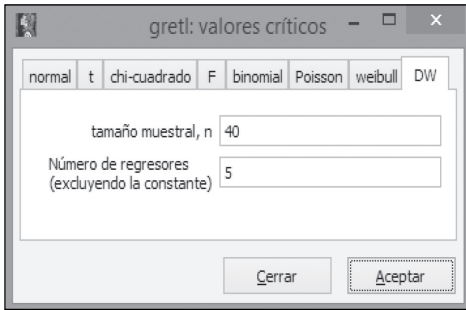


Figura 10.13

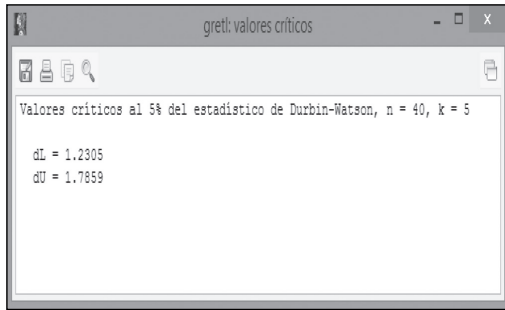


Figura 10.14

10.2.2 Contraste de Breusch-Godfrey.

Desde la ventana del objeto ecuación estimada seleccionamos *View* → *Residual Diagnostics* → *Serial Correlation LM Test* como se muestra en la Figura 10.15. En el cuadro de diálogo que aparece posteriormente (Figura 10.16) introducido el número de retardos (*Lags to include*) de los residuos que queremos que se incluyan en la regresión auxiliar mencionada arriba. Por defecto, Eviews toma 2 retardos. Este valor no es más que el número del orden de la correlación serial que deseamos contrastar.

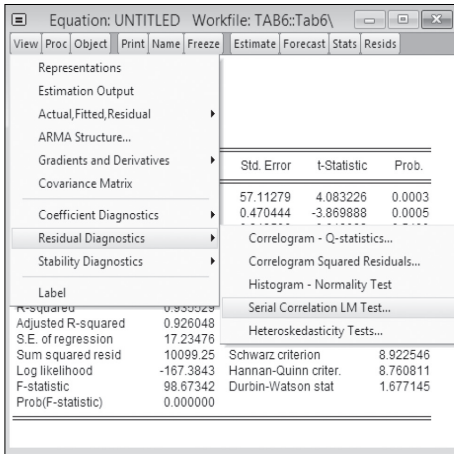


Figura 10.15

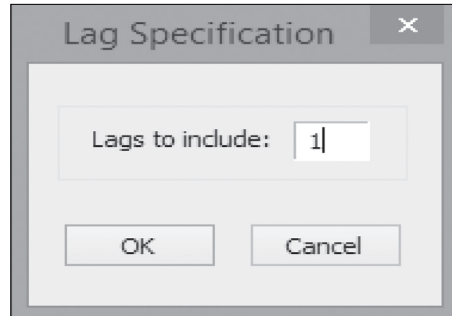


Figura 10.16

El resultado muestra dos estadísticos: El estadístico *Breusch-Godfrey*, *Obs*R-squared* y el llamado estadístico (*F-statistic*) de significación conjunta de los residuos retardados que, en realidad, presenta una distribución

no conocida, pero que se utiliza de forma informal también para realizar el contraste. Adicionalmente, incluye la estimación de la regresión auxiliar, lo que permite observar los retardos que tienen algún poder explicativo sobre los residuos.

En nuestro caso, la realización del contraste Breusch-Godfrey para un retardo muestra un valor del estadístico igual a 0,842, lo que permite no rechazar la hipótesis nula de ausencia de correlación serial igual al 5% de significación mediante un esquema AR(1) y MA(1), como se muestra en la Tabla 10.1.

Tabla 10.1

Breusch-Godfrey Serial Correlation LM Test.

F- statistic	0,70986	Prob. F(1, 33)	0,4056
Obs*R-squared	0,842318	Prob. Chi-Square(1)	0,3587

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Sample: 2009Q1: 2018Q4

Included observations: 40

Presample missing value lagged residuals set to zero

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9,642007	56,48857	-0,164853	0,8701
PE	0,046884	0,47573	0,098551	0,9221
PC	0,041039	0,318598	0,128813	0,8983
GP	-0,000186	0,006108	-0,068267	0,9327
R	-0,000343	0,002722	-0,146471	0,9460
RESID(-1)	0,152058	0,180478	0,842532	0,8844
R-squared	0,021058	Mean dependent var		3,09E-14
Adjusted R-squared	-0,156932	S.D. dependent var		16,09208
S.E. of regression	17,30877	Akaike info criterion		8,697931
Sum squared resid	9.886,582	Schwarz criterion		8,993485
Log likelihood	-166,9586	Hannan-Quinn criterion		8,804794
F-statistic	0,118310	Durbin-Watson stat		1,987721
Prob(F-statistic)	0,993492			

10.2.3 Contraste de Ljung-Box.

De acuerdo con Carrascal et al, (2004:276), el estadístico de Ljung-Box, que denotaremos por Q , sirve para contrastar esquemas generales de autocorrelación de la perturbación. La hipótesis nula continúa siendo la ausencia de correlación serial, mientras que la hipótesis alternativa puede plantearse en términos de cualquier esquema ARMA (p.q.).

Ho: Ausencia de autocorrelación ($p_1 = p_2 = \dots = p_r = 0$)

Ha: ARMA(p.q.)

El estadístico se define como:

$$Q = T(T+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{T-j} \rightarrow \chi^2_{r-p-q}$$

Y se distribuye asintóticamente como un chi-cuadrado con $r-p-q$ grados de libertad, siendo T el número de observaciones utilizadas en la estimación de la ecuación con la que estamos trabajando, p y q el orden del proceso de autocorrelación propuesto y $\hat{\rho}_j$ el coeficiente de autocorrelación muestral de orden j de los residuos mínimo cuadráticos.

EViews calcula el estadístico Q junto con los correlogramas de las funciones muestrales de autocorrelación simple y parcial de los residuos mínimo cuadráticos para cada uno de los retardos. Su ejecución se realiza en la barra de herramientas de la ventana del objeto ecuación estimada, para ello seleccionamos *View* → *Residual Diagnostics – Correlogram – Q statistic*, como se muestra en la Figura 10.17. En el cuadro de diálogo que se abre digitamos 20 y damos OK como se indica en la Figura 10.18. El número de retardos que escribamos en el cuadro de diálogo indica, además de la cantidad de coeficientes a calcular, el número de contrastes a realizar, ya que, para cada retardo, EViews calcula el valor del estadístico Q y su probabilidad (p-valor), es decir, la probabilidad que deja dicho estadístico en el extremo derecho de la distribución chi-cuadrado. De esta forma, podemos contrastar, para cada uno de esos retardos, la hipótesis nula de que no hay autocorrelación de dicho orden. El caso de un modelo en el que las perturbaciones no estuvieran correlacionadas debería presentar en todos los retardos valores muy pequeños del estadístico Q y grandes de su probabilidad (superiores a 0,05 si el nivel de significación es de 5%).

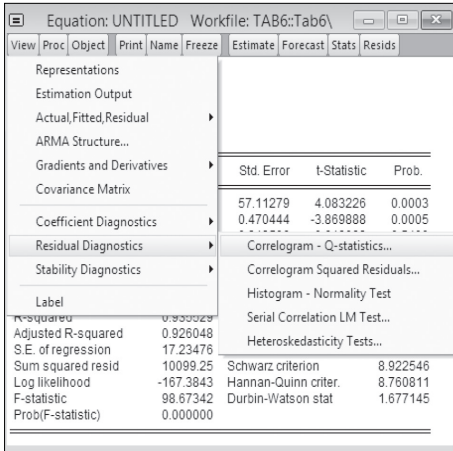


Figura 10.17

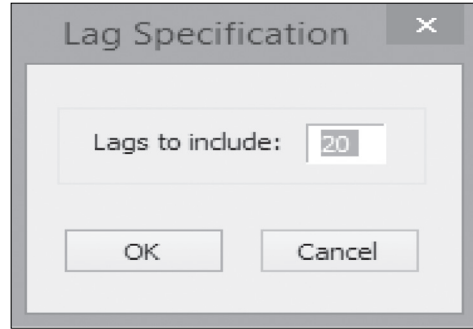


Figura 10.18

En nuestro modelo, los retardos del 1 al 13 presentan un Q-stat pequeño cuya probabilidad (Prob) es superior a 0,05, en cambio, en los retardos del 14 al 20 los valores del Q-stat son más grandes y su probabilidad es inferior a 0,05. Por consiguiente, los resultados permiten no rechazar la hipótesis nula de incorrelación serial. Este hecho es coherente con el gráfico de autocorrelación y correlación parcial donde los coeficientes se encuentran dentro de las bandas respectivas, lo cual confirma la no existencia de autocorrelación, tal como se aprecia en la Figura 10.19.

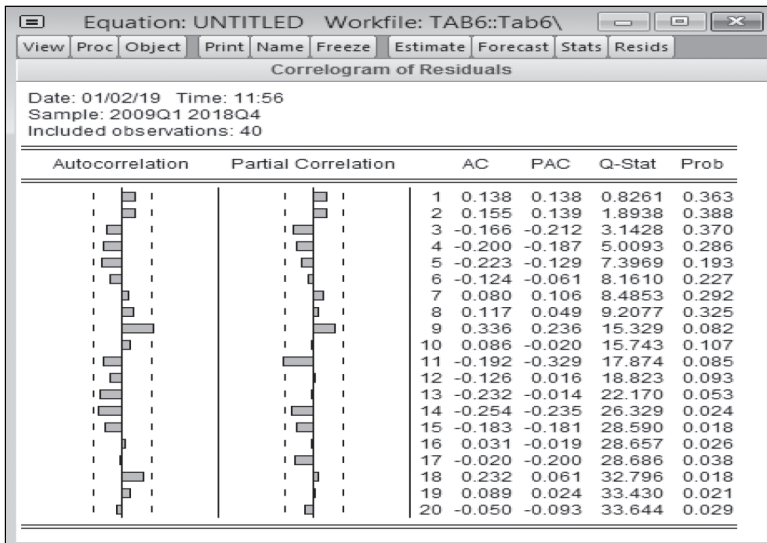


Figura 10.19

Don Pablo comenta a su hijo que cree entender ya por qué, a la hora de discutir los resultados numéricos obtenidos para el modelo de ventas de la editorial, es necesario tener en cuenta las hipótesis que se hayan establecido sobre los diferentes elementos del mismo, puesto que aquéllas son la base de la confianza en los métodos utilizados y, por tanto, en los residuos obtenidos. Por esta razón, Don Pablo, que como veremos aprende rápido, hace el siguiente comentario:

Don Pablo: Dado que nuestras ventas de un mes a otro, o de un trimestre a otro, no son algo errático o aleatorio, sino que, por el contrario, podemos decir que presentan un a cierta inercia, ¿no significa esto que las ventas en los diferentes periodos están correlacionadas? Y la ausencia de esta inercia, ¿no era una de las hipótesis de las que me has hablado? Por cierto, ¿cómo las has denominado?

Emilio: Se denomina técnicamente autocorrelación, es decir, la correlación que una variable presenta consigo misma en diferentes periodos de tiempo. La hipótesis de ausencia de autocorrelación que hemos hecho anteriormente hace referencia a las perturbaciones del modelo. Evidentemente, puede existir inercia en nuestras ventas, que es la variable explicada, sin que tenga por qué existir autocorrelación en las perturbaciones de nuestro modelo. No obstante, haremos bien en comprobar si esta hipótesis, al igual que las demás, parece cumplirse en nuestros datos o no, ya que si resultara que existe autocorrelación deberíamos tener en cuenta este hecho en la estimación del modelo.

Las dudas sobre el comportamiento adecuado o no de los residuos en el modelo (7.8) estimado se pueden resolver ahora utilizando los contrastes que se acaban de presentar. En la estimación mínimo-cuadrática llevada a cabo en el modelo (7.8) se obtiene el siguiente valor para el estadístico de Durbin-Watson: $DW = 1,632$

Si miramos en las tablas correspondientes a este estadístico, para el caso $K = 4$, $n = 40$ a un nivel de significación $\alpha = 0,05$, obtenemos los valores $dL = 1,285$ y $dU = 1,721$. Como puede apreciarse, el valor del estadístico DW obtenido arriba se encuentra dentro del intervalo de indefinición.

Si buscamos el valor de los mismos límites para un nivel de significación $\alpha = 0,01$ encontramos $d_L = 1,098$ y $d_U = 1,518$. Se aprecia que para este nivel de significación el valor del estadístico DW supera la región de indefinición y se coloca con nitidez en la región de aceptación de la hipótesis nula.

Dado este resultado, podemos concluir que no hay evidencia clara de la existencia de autocorrelación de orden 1 en el modelo, aunque al nivel de significación habitual en los contrastes hay una indefinición que no permite la aceptación de la hipótesis nula de que las perturbaciones son ruido blanco.

Podríamos tratar de salir de dudas llevando a cabo un contraste más general para lo cual utilizaremos el estadístico propuesto por Breusch y Godfrey. Con los resultados obtenidos, el estadístico a utilizar toma el valor de:

$$(T - n)R_{\mu}^2 = 37 \times 0,116 = 4,292 < \chi^2_{0,05}(3) = 7,815$$

De donde se deduce que se acepta la hipótesis nula de no existencia de autocorrelación. Por último, los resultados de ambos contrastes son compatibles y señalan que no hay evidencia de que exista autocorrelación en los residuos y pueden, desde esta perspectiva, darse por buenos los resultados obtenidos hasta el momento.

RETO 9:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

REGRESORES ESTOCÁSTICOS Y MODELOS DINÁMICOS

Don Pablo comenta con su hijo los resultados que han obtenido con el modelo propuesto, reconociendo que el experimento le ha interesado.

Don Pablo: Tengo que reconocer que lo he pasado bien con este asunto del modelo y también te diré que he aprendido bastantes cosas con ello. Pero lo que más me ha gustado ha sido que hayas tenido que dar tu brazo a torcer y aceptar que el precio de la competencia (PC) no tiene nada que ver con nuestras ventas.

Emilio: La verdad es que ha sido inesperado para mí ese resultado. Le estoy dando bastantes vueltas y creo que podría tener una explicación. En el modelo estamos suponiendo que el precio de la competencia que influye en nuestras ventas es el que se produce en el mismo periodo. Sin embargo, como sabes bien, hay un retraso entre el momento en que nuestros competidores deciden cambiar su precio y el momento en que nosotros, y no digamos los clientes, son conscientes de ello. Por lo tanto, creo que deberíamos incluir en el modelo el precio de la competencia pero referido a un precio más atrás.

Don Pablo: ¡Pero!, ¿se puede hacer eso?

Emilio: Claro que se puede hacer; es verdad que tendremos una observación menos para poder estimar los coeficientes, pero esto no tiene por qué suponer un problema.

Don Pablo: Pero entonces también a mí me gustaría hacer otra prueba. Mi padre siempre decía que nuestras ventas tenían mucho que ver con lo que hubiera pasado en el mes anterior; nunca supo a qué podría atribuirse ese efecto, pero aseguraba que si en un mes vendíamos mucho, al mes siguiente volvía a repetirse una venta alta, y viceversa. Yo he observado, que algo de esto pasa. Si se puede hacer que el precio de la competencia aparezca retrasado, está claro que lo mismo puede hacerse con las ventas.

Emilio: Lo que dices se puede hacer, pero no es lo mismo hacerlo con el precio de la competencia que con las ventas. En este último caso la cuestión es más compleja.

Don Pablo: Ya estamos. Cada vez que yo sugiero algo, resulta difícil hacerlo según tú. Si en un caso perder una observación no es importante, tampoco lo será en el otro, ¿no!

Emilio: No tiene que ver con el número de observaciones. En eso tienes razón, porque en los dos casos perderemos la información inicial. La diferencia está en que el precio de la competencia es una variable explicativa en el modelo y las ventas, en cambio, es lo que queremos explicar.

Para comprobar si la nueva hipótesis de Emilio es o no correcta, ha comenzado por incluir en el modelo como variable explicativa PC_{t-1} en lugar de PC_t . Para realizar este procedimiento, a partir del archivo de trabajo *Tab6* estimamos la ecuación de las ventas (V) contra el precio de la editorial (PE), el precio de la competencia (PC), el gasto en publicidad (GP), la renta personal (R) y el número de horas semanales de lectura multiplicada por el gasto en publicidad (LGP). Para re-estimar la ecuación con la variable PC retardada un periodo, en la barra de herramientas damos clic en *Estimate* (Figura 11.1) y en el cuadro de diálogo que se abre, reemplazamos la variable PC por $PC(-1)$ editando $PC(-1)$, como se muestra en la Figura 11.2 y damos Aceptar.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	233.2044	57.11279	4.083226	0.0003
PE	-1.820566	0.470444	-3.869888	0.0005
PC	0.194089	0.313506	0.619093	0.5400
GP	0.098176	0.006051	15.89433	0.0000
R	0.003065	0.002702	1.134507	0.2645
LGP	0.004940	0.002293	2.154637	0.0384

R-squared	0.935529	Mean dependent var	344.7425
Adjusted R-squared	0.926048	S.D. dependent var	63.37666
S.E. of regression	17.23476	Akaike info criterion	8.669214
Sum squared resid	10099.25	Schwarz criterion	8.922546
Log likelihood	-167.3843	Hannan-Quinn criter.	8.760811
F-statistic	98.67342	Durbin-Watson stat	1.677145
Prob(F-statistic)	0.000000		

Figura 11.1

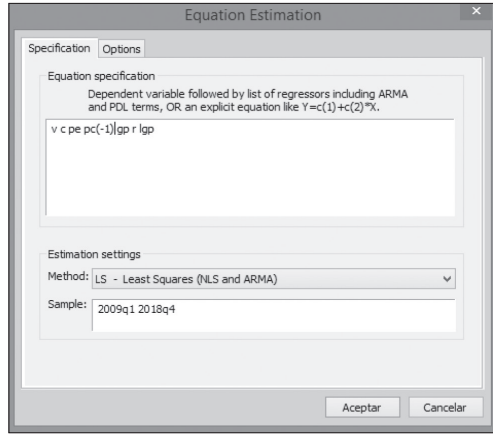


Figura 11.2

A partir de la información proporcionada por la salida de EViews, se obtiene el siguiente modelo:

$$\widehat{Ventas}_t = 306,343 - 1,995PE_t - 0,445PC_{t-1} + 0,098GP_t + 0,005R_t + 0,005LGP_t \quad (11.1)$$

(66,29) (0,48) (0,32) (0,006) (0,003) (0,002)

$R^2 = 93,90\%; \quad \bar{R}^2 = 92,98\%$

Si comparamos este resultado con el obtenido en (7.7) se comprueba que no cambia la conclusión respecto del precio de la competencia: Ni cuando se toma en el mismo periodo que el resto de variables ni cuando se incluye la variable retardada en un periodo se encuentra evidencia de que pueda influir sobre las ventas de la editorial. La probabilidad del precio de la competencia retardada un periodo es mayor que 0,05 ($0,1791 > 0,05$), por lo tanto, esta variable tampoco tiene poder explicativo en las ventas de la editorial.

La hipótesis que Don Pablo ha avanzado, por su parte, supone la inclusión de la propia variable dependiente (ventas) retardada un periodo. Para estimar este modelo en la ventana del objeto ecuación anterior, nos ubicamos en la barra de herramientas en *Estimate* y damos clic. En el cuadro de dialogo que se abre, reemplazamos la variable PC_{t-1} por PC (como en el modelo inicial) y digitamos la nueva variable propuesta por Don Pablo: ventas retardada

un periodo (V_{t-1}), editando $V(-1)$ y damos Aceptar, como se muestra en la Figura 11.3.

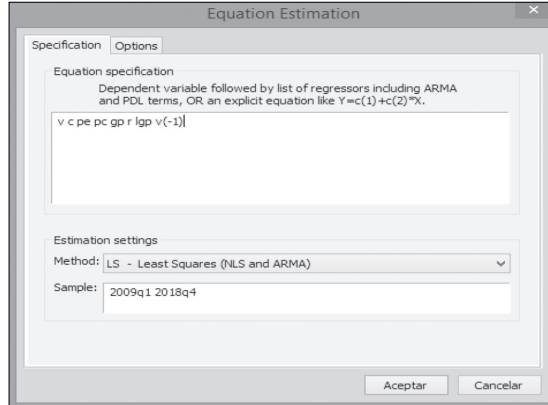


Figura 11.3

A partir de la información proporcionada por la salida de EViews, se obtiene el siguiente modelo:

$$\widehat{Ventas}_t = 225,072 - 1,932PE_t + 0,195PC_{t-1} + 0,096GP_t + 0,002R_t + 0,005LGP_t + 0,069V_{t-1} \quad (11.2)$$

(56,60) (0,48) (0,32) (0,006) (0,003) (0,002) (0,056)

$R^2 = 93,89\%$; $\bar{R}^2 = 92,75\%$

La nueva variable ventas retardada un periodo (V_{t-1}) con una probabilidad de 0,2302 superior al 5% ($0,2302 > 0,05$) no tiene significación estadística individual, por lo tanto, tampoco la hipótesis de Don Pablo parece estar avalada por los datos. Conviene resaltar que, puesto que esta hipótesis requiere la inclusión de la variable endógena retardada un periodo, es importante comprobar que no hay evidencia de autocorrelación entre las perturbaciones.

Los contraste de Durbin-Watson y Breusch y Godfrey que se han realizado anteriormente en el capítulo 10, sin que se encontrara evidencia de que exista autocorrelación entre las perturbaciones. No obstante, al incluir la variable endógena retardada estamos proponiendo un nuevo modelo en el que deberíamos llevar a cabo los contrastes correspondientes. Sin embargo,

la presencia de la variable endógena retardada como regresor invalida el resultado del contraste de Durbin-Watson (que, como ya se indicó, es válido únicamente cuando todos los regresores son fijos y presenta un sesgo hacia el valor 2 cuando se incluye como explicativa la endógena retardada).

El contraste de Breusch y Godfrey, por otra parte, no presenta esta limitación. Por lo tanto, podemos preguntarnos si existe autocorrelación en las perturbaciones mediante el siguiente procedimiento: En el Workfile del archivo *Tab6*, creamos una nueva serie para los residuos de la ecuación de ventas, seleccionamos en la barra de herramientas: *Quick* → *Generate Series* como se muestra en la Figura 11.4 y damos clic. En el cuadro de diálogo que se abre escribimos $u=resid$ como se indica en la Figura 11.5 y damos OK.

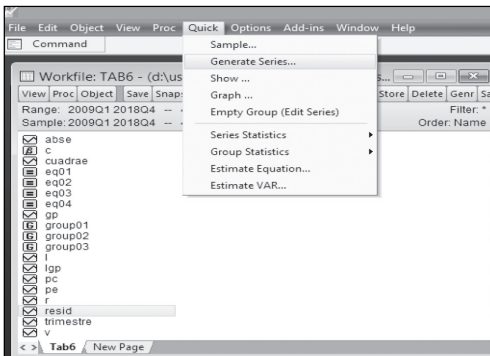


Figura 11.4

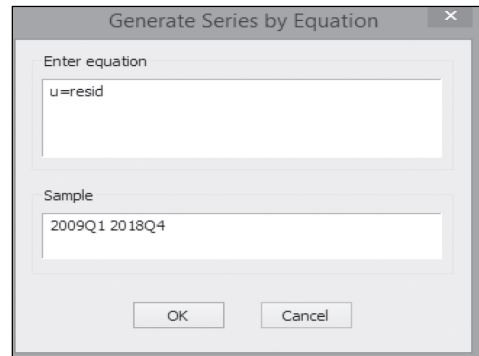


Figura 11.5

Posteriormente, con la tecla CTRL y el ratón seleccionamos las variables en el siguiente orden: u pe pc gp r y lgp y estimamos la regresión de los residuos contra las variables pe, pc, gp, r y lgp como *as Equation* y damos clic. En la ventana que se abre, en particular, en la solapa *Specification* agregamos la variable ventas retardada un periodo y la nueva variable que contiene los residuos (u) retardada uno, dos y tres periodos y damos Aceptar, como se indica en la Figura 10.6.

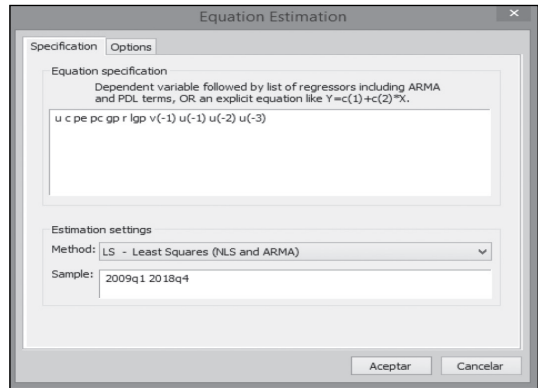


Figura 11.6

De este modo, obtenemos la salida de EViews, estimamos el siguiente modelo:

$$\begin{aligned} \mu_t = & 12,358 - 0,018PE_t - 0,071PC_t - 0,00010GP_t + 0,0008R_t - 0,00063LGP_t - 0,0013V_{t-1} + \\ & (65,99) \quad (0,57) \quad (0,36) \quad (0,007) \quad (0,004) \quad (0,003) \quad (0,065) \\ & 0,138\mu_{t-1} + 0,098\mu_{t-2} - 0,235\mu_{t-3} \\ & (0,22) \quad (0,20) \quad (0,19) \\ & R^2 = 7,43\% \end{aligned}$$

De donde se deduce que: $(T - K)R_{\mu}^2 = 37 \times 0,074 = 1,998 < \chi^2_{0,05}(3) = 7,815^{26}$; Por tanto, $1,998 < 7,815$. Por consiguiente, se acepta se acepta la hipótesis nula de que no existe autocorrelación entre las perturbaciones.

Por otro lado, y tras conocer el contenido de los capítulos precedentes, podemos considerar que quizá ninguno de los dos modelos anteriores debería ser el propuesto ya que con el objeto de evitar posibles sesgos en la estimación, deberíamos haber incluido a la vez ambas variables, en tanto, que potenciales variables explicativas, de las que originalmente formaban parte de él. Para ello, cargamos en el *Workfile* el archivo Tabla 7.6 y con la techa CTRL y con el ratón seleccionamos las variables en el siguiente orden: v pe pc gp r lgp, posteriormente, abrimos como ecuación (*as*

26 Este valor se obtiene de la tabla estadística de la distribución Chi-cuadrado con 3 grados de libertad y una probabilidad en la cola derecha de 0,05.

Equation) y damos clic en la barra de herramientas del workfile en *Estimate*. En la solapa *Specification* del cuadro de diálogo que a continuación aparece, editamos las siguientes variables: precio de la competencia retardado un periodo (PC_{t-1}) escribiendo $pc(-1)$ y ventas de la editorial retardada un periodo (V_{t-1}), escribiendo $v(-1)$, como se muestra en la Figura 11.7 y damos Aceptar.

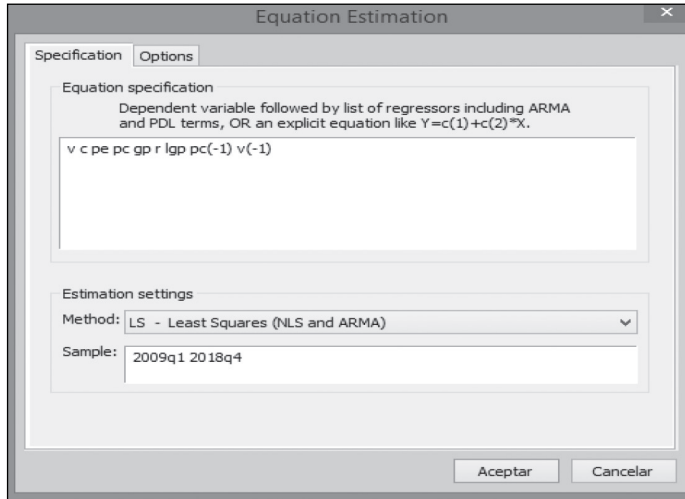


Figura 11.7

A partir de la salida de EViews, estimamos el siguiente modelo, el cual incorpora todas las variables.

$$\widehat{Ventas}_t = 302,969 - 2,168PE_t + 0,068PC_t + 0,097GP_t + 0,003R_t + 0,005LGP_t \quad (11.4)$$

$$\begin{array}{cccccc} (79,18) & (0,50) & (0,32) & (0,006) & (0,003) & (0,002) \\ & & - 0,473PC_{t-1} & + 0,075V_{t-1} & & \\ & & (0,34) & (0,06) & & \end{array}$$

$$R^2 = 94,25\%; \quad \bar{R}^2 = 92,95\%$$

Ahora bien, la inclusión de variables que acabamos de hacer seguramente es excesiva, ya que el precio de la competencia aparece dos veces con sus valores en t ($\text{Prob} \rightarrow 0,8324 > 0,05$) y en $t-1$ ($\text{Prob} \rightarrow 0,1755 > 0,05$), ninguno de los cuales resulta significativo al 5%, mientras que Emilio pensaba que uno debería sustituir a otro. El modelo anterior puede simplificarse un poco más, eliminando la variable precio de la competencia (PC_t); para ello, re-estima-

mos el modelo, como se muestra en la Figura 11.8 y damos Aceptar. La salida de EViews nos permite estimar el siguiente modelo.

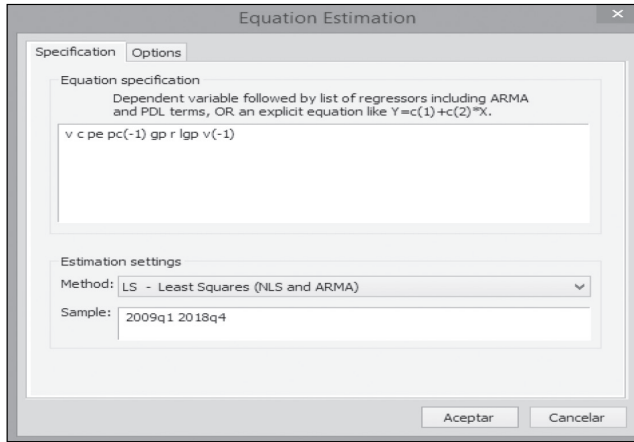


Figura 11.8

$$\widehat{Ventas}_t = 312,123 - 2,174PE_t - 0,493PC_{t-1} + 0,097GP_t + 0,004R_t + 0,005LGP_t + 0,075V_{t-1} \quad (11.5)$$

(65,57) (0,49) (0,32) (0,006) (0,003) (0,002) (0,05)

$R^2 = 94,24\%$; $\bar{R}^2 = 93,16\%$

Ahora pues, re-estimaremos el modelo eliminando las variables precio de la competencia retardado un periodo (PC_{t-1}) y la renta personal (R_t), como se muestra en la Figura 11.9.

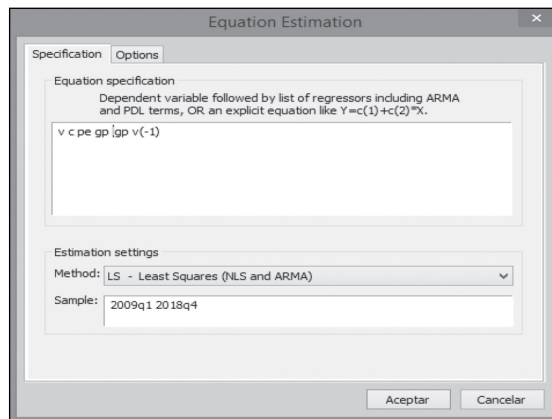


Figura 11.9

El resultado anterior, una vez realizados los contrastes de significación individuales al 5%, muestra que el precio de la competencia ($PC_{t-1} \rightarrow \text{Prob.} \rightarrow 0,1354 > 0,05$) no tiene influencia significativa sobre las ventas aunque se tome la variable con un retardo, tampoco la renta personal ($R_t \rightarrow \text{Prob.} \rightarrow 0,2176 > 0,05$), mientras que hay evidencia de que las ventas retardadas un periodo tienen alguna influencia positiva sobre las ventas actuales. Por lo tanto, la hipótesis de Emilio no se confirma mientras que sí lo hace la de Don Pablo.

De conformidad con la Tabla 11.1, el modelo final adecuado para explicar las ventas de la editorial, resulta el modelo (11.6).

Tabla 11.1

Estimación modelo final de ventas de la editorial.

Variable Dependiente: V

Método: MCO

Muestra ajustada: 2009Q2 2018Q4

Observaciones incluidas: 39 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	259,4275	44,64444	5,810969	0,0000
PE	-2,148630	0,400038	-5,371061	0,0000
GP	0,097627	0,005799	16,83440	0,0000
LGP	0,004441	0,002208	2,011040	0,0523
V(-1)	0,091115	0,047215	1,929804	0,0620
R-squared	0,936672	Mean dependent var		346,5179
Adjusted R-squared	0,929221	S.D. dependent var		63,18935
S.E. of regression	16,81107	Akaike info criterion		8,601161
Sum squared resid	9.608,806	Schwarz criterion		8,814438
Log likelihood	-162,7226	Hannan-Quinn criterion		8,677683
F-statistic	125,7211	Durbin-Watson stat		1,899826
Prob(F-statistic)	0,000000			

Por tanto, se obtiene:

$$\widehat{Ventas} = 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \quad (11.6)$$

$$(44,64) \quad (0,40) \quad (0,006) \quad (0,002) \quad (0,047)$$

$$R^2 = 93,67\%; \quad \bar{R}^2 = 92,92\%$$

Ahora realizaremos el contraste de Breusch y Godfrey. Para ello nos situamos en la ventana del objeto ecuación del archivo *Tab6* y seleccionamos *View* → *Residual Diagnostics* → *Serial Correlation LM Test* como se muestra en la Figura 11.10 y damos clic. En la pestaña del cuadro de diálogo que se abre, la cual solicita se proporcione el número de retardos a incluir (*Lags to include*), escribimos 1 y damos OK (véase Figura 11.12).

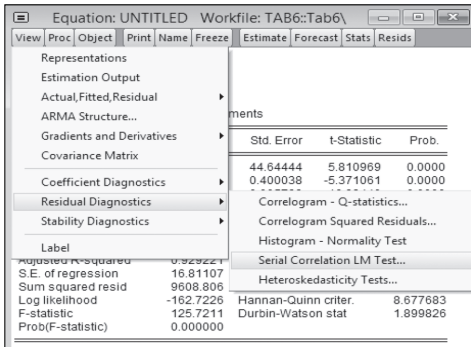


Figura 11.10

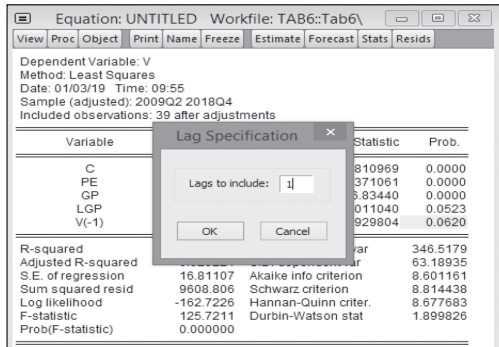


Figura 11.12

La Tabla 11.2 muestra que la probabilidad del estadístico F (F-statistic) y Obs*R-squared es superior a 0,05, por lo tanto, se acepta la hipótesis nula de ausencia de autocorrelación en las perturbaciones.

Tabla 11.2***Contraste de Breusch-Godfrey Serial Correlation LM Test.***

F- statistic	0,098093	Prob. F(1, 33)	0,7561	
Obs*R-squared	0,115584	Prob. Chi-Square(1)	0,7339	
Test Equation:				
Dependent Variable: RESID				
Method: Least Squares				
Sample: 2009Q1: 2018Q4				
Included observations: 39				
Presample missing value lagged residuals set to zero				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2,786338	46,11490	-0,060422	0,9522
PE	0,035280	0,420809	0,083839	0,9337
GP	0,000388	0,006007	0,064617	0,9489
LGP	-0,000167	0,002300	-0,072393	0,9427
V(-1)	-0,004501	0,049965	-0,090076	0,9288
RESID(-1)	0,060571	0,193395	0,313197	0,7561
R-squared	0,002964	Mean dependent var	1,37E-14	
Adjusted R-squared	-0,148102	S.D. dependent var	15,90168	
S.E. of regression	17,03857	Akaike info criterion	8,649475	
Sum squared resid	9.580,328	Schwarz criterion	8,905407	
Log likelihood	-162,6648	Hannan-Quinn criterion	8,741301	
F-statistic	0,019619	Durbin-Watson stat	2,009270	
Prob(F-statistic)	0,999828			

Ahora, obtendremos un correlograma de los residuos. Para ello, nos situamos en la ventana del objeto ecuación del modelo estimado (11.6) y seleccionamos *View* → *Residual Diagnostics* → *Correlogram – Q statistics* y damos clic, como se indica en la Figura 11.10. En el cuadro de diálogo que aparece a continuación damos OK (véase Figura 11.12).

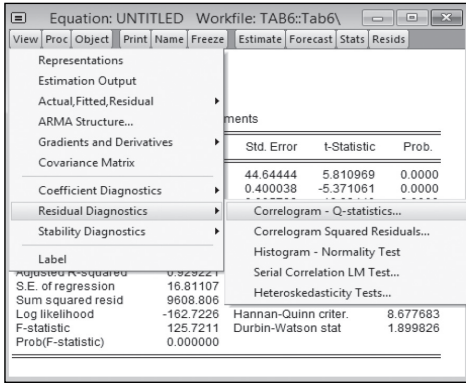


Figura 11.10

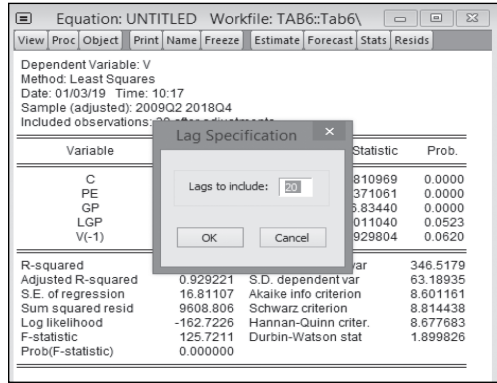


Figura 11.12

Los resultados del correlograma de residuos (véase Figura 11.13) muestran que no existe evidencia en contra de la existencia de incorrelación entre las perturbaciones de este modelo. Así pues, podemos dar por válidos los resultados de la estimación, puesto que los residuos de autocorrelación y correlación parcial se encuentran dentro de las bandas respectivas.

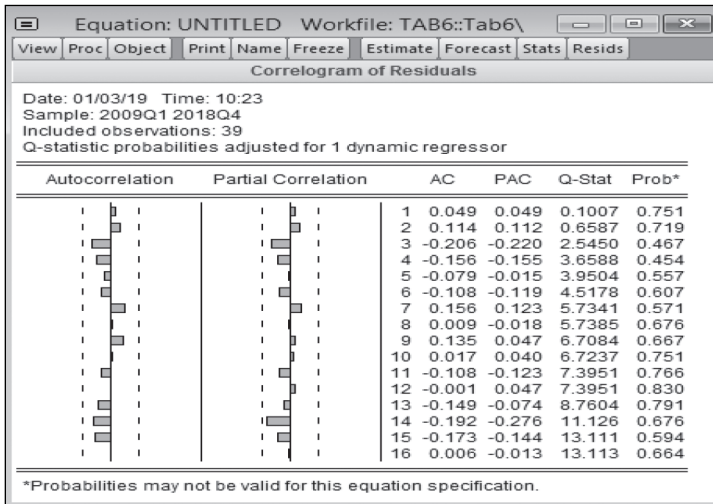


Figura 11.13

Aunque el contraste de Breusch y Godfrey y el correlograma de los residuos realizados en este último modelo señalan la ausencia de autocorrelación en las perturbaciones, antes de decidir adoptar definitivamente el modelo (11.11) trataremos de cerciorarnos de que no contiene ningún error de especificación importante. En particular, y debido a la presencia en el mismo de la variable dependiente ventas retardada un periodo (V_{t-1}), podríamos confirmar el resultado obtenido con el contraste de Breusch y Godfrey llevando a cabo un contraste de Hausman que conforme que este regresor, no está correlacionado con los residuos.

11.1 Contraste de exogeneidad de Hausman.

El contraste de Hausman sirve para detectar la presencia de regresores estocásticos en un modelo. Bajo la hipótesis nula se supone que el regresor no es estocástico, mientras que la hipótesis alternativa establece lo contrario. El contraste compara los estimadores de los parámetros del modelo obtenido por los métodos de mínimos cuadrados ordinarios (MCO) y por variables instrumentales (VI). Bajo el supuesto de que el regresor no es estocástico (hipótesis nula) los estimadores obtenidos por ambos métodos son consistentes, mientras que bajo la hipótesis alternativa (regresor estocástico) solo el estimador de variables instrumentales mantiene esta propiedad. De este modo, bajo la hipótesis nula $\text{plim}(\hat{\beta}_{\text{MCO}} - \hat{\beta}_{\text{VI}}) = 0$, apenas habrá diferencia entre ambos estimadores.

Contraste:

Hipótesis nula	frente	Hipótesis alternativa
H_0 : El regresor no es estocástico		H_a : El regresor es estocástico

Cuanto mayor sea la diferencia entre las estimaciones de los parámetros por ambos métodos (MCO y VI), mayor evidencia existirá en contra de la hipótesis nula.

Si recordamos el modelo final (11.6), el vector de coeficientes β^* se limita en este caso a β_4 , coeficiente que acompaña a la variable V_{t-1} . Los dos estimadores a considerar serán, por una parte, el estimador MCO que acabamos

de obtener en (11.6) y, por otra, un estimador de variables instrumentales que sustituya al regresor potencialmente problemático (V_{t-1}) en el modelo 11.7 por una predicción del mismo derivada de una regresión de esta variable sobre el conjunto de potenciales instrumentos.

Considerando el conjunto de instrumentos a nuestra disposición, en este caso, aparecen de inmediato seis variables PC_t , PC_{t-1} , R_t , R_{t-1} , PE_{t-1} y GP_{t-1} que no aparecen en el modelo y pueden considerarse fijas; podríamos por lo tanto, utilizarlas como instrumentos de V_{t-1} . Sin embargo, las variables PC_t , PC_{t-1} , R_t y R_{t-1} presentan el inconveniente de estar poco correlacionadas con la variable dependiente V_t , que es la razón por la que se han eliminado del modelo; por lo que es de esperar que su correlación con V_{t-1} sea asimismo reducida. Por otro lado, es claro que la eficiencia asintótica de este estimador aumenta cuando los instrumentos están muy correlacionados con la varianza a instrumentar. Nos conviene, por tanto, buscar instrumentos lo más correlacionados posible con las ventas retardadas. Dado que al incluir un retardo en la ecuación se pierde la primera observación, resulta posible utilizar el precio de la editorial y el gasto en publicidad retardados un periodo como variables instrumentales. Por todo ello, finalmente se ha decidido utilizar las variables PC_t , PC_{t-1} , R_t , R_{t-1} , PE_{t-1} y GP_{t-1} como instrumentos.

En la barra de herramientas de la ventana del objeto ecuación del modelo final (11.6) seleccionamos *Estimate* y damos clic. Posteriormente, en la pestaña *Method* del el cuadro de diálogo de especificación (*Specification*) del modelo, más específicamente, en *Estimation Setting*, en la pestaña de método (*Method*) elegimos el método de mínimo cuadrados en dos etapas (*TSLS – Two Stage Least Squares – TNSLS and ARMA*) y en el cuadro de variables instrumentales (*Instrument list*) editamos separadas por un espacio las siguientes variables: PC_t PC_{t-1} R_t R_{t-1} PE_{t-1} y GP_{t-1} y damos Aceptar, como se indica en la Figura 11.14.

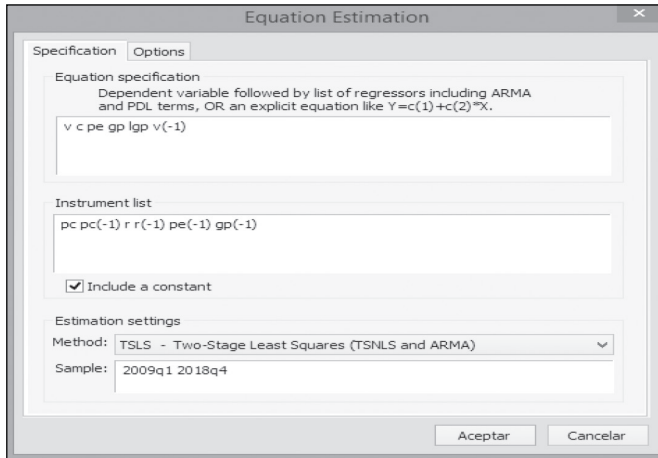


Figura 11.14

Los resultados de la regresión realizada por EViews se muestran en la Tabla 11.3, información que proporciona la siguiente ecuación:

Tabla 11.3

Estimación del modelo por el método de mínimos cuadrados en dos etapas.

Variable Dependiente: V
 Método: Two-Stage Least Squares
 Muestra ajustada: 2009Q2 2018Q4
 Observaciones incluidas: 39 after adjustments
 Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	122,1993	253,5420	0,481969	0,6329
PE	-1,350207	1,968580	-0,685879	0,4974
GP	0,144080	0,052275	2,756225	0,0093
LGP	-0,000148	0,020423	-0,007252	0,9943
V(-1)	-0,085129	0,218629	-0,389376	0,6994
R-squared	0,813603	Mean dependent var		346,5179
Adjusted R-squared	0,791674	S.D. dependent var		63,18935
S.E. of regression	28,84132	Sum squared resid		28281,94
F-statistic	20,68965	Durbin-Watson stat		1,617791
Prob(F-statistic)	0,000000	Second-Stage SSR		82.889,56
J-statistic	0,033301	Instrument rank		6
Prob(J-statistic)	0,855202			

Por consiguiente, el modelo obtenido es:

$$\widehat{Ventas}_{VI} = 122,199 - 1,350PE_t + 0,144GP_t - 0,00015LGP_t - 0,085V_{t-1} \quad (11.7)$$

$$(253,54) \quad (1,97) \quad (0,052) \quad (0,020) \quad (0,22)$$

Para calcular \hat{q} procedemos a restar del coeficiente de la variable $V(-1)$ del modelo de variables instrumentales (11.7) el valor del coeficiente de la variable $V(-1)$ del modelo de MCO (11.6), obteniendo:

$$\hat{q} = \hat{\beta}_4^{VI} - \hat{\beta}_4^{MCO} = -0,085129 - 0,091115 = -0,070438$$

y su varianza se deduce de

$$\text{Var}(\hat{q}) = \text{Var}(\beta_4^{VI}) - \text{Var}(\beta_4^{MCO})$$

$$\text{Var}(\hat{q}) = \sigma^2((X'Z(Z'Z)^{-1}Z'X)_{[6,6]}^{-1} - (X'X)_{[6,6]}^{-1})$$

Será preciso estimar σ^2 : de la estimación realizada por VI obtenemos $\hat{\sigma}_{VI}^2 = 28,84132$ que es diferente a la obtenida en la estimación MCO (16,81107). Por lo tanto, para este caso.

$$\text{Var}(\hat{q}) = (0,218629)^2 - (0,047215)^2 = 0,04556938$$

de donde el estadístico de contraste toma el valor:

$$m = \frac{(-0,070438)^2}{0,04556938} = 0,1088782 < X_{0,05}^2(1) = 3,84146$$

El valor del estadístico obtenido, es decir, 0,1088782 se compara con el valor de la distribución chi-cuadrado con un grado de libertad y una probabilidad en la cola derecha de 0,05, cuyo valor es 3,84146, como se muestra en las Figuras 11.15 y 11.16 del buscador de GRETl.

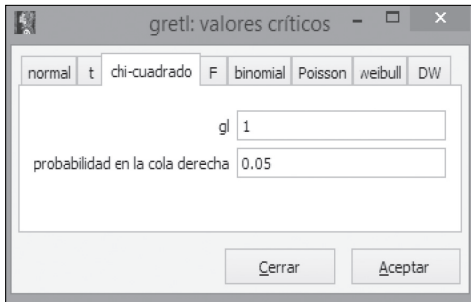


Figura 11.15

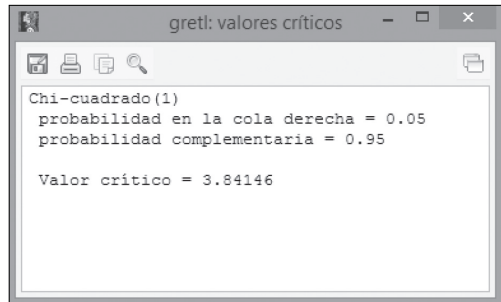


Figura 11.16

Como $0,1088782 < 3,84146$, por consiguiente, rechazamos la hipótesis nula de que el regresor no es estocástico. El resultado del contraste señala que no hay evidencia de que exista correlación entre la variable dependiente retardada que se utiliza como explicativa y las perturbaciones de la ecuación, por lo que la estimación MCO realizada en (11.6) resulta satisfactoria.

Y llegamos al final de la historia de la Editorial, que nos ha permitido ilustrar la interpretación que puede darse a los resultados así como la utilidad que un modelo puede tener como instrumento para ayudar en la toma de decisiones. Los pasos sucesivos, cuidadosamente estudiados, nos han permitido llegar a formular un modelo en el que tanto Don Pablo como Emilio han tenido que ponerse de acuerdo. Don Pablo ha cedido en su opinión inicial, reconociendo que era excesivamente simplista, y Emilio no ha encontrado evidencia que sustente su idea inicial de que el precio de la competencia y la renta personal tienen mucha importancia en las ventas de la editorial. Lo que han aprendido ambos en este proceso les permitirá encarar la marcha de la editorial con evidencia empírica.

RETO 10:

Se invita al lector a realizar las rutinas econométricas expuestas en este capítulo, utilizando las bases de datos que acompañan este texto con el apoyo del software EViews.

USO DEL MODELO ECONOMÉTRICO PARA ANÁLISIS DE POLÍTICA

Don Pablo y su hijo Emilio se sienten satisfechos por el trabajo realizado hasta el momento. Aunque sus discusiones han sido bastante acaloradas (cada quien defendiendo su postura respecto a los determinantes de las ventas de la editorial), aún queda por resolver un interrogante: El trabajo econométrico realizado nos ha conducido a la obtención de un modelo final que explica las ventas de la editorial, pero ambos se preguntan: ¿Qué haremos con este modelo? Don Pablo como hombre de negocios excesivamente pragmático, sabe que tanto esfuerzo no debe desperdiciarse, por ello, decide romper el silencio y preguntar a su hijo:

Don Pablo: Emilio, tú has sido testigo de los airados debates que hemos sostenido ambos en torno a la determinación de las variables que explican las ventas de la editorial. Al final, el uso de la econometría que tú tanto defiendes, nos ha generado respuestas sorprendentes. ¡Hijo!, sin titubear, dime ¿qué vamos a hacer con este modelito?

Emilio: ¡Padre!, tu pregunta es apenas lógica. Desde el momento que estimamos el modelo econométrico final de las ventas de la editorial, yo también me pregunté sobre el uso que le íbamos a dar. Para no enredar más la respuesta, con mucha seguridad te diré que el último paso de la metodología econométrica tradicional es la utilización del modelo con fines de control o de política empresarial; en nuestro caso, este modelito como tú lo llamas, nos permitirá diseñar una política de ventas y una política de precios.

Don Pablo: Entonces adelante, explícame cómo utilizar este modelo en la toma de decisiones empresariales.

Partiremos del modelo final obtenido:

$$\widehat{Ventas} = 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \quad (12.1)$$

$$(44,64) \quad (0,40) \quad (0,006) \quad (0,002) \quad (0,047)$$

$$R^2 = 93,67\% \quad \bar{R}^2 = 92,92\%$$

En la terminología y notación utilizada en la literatura especializada en econometría a la variable dependiente (Y_t) también se le suele llamar “variable controlada” y a las variables explicativas (X_t), se les denomina alternativamente, “variables de control”. Por ende, mediante una mezcla apropiada de políticas comerciales, las directivas de la editorial pueden manejar las variables de control X_t para producir el nivel deseado de la variable objetivo Y_t . Para ser más exactos, la editorial es capaz de manipular las variables precio de los libros y el gasto en publicidad para elevar las ventas en un monto determinado. También debe recordarse que la introducción de variables ficticias en el modelo permitió comprender que la campaña publicitaria de la editorial debe intensificarse en periodo vacacional (tercer trimestre del año) y focalizarse, en particular, a mujeres con estudios universitarios, por ser las personas que más leen.

Don Pablo: Para ser honesto, puede comprender qué es una variable de control y una variable controlada u objetivo, pero Emilio, te sugiero ser más aplicado. Con esto me refiero a que trates de trasladar este concepto a un ejemplo práctico para que lo pueda comprender mejor.

Emilio: ¡Papá, cálmate! Eso es precisamente, lo que trato de hacer. Entonces, ¡manos a la obra!

El lector recordará que la información de la editorial correspondiente al cuarto trimestre de 2018 es la siguiente:

Ventas (V_{t-1}) = \$407,0 millones; Precio promedio (PE_t) = \$97,7 miles y Gasto en publicidad (GP_t) = \$3.352,0 miles..

Si la editorial se fija como meta para el primer trimestre de 2019, incrementar sus ventas (V_{t+1}) en 10%, por lo tanto, $\$407,0 \times 1,10 = \$447,7$, manteniendo constante el gasto en publicidad, $LGP = 0$ (correspondiente al primer trimestre de 2019) y $V_{t-1} = \$407,0$ (ventas retardadas un periodo, es decir, las ventas correspondientes al cuarto trimestre de 2018). Por lo tanto, la pregunta a resolver es la siguiente: ¿Cuál debe ser el precio promedio de los libros para obtener ventas por valor de $\$447,7$ millones en el primer trimestre de 2019?

Como puede observarse, la variable objetivo corresponde a las ventas de la editorial. En el modelo econométrico final procederemos a reemplazar los valores en las variables de control manteniendo como incógnita el precio promedio, así:

$$\begin{aligned} \widehat{Ventas} &= 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \\ 447,7 &= 259,428 - 2,149PE_t + 0,098(3.352) + 0,0044(0) + 0,091(407,0) \\ 447,7 &= 259,428 - 2,149PE_t + 328,5 + 0 + 37,04 \\ 447,7 - 259,428 - 328,5 - 0 - 37,04 &= -2,149PE_t \\ 82,5 &= 2,149PE_t \\ PE_t &= \frac{-177,26}{-2,149} = \$ 82,5 \text{ miles} \end{aligned}$$

Si Don Pablo pretende incrementar las ventas de la editorial en el primer trimestre de 2019 en un 10%, deberá reducir el precio promedio de los libros en un 15,6% en el mismo periodo.

Ahora, simularemos otra situación: ¿cuál debe ser el gasto en publicidad, si la editorial pretende obtener ventas por valor de $\$447,7$ millones en el primer trimestre de 2019?

La incógnita es el gasto en publicidad (GP_t), por lo tanto:

$$\begin{aligned} \widehat{Ventas} &= 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \\ 447,7 &= 259,428 - 2,149(97,7) + 0,098 GP_t + 0,0044(0) + 0,091(407,0) \\ 447,7 &= 259,428 - 210,0 + 0,098GP_t + 0 + 37,04 \\ 447,7 - 259,428 + 210,0 - 37,04 &= 0,098GP_t \\ 361,19 &= 0,098GP_t \\ GP_t &= \frac{361,19}{0,098} = \$ 3.685,6 \text{ miles} \end{aligned}$$

En fin, podemos simular diferentes situaciones, por ejemplo, determinar el valor del gasto en publicidad en el tercer trimestre de 2019, conociendo de antemano, que en periodo vacacional las ventas se elevan por efecto del incremento de las horas de lectura de los clientes de la editorial, en especial, en el grupo de mujeres con estudios universitarios.

Si Don Pablo se propone incrementar las ventas de la editorial en el primer trimestre de 2019 en un 10%, deberá también aumentar el gasto en publicidad en un 10% en el mismo periodo.

Don Pablo: Emilio, en verdad, me ha sorprendido la versatilidad de tu modelito. En realidad, nunca pensé en la potencialidad de este modelo como simulador para la toma de decisiones. Honestamente, veo que ahora tenemos en las manos una herramienta poderosa para planear nuestras metas de ventas en función del precio de los libros y el gasto en publicidad como variables que juntos podemos controlar.

Emilio: ¡Papá!, celebro mucho que al final, hayas podido comprender la importancia de un buen modelo econométrico. Desde hoy, no tomaremos decisiones empresariales a ciegas; en cambio, lo haremos apoyados en una herramienta versátil como ésta.

Don Pablo: ¿Pero que sigue en adelante?

Emilio: El modelo econométrico que hemos construido debe ser objeto de constante verificación, calibración y actualización para conservar su poder explicativo, de simulación y de predicción. Como los fenómenos económicos cambian constantemente, también nuestro modelo debe ser capaz de adaptarse a los nuevos cambios para responder a los nuevos retos comerciales en un mundo altamente competitivo.

Don Pablo: Estoy seguro que vamos un paso más adelante que la competencia. Probablemente, nuestros competidores continúen tomando decisiones basadas en la intuición, desconociendo el poderío de la econometría. Ahora entiendo, cómo el conocimiento puede marcar la diferencia.

Emilio: Comparto contigo los alcances de la econometría, pero también soy consciente, de sus limitaciones y frecuentes críticas. Por cierto, esto lo discutiremos en el siguiente capítulo.

Don Pablo: ¡Hijo!, delego en ti, el diseño y la implementación de la campaña publicitaria de la editorial. No subestimes el conocimiento adquirido, que tanto esfuerzo, dedicación y discusión ha requerido. Hoy, comprendo mejor que la publicidad es un medio de producir información a bajo costo, puesto que la recopilación de información tiene un costo en términos del tiempo que hay que dedicar y que el tiempo tiene un valor implícito. De la forma más sencilla, la publicidad ahorra tiempo a nuestros clientes, en su esfuerzo por adquirir información sobre los precios o calidades de nuestros productos. En el caso de la editorial, estoy más que seguro, que la campaña publicitaria que emprendamos debe apoyarse en el uso de la red de internet y de otras herramientas virtuales.

Emilio: ¡Sabes!, siento gran satisfacción por haber comprendido, después de todo, que la existencia de publicidad reduce la cantidad de tiempo que nuestros clientes invierten en la búsqueda de precios más bajos. Si ellos invierten menos tiempo buscando precios más baratos, les queda necesariamente más tiempo para dedicar a otras actividades más deseables, como trabajar un poco más, compartir en familia o disfrutar del ocio.

UNA SÍNTESIS DE LA HISTORIA DE LA EDITORIAL

Para visualizar las etapas que ha atravesado el modelo econométrico de ventas de la editorial a lo largo de los capítulos abordados, se incluye un resumen muy sintético del desarrollo de nuestro modelo en el contexto de la aplicación de la metodología econométrica habitual según Gujarati (2010).

13.1 Planteamiento de la teoría o hipótesis relacionada con las ventas de la editorial.

Don Pablo considera que las ventas de la editorial están determinadas por el precio de los libros (a mayor precio menor volumen de ventas y viceversa), los demás factores son irrelevantes. Emilio, en cambio, percibe que el precio es un factor que explica las ventas de la editorial, en alguna medida; pero también cree que existen otras variables con igual o mayor poder explicativo, tales como: el precio de la competencia, el gasto en publicidad y la renta personal de los clientes actuales y potenciales de la editorial.

La hipótesis anterior está contextualizada en el siguiente razonamiento lógico: *“la ciencia económica considera que en un mercado ideal, de libre competencia de bienes y servicios como de factores productivos, llegarán los demandantes tratando de obtener la mayor cantidad de bienes al precio más bajo posible. Aunque el precio es un factor preponderante en la configuración de la demanda, no es el único. Los oferentes intentarán identificar los factores que determinan la demanda para sacar mayor provecho de su actividad económica”*.

13.2 Especificación del modelo matemático de ventas de la editorial.

La ley de la demanda de amplia aceptación en la economía de mercado establece que existe una relación negativa o inversa entre el precio de un bien, servicio o factor productivo y la demanda de un bien, servicio o factor productivo. Aunque existen ciertas excepciones -como los bienes Giffen y Veblen-, las dos variables en cuestión se comportarán de esta manera.

Por simplicidad, un matemático podría proponer la siguiente relación funcional del modelo (ecuación matemática) de ventas en relación con el precio de la editorial (PE_t), el precio de la competencia (PC_t), el gasto en publicidad (GP_t) y la renta personal (R_t).

$$\text{Ventas}_t = \beta_0 - \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t \quad t = 1, 2, \dots, T \quad (13.1)$$

Se supone que en este modelo existen relaciones deterministas o exactas entre la variable dependiente y las variables explicativas.

13.3 Especificación del modelo econométrico de ventas de la editorial

Emilio consiente de la imposibilidad de establecer relaciones exactas entre las variables económicas (sin opción al error), especifica el modelo en función de las variables que lo determinan, teniendo en cuenta que los datos tienen errores de medición, por lo tanto, las relaciones entre las variables son inexactas, es decir, existen residuos.

$$\text{Ventas}_t = \beta_0 - \beta_1 PE_t + \beta_2 PC_t + \beta_3 GP_t + \beta_4 R_t + \mu_t \quad t = 1, 2, \dots, T \quad (13.2)$$

Emilio, sabiamente ha incluido en el modelo el término de perturbación o de error estocástico μ_t , el cual recoge los errores o residuos en la medición de la relación entre las variables en estudio.

Los signos (\pm) que preceden a cada coeficiente indican a priori la dirección existente entre cada par de variables (dependiente y explicativa), por ejemplo, el signo menos indica que se sospecha que existe una relación inversa (negativa) entre las ventas y el precio promedio de la editorial (PE_t) y una relación directa (positiva) de las ventas con el precio de la competencia (PC_t), el gasto en publicidad (GP_t) y la renta personal de los clientes (R_t).

13.4 Estimación de los parámetros del modelo econométrico de las ventas de la editorial.

Para medir el efecto marginal que ejerce el precio sobre las ventas de la editorial, Emilio a partir de la combinación de teoría económica, métodos estadísticos y datos estimará el siguiente modelo, el cual permitirá la dirección y el efecto marginal que ejerce cada variable explicativa sobre la variable dependiente, así:

$$\widehat{Ventas} = \hat{\beta}_0 - \hat{\beta}_1 PE_t + \hat{\beta}_2 PC_t + \hat{\beta}_3 GP_t + \hat{\beta}_4 R_t + \hat{\mu}_t \quad t = 1, 2, \dots, T \quad (13.3)$$

Donde: V_t representa las ventas de libros, PE_t el precio promedio al que los vende la editorial, PC_t el precio promedio de la competencia (precio al que venden otras editoriales), GP_t el gasto realizado en publicidad por la editorial y R_t la renta o ingreso promedio de los clientes de la editorial.

La estimación del modelo de ventas de la editorial proporciona el siguiente resultado:

$$\widehat{Ventas} = 243,20 - 1,91PE_t + 0,127PC_t + 0,100GP_t + 0,003R_t \quad (13.4)$$

$$(59,8) \quad (0,49) \quad (0,33) \quad (0,006) \quad (0,003)$$

$$R^2 = 92,67\%; \quad \bar{R}^2 = 91,84\%$$

En este modelo, el coeficiente de determinación indica que el 92,6% de la variabilidad de la media de las ventas de la editorial son explicadas por el precio de la editorial (PE_t), el precio de la competencia (PC_t), el gasto en publicidad (GP_t) y la renta personal (R_t); pero aún queda un remanente de 7,3% ($100\% - 92,7\% = 7,3\%$) de las ventas sin explicación, lo que supone que existen otros factores determinantes de las ventas, así como errores en la medición

13.5 Prueba de hipótesis del modelo econométrico de ventas de la editorial.

A través de la prueba de hipótesis Emilio intentará demostrar a su padre que el valor de los coeficientes que acompañan a las variables incorporadas en el modelo pueden o no ser relevantes al momento de explicar las ventas. La significación estadística de cada variable se demostrará mediante la ejecución de contrastes individuales y conjuntos.

Don Pablo sostiene que el precio de los libros es determinante en la explicación de las ventas de la editorial, los demás factores son poco significativos. De otro lado, Emilio sospecha, que si bien, el precio es una variable con poder explicativo en las ventas de la editorial, también hay otros factores explicativos como el precio de la competencia, el gasto en publicidad (con el cual se identifica mucho) y la renta personal.

Se tratará de resolver esta discusión aplicando algunos contrastes individuales y conjuntos. Mediante la distribución t de Student se concluye que el precio de la editorial y el gasto en publicidad tienen significancia estadística al 5%, mientras que el precio de la competencia y la renta personal de los clientes, han demostrado no ser significativos a este mismo nivel de significancia. Parece que Don Pablo ha ganado parcialmente esta discusión al sostener que el precio de la editorial es la variable que explica las ventas de la editorial, pero también Emilio resulta vencedor al comprobar que el gasto en publicidad tiene poder explicativo.

Emilio no se siente derrotado por la postura de su padre. Admite que si bien el precio de la competencia y la renta personal han demostrado individualmente no ser estadísticamente significativas, considera que conjuntamente que sí lo son. Para demostrar esto, Emilio efectúa un contraste conjunto apoyado en la tabla de distribución de Snedecor; los resultados confirman que globalmente las cuatro variables tienen capacidad explicativa sobre las ventas de la editorial. Padre e hijo, pese a sus diferencias, deciden conservar en el modelo estimado (13.4), las variables en discusión y continuar con la revisión minuciosa de este modelo.

A continuación, Emilio ante la mirada intrigante de su padre, efectúa un análisis de residuos gráfico y numérico, prueba la inclusión de variables irrelevantes y la omisión de variables relevantes en el modelo, realiza las pruebas de estabilidad estructural pertinentes, así como la prueba de especificación de la forma funcional del mismo. Al final de cuentas, el modelo estimado hasta ahora, no adolece de ningún problema importante de especificación. Hasta aquí, todo parece ir bien.

13.6 Predicción.

Don Pablo, aunque siente satisfacción al saber de qué el modelo (13.4) superó satisfactoriamente las pruebas de diagnóstico a las que fue sometido por Emilio, no deja de pensar en cuál será la utilidad que tendrá este modelo. Sin vacilación alguna, le pregunta a su hijo: ¡Hola hijo!, dime ¿qué harás con este modelito que tanta discusión ha provocado?

De inmediato, Emilio responde: ¡Papá!, usaremos este modelo para predecir el valor que se espera alcancen las ventas de la editorial en el supuesto de que las variables explicativas tomen ciertos valores que nosotros podamos fijar y controlar.

Como Don Pablo está muy interesado en reducir el gasto en publicidad, en al menos, el 10%, le pide a Emilio que pronostique las ventas en un escenario en que la editorial realice un gasto en publicidad de \$ 3.016,8 en el primer semestre de 2019, manteniendo los demás factores constantes. Don Pablo quiere conocer el valor de las ventas en el primer trimestre de 2019, a partir de los valores de las variables explicativas correspondientes al cuarto trimestre de 2018.

$$X'_p = [97,7 \quad 113,5 \quad 3.016,8 \quad 5.712,1]$$

Emilio procede a reemplazar estos valores en el modelo obtenido, por lo tanto, la predicción resulta ser:

$$\hat{Y}_p = 243,20 - 1,91x97,7 + 0,127x113,5 + 0,100x3.016,8 + 0,003x5.712,1 = 388,53$$

con un intervalo de:

$$IC(Y_P)_{(1-\alpha)} = [\hat{Y}_P \pm \hat{\sigma}_e t_{\alpha/2}]$$

$$\begin{aligned} IC(Y_P)_{0,95} &= [388,53 \pm (19,73)(2,030)] \\ &= [348,48 \quad 428,58] \end{aligned}$$

Los valores [348,48 428,58] establecen los límites entre los que se espera que esté el volumen de ventas de la editorial, en el evento de que las variables explicativas tomen los siguientes valores: PE=97,7; PC=113,5; GP=3.016,8 y R=5.712,1, al tiempo que el valor central 388,53 $(348,48 + 428,58/2 = 388,53)$, constituye la mejor aproximación puntual al valor que se espera obtener.

Emilio le comenta a su padre, que el valor pronosticado de las ventas de la editorial en el primer semestre de 2019 será de \$ 388,53, con una disminución de los gastos en publicidad del 10% con respecto al cuarto trimestre de 2018 (\$ 407,0), manteniendo las demás variables constantes. Emilio, además agrega: Como tú solo vives preocupado por el comportamiento de los gastos de publicidad, debes tener presente que una reducción del 10% en este tipo de gastos traerá consigo una disminución en las ventas de libros de 4,54%. ¡Papá!, ¿acaso, esto es lo que tú quieres?

Don Pablo responde: Pero si disminuimos el 10% en el gasto en publicidad nos ahorraremos un montón de dinero, así caigan las ventas en 4,54%. ¡Creo que esta podría ser una buena decisión!

Emilio replica: Otra vez, veo que cometes ligerezas en tu análisis financiero. El 10% de ahorro en publicidad equivale a \$ 0,34 millones, mientras que una caída de 4,54% en las ventas generará una disminución de \$ 18,5 millones en nuestros beneficios brutos. Ahora, ¿si entiendes la importancia de un buen modelo econométrico para la toma de decisiones empresariales?

Don Pablo y Emilio caen en cuenta que en el modelo de ventas de la editorial se les ha pasado por alto el fenómeno del incremento que experimentan las ventas en el periodo vacacional (julio a septiembre). Dicho fenómeno puede asociarse a que en la época veraniega los clientes disponen de más tiempo libre para la lectura.

Para comprobar esta hipótesis Emilio especifica el siguiente modelo:

$$V_t = \beta_0 + \beta_1 L_t + \mu_t \quad (13.5)$$

Donde: V_t = Ventas de la editorial y L_t = Número de horas semanales de lectura

Emilio mediante MCO estima este nuevo modelo, obteniendo los siguientes resultados:

$$\widehat{Ventas} = 332,943 + 47,197L_t \quad (13.6)$$

$$(11,08) \quad (22,16)$$

$$R^2 = 10,67\%; \quad \bar{R}_2 = 8,31\%$$

Con la prueba de hipótesis a un nivel de significación estadística de 5%, se concluye que existe un diferencial significativo en las ventas de la editorial cuando es verano con respecto al resto del año.

En el modelo se incluye el efecto estacional a través de la variable LG_t , evitando así sesgar las estimaciones y obteniendo al mismo tiempo una estimación precisa de los coeficientes, por lo tanto, el resultado obtenido es:

$$\widehat{Ventas} = 233,204 - 1,821PE_t + 0,194PC_t + 0,096GP_t + 0,003R_t + 0,0049LGP_t \quad (13.7)$$

$$(57,11) \quad (0,47) \quad (0,31) \quad (0,006) \quad (0,003) \quad (0,002)$$

$$R^2 = 93,55\%; \quad \bar{R}^2 = 92,60\%; \quad \Sigma \hat{\mu}^2 = 10.099,25$$

Para evitar más dudas de su padre, Emilio aplica al modelo obtenido, las pruebas disponibles para la detección de problemas de heteroscedasticidad y autocorrelación. Dichas pruebas confirman que el modelo no adolece de tales problemas que pueden poner en tela de juicio la eficiencia de los estimadores, y por ende, su capacidad explicativa y predictiva.

Aunque Don Pablo reconoce que el modelo estimado por Emilio cada vez ha adquirido más solidez y capacidad explicativa, aún le preocupa algo. Recuerda las palabras de su padre, quien solía decir que “*las ventas de la editorial en el siguiente trimestre tenían mucho que ver con lo que hubiera pasado en el trimestre anterior*”; aunque nunca supo a qué podría atribuirse ese efecto, pero aseguraba que si en un trimestre se vendía mucho, al trimestre siguiente, las ventas se incrementaban en algo más y viceversa.

Emilio para comprobar que tan cierto puede ser este efecto, en el modelo de ventas de la editorial introduce como variable explicativa, las ventas retardadas un periodo (V_{t-1}), obteniendo así, el modelo final.

Don Pablo expresa su alegría por el logro obtenido. Se resiste a creer que al final de cuentas, el modelito como lo llama él, haya dado respuesta a cada uno de los interrogantes expresados en su debido momento por cada uno de los dos. Ahora él entiende que las sucesivas discusiones, el tiempo y el esfuerzo invertidos valieron la pena. La editorial cuenta con un modelo robusto que explica y predice las ventas y permitirá tomar mejores decisiones comerciales.

$$\widehat{Ventas} = 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \quad (13.8)$$

$$(44,64) \quad (0,40) \quad (0,006) \quad (0,002) \quad (0,047)$$

$$R^2 = 93,67\%; \quad \bar{R}^2 = 92,92\%$$

13.7 Uso del modelo para fines de control o de política.

Como exigencia final, Don Pablo le pide a Emilio que simule un aspecto que siempre le ha inquietado: Si la editorial fijará como meta para el primer trimestre de 2019, incrementar sus ventas (V_{t+1}) en 10% ($\$407,0 \times 1,10 = \$447,7$), manteniendo constante el gasto en publicidad, considerando que el primer trimestre de 2019 no periodo vacacional, por consiguiente, $LGP = 0$ y $V_{t-1} = \$407,0$ (ventas retardadas un periodo, es decir, las ventas correspondientes al cuarto trimestre de 2018). ¿Cuál debe ser el precio promedio de los libros para obtener ventas por valor de $\$447,7$ en el primer trimestre de 2019?

Emilio procede a despejar la incógnita de interés de su padre, así:

$$\begin{aligned}\widehat{Ventas} &= 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \\ 447,7 &= 259,428 - 2,149PE_t + 0,098(3.352) + 0,0044(0) + 0,091(407,0) \\ 447,7 &= 259,428 - 2,149PE_t + 328,5 + 0 + 37,04 \\ 447,7 - 259,428 - 328,5 - 0 - 37,04 &= -2,149PE_t \\ 82,5 &= 2,149PE_t \\ PE_t &= \frac{-177,26}{-2,149} = \$ 82,5 \text{ miles}\end{aligned}$$

Emilio responde a su padre: ¡Mira papá!, si tu intención es incrementar las ventas de la editorial en el primer trimestre de 2019 en un 10%, deberás reducir el precio promedio de los libros en un 15,6% en el mismo periodo, manteniendo las demás variables constantes.

Emilio también invita a su padre a simular otra situación que le parece interesante: ¿cuál debe ser el gasto en publicidad, si la editorial pretende obtener ventas por valor de \$ 447,7 millones en el primer trimestre de 2019?

La incógnita es el gasto en publicidad (GP_t), por lo tanto:

$$\begin{aligned}\widehat{Ventas} &= 259,428 - 2,149PE_t + 0,098GP_t + 0,0044LGP_t + 0,091V_{t-1} \\ 447,7 &= 259,428 - 2,149(97,7) + 0,098 GP_t + 0,0044(0) + 0,091(407,0) \\ 447,7 &= 259,428 - 210,0 + 0,098GP_t + 0 + 37,04 \\ 447,7 - 259,428 + 210,0 - 37,04 &= 0,098GP_t \\ 361,19 &= 0,098GP_t \\ GP_t &= \frac{361,19}{0,098} = \$ 3.685,6 \text{ miles}\end{aligned}$$

Si la editorial quiere vender \$ 447,7 millones en el primer trimestre de 2019, deberá gastar en publicidad \$ 3.685,6 miles en el mismo trimestre, manteniendo fijas las demás variables. Si Don Pablo se propone incrementar

las ventas de la editorial en el primer trimestre de 2019 en un 10%, deberá también aumentar el gasto en publicidad en un 10% en el mismo periodo.

Emilio complementa su intervención diciendo: ¡Querido papá!, estos pequeños ejercicios nos permiten concluir que a partir del modelo final de ventas de la editorial, podríamos simular diferentes situaciones, por ejemplo, determinar el valor del gasto en publicidad en el tercer trimestre de 2019, conociendo de antemano, que en el periodo vacacional las ventas se elevan por efecto del incremento de las horas de lectura de los clientes de la editorial, en especial, en el grupo de mujeres con estudios universitarios.

Don Pablo cierra esta discusión reconociendo que el modelito final le sorprende aún más. En verdad, este modelito nos será muy útil para la futura toma de nuestras decisiones empresariales. Solo nos resta estar atentos a los cambios que surjan en el desempeño de la editorial para calibrarlo, actualizarlo y probar su poderío explicativo y predictivo.

LIMITACIONES DE LA ECONOMETRÍA

Como ha podido apreciarse a lo largo de este texto, la econometría es la combinación de teoría económica, métodos estadísticos y datos empíricos en el análisis económico moderno. Muchos estudiosos de la econometría concuerdan que esta disciplina se concibió como un intento sistemático y cuantitativo de otorgar mayor carácter científico a la ciencia económica con el propósito de acercarla más a las ciencias puras, es decir, a las ciencias naturales como la física, la química, la biología, la astronomía, la paleontología y la geología, que se apoyan en el razonamiento lógico y en el método experimental de las ciencias formales como la matemática y la lógica. Desde esta perspectiva, parece ser que la ciencia económica es muy atrevida en pretender parecerse a las ciencias formales; precisamente, esto es lo que busca al usar el instrumental matemático y estadístico en el estudio de la conducta humana.

Ekelund y Hébert (2005) afirman que, “el principal argumento para la continua formalización (es decir, matematización) de la economía es que la disciplina no puede llegar a ser verdaderamente científica hasta que no sea lo rigurosa y completa que debe ser una ciencia; en otras palabras, hasta que sus proposiciones fundamentales hayan sido contrastadas y probadas. La teoría sin verificación (o potencial verificación) tiene una utilidad limitada. Los hechos, sin una teoría forjada en la lógica de las matemáticas, carecen de significado. Por lo tanto, los economistas de la corriente principal argumentan que un mayor respeto por la economía, como disciplina científica e independiente, solo se conseguirá por medio de la constante aplicación de rigurosos instrumentos matemáticos y estadísticos” (p.643).

Continuando con estos autores, el moderno campo de la economía empírica, o econometría, hablando en términos generales, consiste en la aplicación de métodos matemáticos y estadísticos a los datos económicos, a fin de verificar y mejorar la teoría económica. Su objeto es explicar y predecir el comportamiento económico en el contexto de la teoría. Dentro de los límites de la inferencia estadística, la econometría intenta contrastar la teoría económica utilizando los datos históricos y prever los acontecimientos económicos utilizando una combinación de teoría económica y datos económicos.

Los métodos econométricos son fuertemente criticados por la corriente económica heterodoxa, en particular, la causalidad como soporte básico. Según Lorente (2018:23), la causalidad afirma que un evento genera o produce al otro, enunciado que va más allá de una simple asociación o de una correlación estadística, y que permite integrar la relación funcional dentro de un esquema más general y de tipo explicativo, es decir, en una teoría que podría señalar límites de validez para esa relación funcional, o sugerir nuevas variables y relaciones que enriquezcan la explicación o, en ciertos casos, descubrir alguna inconsistencia que obligue a revisar todos los supuestos fundamentales.

A continuación en términos jocosos, citaremos algunas críticas sobre la econometría de labios de verdaderos expertos en esta disciplina. Para Cuthberston et al, (1992), la econometría aplicada no puede concebirse de manera mecánica: necesita comprensión, intuición y habilidades. Para Hendry (1995:68) (...) por lo general atravesamos puentes sin preocuparnos por la solidez de su construcción, pues tenemos la certeza razonable de que alguien verificó con rigor los principios y prácticas de la ingeniería. Los economistas deben hacer esta verificación con los modelos, o al menos anexar la siguiente advertencia a su modelo: *“no nos hacemos responsables si al emplearse se provoca un colapso”*; y finalmente, Kennedy (1992:82), considera que la búsqueda de la “verdad” por parte de los economistas a lo largo de los años ha dado origen al punto de vista según el cual los economistas son personas que buscan en un cuarto oscuro un gato negro que no existe; a los econométricos por lo general se les acusa de haberlo encontrado”.

Parafraseando a Ekelund y Hébert (2005), algunos críticos mantienen serias reservas sobre esta opinión. Argumentan que la naturaleza de la ciencia social, de la que la economía es una parte, hace imposible la formulación y verificación exactas. Algunos de los problemas centrales de la econometría contemporánea se refieren a las formalizaciones inexactas o incompletas de la teoría económica y a diversas insuficiencias de las muestras de datos y de los errores aleatorios, inherentes a la medición de las variables. Hablando en general, las técnicas econométricas modernas son más apropiadas cuando las muestras de datos son grandes; con todo, en muchos casos, no existen grandes muestras de datos. Así pues, la cantidad y calidad de los datos económicos son a menudo insuficientes para la tarea que debe desarrollarse. En contraste con las condiciones que se dan en las ciencias físicas y naturales, la recogida de la mayoría de datos económicos no está determinada o diseñada de antemano para satisfacer contrastaciones de la teoría económica.

La economía es una ciencia social que estudia el comportamiento o conducta de los agentes económicos (hogares, empresas y gobierno). La escuela neoclásica considera que estos agentes son racionales y optimizadores, que toman decisiones económicas ancladas en la maximización de sus intereses y objetivos, esto no siempre resulta ser cierto. Los agentes económicos interactúan en ambientes de alta incertidumbre guiados por expectativas cambiantes y por pasiones, muchas veces. Para Ariely (2008), la irracionalidad también puede ser una fuente importante de aprendizaje en la toma de decisiones. Este autor no defiende que seamos esencialmente racionales sino que somos previsiblemente irracionales, es decir, somos constantes en nuestra irracionalidad o sistemáticamente irracionales. Sus experimentos le han permitido concluir que vivimos en dos mundos: uno regido por normas sociales y otro por normas mercantiles. Cuando nos comportamos con normas sociales solemos ser generosos y solidarios; cuando lo hacemos por las mercantiles, nos volvemos autónomos y egoístas.

La conducta económica no se comporta de manera lineal y de forma predecible. La irracionalidad económica es lo que mejor caracteriza a los agentes económicos. Si la conducta económica es más irracional que racional, los datos que se recojan del comportamiento humano no son experimentales, por ende, están sujetos a error y se constituyen en aproximaciones de los fenómenos económicos.

La econometría, si bien establece relaciones estadísticas entre variables a través de contrastes de significancia estadística como los estudiados en el capítulo 4 para rechazar o no la hipótesis nula, muchas veces es incapaz de establecer relaciones prácticas entre dichas variables. La significación estadística no siempre puede interpretarse como significación práctica o significación absolutamente económica.

Se cuestiona con bastante frecuencia la poca confiabilidad de los datos económicos. Estos no se derivan de experimentos controlados, sino que corresponde a datos observacionales aproximados del comportamiento humano, basados en alguna medida, en percepciones; por consiguiente, si los datos adolecen de deficiencias en su obtención y la teoría económica no es totalmente firme, no es justo que los métodos estadísticos (si es que en verdad, son perfectos) asuman toda la responsabilidad en la explicación de los fenómenos económicos.

La econometría pretende explicar el presente y pronosticar el futuro a partir de datos del pasado. La econometría da cuenta de su capacidad explicativa y predictiva en un rango de datos históricos (cuando se trabaja con series temporales) previamente definido por el investigador. Si este rango se modifica, también se alterarán los coeficientes de la regresión. Esto conduce a que los pronósticos realizados por la econometría sean de corto plazo y que los modelos estén en constante verificación, actualización y re-especificación.

En una entrevista publicada en *The Economic Journal* (mises.org), en 1939, John Keynes criticó el trabajo econométrico realizado por Jan Tinbergen - Premio Nobel de Economía 1969, argumentando que: a) la economía no es homogénea a lo largo del tiempo, lo que equivale a decir que los factores ignorados (nos referimos a las variables contenidas en el término de perturbación) en la modelación econométrica tienen que permanecer constantes, de forma que no alteren los descubrimientos empíricos; b) No todos los factores relevantes (factores políticos, sociales y psicológicos, incluyendo cosas como políticas públicas, el progreso de la invención y el estado de las expectativas) para explicar un fenómeno económico no tienen que ser cuantificables y medibles estadísticamente; c) las variables explicativas en un modelo no siempre son independientes (ausencia de multicolinealidad), sino

son independientes se corre el riesgo de obtener correlaciones espurias; y d) no siempre existe linealidad entre las variables, aun cuando la econometría ha desarrollado técnicas para transformar variables, y de este modo, linealizar ecuaciones, este procedimiento corre el riesgo de falsar teorías; en otras palabras, “torturar los datos para ajustar la realidad a una teoría pre-existente, convirtiendo a la econometría en una técnica de abuso y manipulación”.

Las anteriores consideradas no suponen en ningún momento que la econometría carezca de validez teórica y aplicada, sino más bien, que los modelos econométricos son limitados por la misma naturaleza social de la economía. Ahora, escuchemos las opiniones de los protagonistas de ésta didáctica y divertida historieta.

Don Pablo: Si en el proceso de construcción del modelo econométrico de las ventas de la editorial se pudo constatar, el poderío de la econometría para explicar y predecir fenómenos económicos cotidianos; no entiendo, ¿dónde estriban las debilidades de esta disciplina?

Emilio: ¡Papá!, tú recuerdas que cuando estimamos el modelo de ventas de la editorial, hablamos de los parámetros o coeficientes de la pendiente que acompañan las variables explicativas, ¿si lo recuerdas? Así pues, viene la primera limitación. Dichos coeficientes son solo una estimación de la relación causa-efecto que se teoriza. Dicho de otra forma, el investigador que emplee esta técnica no puede estar nunca enteramente seguro de que los coeficientes reflejen la relación verdadera. Solo puede confiar en que la estimación sea correcta dentro de determinados intervalos probabilísticos, es decir, un intervalo de confianza de un 5% significa que la técnica proporcionará la respuesta correcta el 95% de las veces. Las técnicas de regresión son incapaces de proporcionar una prueba concluyente de las hipótesis seleccionadas; no podemos estar nunca completamente seguros de que un estimador capta las auténticas relaciones. No obstante, la econometría puede desarrollar técnicas estadísticas que aumentan la confianza en las estimaciones.

Don Pablo: Estos argumentos esbozados por críticos de las imperfecciones y limitaciones de la econometría, ¿estarían señalando que no se debe utilizar esta disciplina en el análisis económico?

Emilio: Nunca pretendí decir esto. El desarrollo de la econometría le ha permitido a la ciencia económica explicar fenómenos que hubiese sido imposible comprenderlo sin la existencia de las técnicas econométricas hoy disponibles. Por ejemplo, hoy las autoridades económicas y los bancos centrales de los países pueden tomar mejores decisiones de política económica gracias a los modelos macroeconómicos disponibles; igualmente, nosotros podemos programar metas más ambiciosas de ventas de la editorial, manipulando variables clave como el precio de los libros y el gasto en publicidad.

Don Pablo: Pero al fin, ¿es recomendable o no utilizar las técnicas econométricas disponibles?

Emilio: Por supuesto que sí. El uso de la econometría en los estudios empíricos permite explicar fenómenos económicos y generar conocimiento valioso. Solo considero oportuno, que en su uso, se tengan en cuenta sus limitaciones y la naturaleza misma de la economía como ciencia social.

Don Pablo: He escuchado que la econometría ha tenido un desarrollo vertiginoso desde comienzos del siglo XX. ¿Este desarrollo aumenta el poderío de las técnicas econométricas?

Emilio: Ekelund y Hébert (2005), sostienen que aunque una teoría inadecuada y unos datos incompletos no son razones suficientes en sí mismas para rechazar los métodos cuantitativos, algunos críticos argumentan que los costos de diseño y recogida necesarios para asegurar una calidad elevada de los datos son prohibitivos. El trabajo incesante de expertos en esta disciplina está conduciendo a mejorar los métodos de estimación y de contrastación de hipótesis, y las agencias gubernamentales y organizaciones privadas están preocupadas por producir bases de datos confiables y actualizados para el ejercicio empírico. En este orden de ideas, es recomendable que el economista y el profesional de las ciencias sociales disponga de un arsenal de técnicas y métodos cuantitativos y cualitativos que bajo una combinación sabia, le permitan encontrar respuestas a problemas existentes. Debe recordarse que la economía incorpora proposiciones normativas y positivas.

Don Pablo: ¡Ya logré captar el mensaje! El uso de la econometría es valioso para el trabajo empírico y la investigación económica, pero hay que

comprender sus limitaciones y deficiencias en el campo científico, por lo tanto, hay que actuar con cautela y buen juicio. Si estuviésemos trabajando en física, en química o en astronomía, ¡créeme papá!, que las cosas serían completamente diferentes.

Emilio: ¡Papá!, veo que asimilaste magistralmente el mensaje. Pese a sus críticas y limitaciones, la economía para muchos, es la ciencia más dura entre las blandas. Precisamente, la econometría, les ha dado el carácter de científicidad, y que años tras año, la Real Academia de las Ciencias de Suecia galardone con el Premio del Banco de Suecia en Ciencias Económicas en memoria de Alfred Nobel a destacados científicos sociales.

Don Pablo: De este modo, damos por terminada nuestra discusión. Sin pretenderlo, a lo largo de este interesante ejercicio hemos puesto a prueba nuestras hipótesis y a través de instrumentos econométricos de fácil apropiación, especialmente con el uso de software, se han generado respuestas y conclusiones para mejorar el proceso de toma de decisiones comerciales. Si no hubiésemos emprendido esta tarea, de seguro, continuaríamos tomando decisiones a ciegas basadas en la intuición. ¿No te parece?

Emilio: Como la economía es cambiante, sólo nos resta mantener en alerta nuestro modelo. Este aprendizaje nos servirá para construir otros modelos de interés para la editorial, en procura de que este negocio pueda sostenerse en un mercado turbulento.

Don Pablo y Emilio: Ahora, comprendemos la importancia de los métodos cuantitativos en la gestión de negocios. En adelante, nos comprometemos a utilizarlos de manera responsable. Enseguida, estos dos pintorescos personajes se felicitan y abrazan por este aporte realizado a la disciplina de la econometría y por el gran legado que dejan a los estudiosos y usuarios de esta disciplina. ¡Colorín colorado...!

BIBLIOGRAFÍA

- Alonso Antón A.; Fernández Macho, J. y Gallastegui Zulaica, I. (2005). *Econometría*. Madrid, España: Pearson Prentice Hall Educación S.A.
- Angrist, J. D. y Pischke, J.S. (2016). *Dominar la econometría. El camino entre el efecto y la causa*. Barcelona, España: Antoni Bosch Editor.
- Ariely, Dan (2008). *Las trampas del deseo: cómo controlar los impulsos irracionales que nos llevan al error*. Madrid, España: Editorial Ariel.
- Carrascal Arranz, U.; González González, Y. y Rodríguez Prado, B. (2004). *Análisis econométrico con EVIEWS*. México, D.F., México: Alfaomega Grupo Editor S.A. de C.V, primera edición.
- Cuthberston, K.; Hall, S. G. & Taylor, M. P. (1992), *Applied Econometrics Techniques*. Michigan, USA: Michigan University Press.
- Dresdner Cid, J. y Vásquez Lavin, F. (2007). *Nociones de econometría intermedia*. Concepción, Chile: Universidad de Concepción, segunda edición.
- Ekelund R. B. y Hébert, R. F. (2005). *Historia de la teoría económica y su método*. México, D.F., México: McGraw-Hill/Interamericana Editores S.A. de C.V.
- Esteban González, M. V.; Moral Zuazo, M. P.; Orbe Mandaluniz, S.; Regúlez Castillo, M.; Zarraga Alonso, A. y Zubia Zubiaurre, M. (2009). *Econometría básica Aplicada con Gretl*. Bilbao, España: Departamento de Economía Aplicada III. Econometría y Estadística, Facultad de Ciencias Económicas y Empresariales. Universidad del País Vasco/Euskal Herriko Unibertsitatea. Recuperado de https://addi.ehu.es/bitstream/handle/10810/12496/08_09est.pdf?sequence=1&isAllowed=y
- Freedman, D., Pisani, R., Purves, R. y Adhikari, A. (1993). *Estadística*. Barcelona, España: Antoni Bosch editor.
- Greene, W. H. (1998). *Análisis econométrico*. Madrid, España: Prentice Hall, tercera edición.
- Gujarati, D. N. y Porter, D. C. (2010). *Econometría*. México, D.F., México: McGraw-Hill/Interamericana Editores S.A. de C.V, quinta edición.

- Hans Franses, P. (2004). *A concise introduction to econometrics. An intuitive guide*. Rotterdam. Institute Erasmus University, Cambridge University Press.
- Hendry, D. F. (1995). *Dynamic Econometrics*, Londres, Inglaterra. Oxford:University Press.
- Judge, G.G.; Hill, R.C.; Griffiths, W.E.; Lütkepohl, H. & Lee, T.C. (1988). *Introduction to the Theory and Practice of Econometrics*. USA. John Wiley & Sons, second edition.
- Kennedy, P. (1992). *A Guide to Econometrics*. The MIT Press, Cambridge Massachusetts,
- Lorente, L. (2018). La explicación causal en economía. *Revista de Economía Institucional*. 20(39). 9-51. DOI: <https://doi.org/10.18601/01245996.v20n39.02>.
- Maddala, G.S. (1992). *Introduction to econometrics*. New York, USA: MacMillan Publishing Company.
- Mendoza Bellido, W. (2014). *Cómo investigan los economistas. Guía para elaborar y desarrollar un proyecto de investigación*. Lima, Perú: Fondo Editorial Pontificia Universidad Católica del Perú, primera edición.
- Karl, F. I. (2016). *Keynes's Critique of Econometrics Is Surprisingly Good*. Mises Institute. Recuperado de: <https://mises.org/wire/keynes%E2%80%99s-critique-econometrics-surprisingly-good>.
- Pena Trapero, J. B.; Estavillo Dorado, J. A.; Galindo Frutos, M. E.; Leceta Rey, M. J. y Zamora Sanz, M. (1999). *Cien ejercicios de econometría*. Madrid, España: Ediciones Pirámide, primera edición.
- Pérez López, C. (2012). *Econometría básica. Aplicaciones con EVIEWS, STATA, SAS y SPSS*. Madrid, España: Editorial Garceta, Grupo Editorial.
- Pérez López, C. (2006). *Problemas resueltos de econometría*. Madrid, España: Ediciones Thomson Paraninfo S.A.
- Pérez López, C. (2016). *Técnicas avanzadas de predicción*. Madrid, España: Garceta Grupo Empresarial.
- Pindyck, R. S. y Rubinfeld, D. L. (2001). *Econometría: Modelos y pronósticos*. México, D.F., México: Mc-Graw-Hill/Interamericana Editores, S.A. de C.V.
- Portillo, F. (2006). *Concepto, método y evolución de la econometría*. La Rioja, España: Universidad de La Rioja, Departamento de Economía y Empresa. Recuperado de: <https://www.unirioja.es/cu/faporti/ieTEMA01.pdf>.

- Pulido San Román, A. y Pérez García, J. (2001). *Modelos econométricos*. Madrid, España. Ediciones Pirámide.
- Sampieri Hernández, R.; Fernández Collado, C. y Baptista Lucio, M. P. (2010). *Metodología de la investigación*. México, D.F., México: Mc-Graw-Hill/ Interamericana Editores, S.A. de C.V.
- Stock, J. H. y Watson, M. M. (2012). *Introducción a la econometría*. Madrid, España: Pearson educación, S.A.
- Varian, H. R. (2016). *Microeconomía intermedia. Un enfoque actual*. Bogotá, Colombia. Antoni Bosch editor – Alfaomega Colombiana S.A. Novena edición.
- Wooldridge, J. M. (2015). *Introducción a la econometría*. México, D.F., México: Cengage Learning.

Anexos: Bases de datos en medio magnético (formatos Excel y EViews)

- Tab1. Ventas de la editorial (modelo univariado).
- Tab2. Ventas de la editorial (modelo multivariado).
- Tab3. Efecto estacional en las ventas.
- Tab4. Horas semanales dedicadas a la lectura.
- Tab5. Modelo multivariado con horas dedicadas a la lectura.
- Tab6. Modelo multivariado con gasto en publicidad en época de verano.
- Tab7. Modelo de lectura con variables cualitativas y cuantitativas_1.
- Tab8. Modelo de lectura con variables cualitativas y cuantitativas_2.
- Tab9. Modelo de lectura con variables cualitativas y cuantitativas_3.

Este libro es un complemento a los excelentes manuales de econometría y de economía aplicada disponibles en las librerías en el mundo que sirven de soporte para la enseñanza-aprendizaje de la econometría, la cual forma parte de los planes de estudio de pregrado y posgrado en economía y en otras ciencias sociales. El propósito de esta obra es que el estudiante logre vencer el miedo cuando se enfrenta al estudio de esta disciplina y de servir de guía para el aprendizaje de algunos de sus aspectos fundamentales, considerando que la econometría es esencial en la investigación empírica moderna y en el ejercicio profesional del economista. Este objetivo académico se pretende lograr mediante el desarrollo de una historieta divertida que narra un hecho económico cotidiano en la economía de mercado, cuyos interrogantes son resueltos por los personajes que participan en ella, de manera secuencial y didáctica mediante el análisis econométrico. La pretensión del autor es que los estudiantes y otras personas interesadas en el estudio de este complejo tema, puedan comprender de manera fácil y amena, el instrumental básico de que dispone la econometría para el análisis económico aplicado.

ISBN: 978-958-5123-33-5

