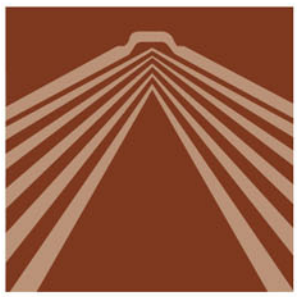
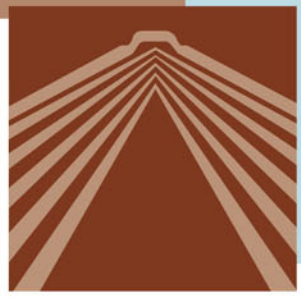
4.ª edición

Econometría

Montserrat Díaz Fernández
María del Mar Llorente Marrón





PIRÁMIDE

4.ª edición

Econometría

MONTSERRAT DÍAZ FERNÁNDEZ

CATEDRÁTICA DE UNIVERSIDAD DEL DEPARTAMENTO DE ECONOMÍA CUANTITATIVA DE LA UNIVERSIDAD DE OVIEDO

MARÍA DEL MAR LLORENTE MARRÓN

PROFESORA TITULAR DE UNIVERSIDAD DEL DEPARTAMENTO DE ECONOMÍA CUANTITATIVA DE LA UNIVERSIDAD DE OVIEDO

4.ª edición

Econometría

EDICIONES PIRÁMIDE

COLECCIÓN «ECONOMÍA Y EMPRESA»

Director:
Miguel Santesmases Mestre
Catedrático emérito de la Universidad de Alcalá

Edición en versión digital

Está prohibida la reproducción total o parcial de este libro electrónico, su transmisión, su descarga, su descompilación, su tratamiento informático, su almacenamiento o introducción en cualquier sistema de repositorio y recuperación, en cualquier forma o por cualquier medio, ya sea electrónico, mecánico, conocido o por inventar, sin el permiso expreso escrito de los titulares del copyright.

© Montserrat Díaz Fernández y María del Mar Llorente Marrón, 2013
© Primera edición electrónica publicada por Ediciones Pirámide (Grupo Anaya, S. A.), 2013
Para cualquier información pueden dirigirse a piramide_legal@anaya.es
Juan Ignacio Luca de Tena, 15. 28027 Madrid
Teléfono: 91 393 89 89
www.edicionespiramide.es
ISBN digital: 978-84-368-2876-4

«Todo economista es un económetra, le agrade o no le agrade, porque mientras no seamos capaces de explicar nuestros argumentos en cifras, la voz de nuestra ciencia, aunque ocasionalmente pueda ayudar a dispersar errores groseros, nunca será oída por los hombres prácticos. Son, por instinto, económetras todos, en su desconfianza de las cosas no sujetas a una prueba exacta.»

SCHUMPETER (1933): «The common sense in Econometrics».
Econometrica, vol. 1

A Víctor y a mis hijas, Elia y Montse
(Montserrat Díaz)

A Manuel y a mis hijas, Lidia y Andrea
(María del Mar Llorente)

Índice

Prólogo 17

PARTE PRIMERA
El modelo de regresión

1. Metodología de la investigación econométrica 23

1. Modelos econométricos 23
2. Especificación del modelo 25
   2.1. Variables del modelo 26
   2.2. Signos y magnitudes de los parámetros 28
   2.3. Forma matemática del modelo 29
3. Estimación del modelo 31
4. Evaluación de las estimaciones 32
5. Evaluación del poder predictivo del modelo estimado 33

2. El modelo de regresión lineal simple 35

1. Introducción 35
2. El modelo de regresión lineal simple 35
3. Supuestos del modelo de regresión 41
   3.1. Supuestos estocásticos 41
   3.2. Supuestos no estocásticos 42
4. Obtención de los estimadores mínimo-cuadráticos 43
5. Contrastes estadísticos de los estimadores mínimo-cuadráticos 52
   5.1. Media de los estimadores mínimo-cuadráticos 52
   5.2. Varianza de los estimadores mínimo-cuadráticos 54
   5.3. Covarianza de los estimadores mínimo-cuadráticos 56
   5.4. Estimación de la varianza de la variable aleatoria 57

© Ediciones Pirámide

6. Propiedades de los estimadores mínimo-cuadráticos 62
7. Coeficiente de determinación 68
8. Distribución de probabilidad de la variable aleatoria 75
8.1. Supuesto de normalidad 75
8.2. Propiedades de los estimadores mínimo-cuadráticos bajo el supuesto de normalidad 76
9. Formas funcionales de los modelos de regresión 78
9.1. Regresión a través del origen 78
9.2. Modelos log-log, doble logaritmo o de elasticidad constante 82
9.3. Modelos semilogarítmicos 83
9.4. Transformaciones recíprocas 84

3. Inferencia estadística en el modelo de regresión lineal simple 97

1. Introducción 97
2. Intervalos de confianza para los coeficientes de regresión 99
2.1. Intervalo de confianza para β_1 99
2.2. Intervalo de confianza para β_0 101
3. Intervalo de confianza para σ_u^2 102
4. Contrastes de hipótesis 106
4.1. Introducción 106
4.2. Contrastes de hipótesis para los parámetros de la regresión 108
5. Análisis de la varianza. 115

4. El modelo lineal general 125

1. Introducción 125
2. Supuestos del modelo 127
2.1. Supuestos estocásticos 127
2.2. Supuestos no estocásticos 129
3. El estimador mínimo-cuadrático. Propiedades 132
3.1. Ajuste mínimo-cuadrático 132
3.2. Estimación del modelo en desviaciones 135
3.3. Propiedades de la función de regresión muestral 137
3.4. Propiedades de los estimadores mínimo-cuadráticos 138
3.5. Matriz de varianzas-covarianzas del vector $\hat{\bar{\beta}}$ 141
3.6. Estimación de la varianza del término de perturbación aleatoria 143
4. El coeficiente de determinación 148
5. Inferencia estadística en el modelo de regresión lineal general 156
5.1. Distribución de probabilidad del vector $\bar{u}$ 156
5.2. Intervalos de confianza y contrastes de hipótesis en el modelo de regresión lineal general 157
5.3. Análisis de la varianza en el modelo general 159
5.4. Otros contrastes con relación al comportamiento de los parámetros del modelo de regresión 164
5.4.1. Contribución marginal de un subconjunto de variables explicativas 164

© Ediciones Pirámide

5.4.2. Prueba de Chow 166
5.4.3. Restricciones lineales sobre los parámetros del modelo 167
6. Estimación por máxima verosimilitud 170

5. Predicción 183

1. Introducción 183
2. Predicción puntual 184
3. Predicción por intervalos 185
3.1. Intervalo de confianza para el valor medio teórico de predicción, μ_p .. 185
3.2. Intervalo de confianza para un valor individual de predicción, Y_p 189
4. Evaluación de la capacidad predictiva de un modelo 191

6. Variables ficticias 201

1. Introducción 201
2. Interpretación y uso de las variables ficticias 202
2.1. Variables ficticias como factor explicativo del modelo 202
2.2. Especificación aditiva de las variables ficticias 205
2.3. Especificación multiplicativa de las variables ficticias 207
2.4. Otros usos de las variables ficticias 209
2.4.1. Las variables ficticias en el análisis estacional 210
2.4.2. Regresión lineal por tramos 211
3. Variables endógenas cualitativas 212
3.1. Modelo de probabilidad lineal 212
3.2. Modelo probit 214
3.3. Modelo logit 215

Ejercicios. El modelo de regresión lineal 231

PARTE SEGUNDA
Vulneración de las hipótesis básicas

7. Relajación de las hipótesis básicas 245

1. Hipótesis básicas en el modelo de regresión 245
2. Errores de especificación en el modelo 248
3. La hipótesis de normalidad. Contraste de Jarque-Bera 249

8. Autocorrelación 253

1. Naturaleza del problema 253
2. Causas y efectos 254
3. Consecuencias de la autocorrelación 255

© Ediciones Pirámide

4. Formas de detectar el problema 258
4.1. Método gráfico 259
4.2. Prueba Durbin-Watson 260
4.3. Contraste de Wallis 264
4.4. Prueba *h* de Durbin 265
4.5. Contraste de Breusch-Godfrey 266
5. Estimación del modelo bajo un esquema AR(1) 267
5.1. Ecuación de diferencias generalizadas 267
5.2. Métodos de estimación de ρ 269
5.2.1. El método de primera diferencia 269
5.2.2. ρ basado en el estadístico Durbin-Watson 270
5.2.3. Método iterativo de Cochrane-Orcutt 270
5.2.4. Método Durbin-Watson en dos etapas 271
5.2.5. Método de Hildreth-Lu 272

9. Multicolinealidad 289

1. Introducción 289
2. Consecuencias de la multicolinealidad 291
2.1. Ausencia de relación lineal entre las variables explicativas 292
2.2. Relación lineal exacta entre las variables explicativas 293
2.3. Relación lineal no exacta entre las variables explicativas 294
3. Cómo detectar la multicolinealidad 297
4. Soluciones a la multicolinealidad 299
4.1. Actuaciones sobre la muestra 299
4.2. Establecimiento de restricciones sobre el comportamiento de los parámetros poblacionales 300
4.3. Transformación de variables 300
4.4. Eliminación de variables 301

10. Heteroscedasticidad 305

1. Naturaleza de la heteroscedasticidad 305
2. Consecuencias de la heteroscedasticidad 308
3. Formas de detectar el problema. 309
3.1. Prueba de Park 310
3.2. Prueba de Glejser 310
3.3. Contraste de Goldfeld y Quandt 311
3.4. Contraste de White 313
3.5. Contraste de picos 314
3.6. Contraste de rangos de Spearman 314
3.7. Contraste de Breusch-Pagan 315
4. Soluciones a la heteroscedasticidad 316
4.1. Varianza del término aleatorio conocida. Método de los mínimos cuadrados generalizados 316
4.2. Varianza del término aleatorio desconocida 317

© Ediciones Pirámide

Ejercicios. El modelo de regresión lineal. Vulneración de hipótesis básicas 339

PARTE TERCERA
Introducción a los modelos de ecuaciones simultáneas

11. Modelos de ecuaciones simultáneas 353

1. Los modelos de ecuaciones simultáneas 353
2. El método mínimo-cuadrático para la estimación de los modelos de ecuaciones simultáneas 355
3. Especificación de un modelo de ecuaciones simultáneas. Notación y definiciones 358
4. El problema de la identificación 361
5. La estimación de los modelos de ecuaciones simultáneas 365
   - 5.1. Mínimos cuadrados indirectos 365
   - 5.2. Estimación por variables instrumentales 366
   - 5.3. Mínimos cuadrados en dos etapas 367
   - 5.4. Otros métodos de estimación 367
   - 5.5. Sistemas recursivos 368
   - 5.6. Comparación entre los distintos estimadores 368
   - 5.7. Evaluación de resultados en los modelos de ecuaciones simultáneas 369

Ejercicios. Introducción a los modelos de ecuaciones simultáneas 371

Algunas cuestiones de seguimiento tipo test 373

Anexo. Tablas estadísticas 447

Bibliografía 461

© Ediciones Pirámide

Prólogo

La Econometría constituye un aspecto fundamental en la formación del economista de cara a su futuro ejercicio profesional, centrado en el ámbito del asesoramiento, tanto a entidades públicas como privadas.

El manual que ahora se presenta, *Econometría*, nace como resultado de la experiencia docente desarrollada en esta disciplina en la Facultad de Ciencias Económicas y Empresariales de la Universidad de Oviedo, desde el curso 1984-1985. Merced a los comentarios, sugerencias, preguntas y opiniones de alumnos y profesores se han ido perfilando los aspectos fundamentales que ahora se abordan y que pretenden constituir una modesta contribución a la enseñanza de esta disciplina, sin duda fundamental en el futuro ejercicio profesional de los actuales estudiantes de *Economía, Administración y Dirección de Empresas* y *Contabilidad y Finanzas*.

El texto, que intenta sintetizar los aspectos más relevantes que constituyen un curso básico de Econometría, no exige unos conocimientos de Álgebra matricial, Cálculo o Estadística sustancialmente diferentes de los establecidos en los actuales planes de estudio. Su planteamiento es, no obstante, riguroso en el sentido formal del lenguaje, en sus aspectos matemáticos, estadísticos y económicos, a la vez que comprensible y asequible para los alumnos que se enfrentan por primera vez al estudio de esta asignatura.

El texto ha sido concebido para ser utilizado, fundamentalmente, como manual básico del curso de Econometría que los alumnos de *Economía*, *Administración y Dirección de Empresas* y *Contabilidad y Finanzas* deben cursar en su licenciatura, como asignatura troncal obligatoria. No obstante, dado que su contenido ofrece una visión amplia y de conjunto de los aspectos básicos a incluir en un curso de Econometría, el manual puede tener otros usos alternativos, como libro de introducción a esta disciplina en cursos dirigidos

© Ediciones Pirámide

a profesionales no especialistas, cuyo ejercicio profesional se desarrolle tanto en el sector público como privado.

El libro se ha estructurado en tres bloques correspondientes a once capítulos que satisfacen el temario tradicional de un manual de Econometría.

Después de introducir, en el primer capítulo, los aspectos metodológicos a tener en cuenta en la construcción de un modelo econométrico, pasaremos a analizar la especificación más sencilla, el modelo de regresión lineal simple, y posteriormente el múltiple. El contenido del libro se refiere fundamentalmente a los modelos uniecuacionales. Tendremos en cuenta aspectos relativos tanto a la estimación del modelo como a la evaluación de resultados, bajo la adopción de criterios económicos, estadísticos y econométricos. Estas cuestiones las desarrollaremos ampliamente entre los capítulos 2 y 4.

En el capítulo 5 abordaremos un aspecto esencial en el ámbito económico: la predicción. La adopción de cualquier medida de Política Económica se apoya precisamente en esta cuestión.

La consideración de factores cualitativos en la especificación del modelo se aborda en el capítulo 6. Nos centraremos, fundamentalmente, en su consideración como variables explicativas. En el contexto social en el que desarrollamos nuestra actividad, la consideración de factores de carácter tanto cuantitativo como cualitativo se hace imprescindible de cara a obtener una especificación correcta del modelo econométrico.

En la segunda parte del texto, capítulos 7-10, analizaremos pormenorizadamente las hipótesis básicas establecidas en el modelo de regresión, bajo distintos aspectos. Trataremos de cuestionar su plausibilidad teórica, analizaremos las consecuencias que se derivan de su vulneración, el establecimiento de criterios que permitan diagnosticar satisfactoriamente el problema, así como la adopción de medidas que permitan recuperar la salud del modelo. Analizaremos, individualmente, los problemas más habituales dentro del campo de la Econometría Aplicada, esto es, autocorrelación, multicolinealidad y heteroscedasticidad.

Por último, en el tercer bloque, capítulo 11, haremos una breve introducción al análisis de los modelos de ecuaciones simultáneas. El establecimiento de los conceptos fundamentales y las condiciones de identificabilidad del modelo, así como de alguna de las técnicas más habituales para su estimación, completará este breve análisis.

Hemos tratado de abordar cada capítulo también desde una perspectiva empírica, mediante la realización detallada de algunos ejercicios. Los programas informáticos permiten resolver de forma simplificada dichos ejemplos. De cara al alumno, la utilización del programa informático resulta mucho más atractiva que la resolución «manual» de los ejercicios; sin embargo, con-

© Ediciones Pirámide

sideramos esta última mucho más fructífera, tanto para la consolidación de los contenidos teóricos como para la interpretación adecuada de los resultados.

Por último, únicamente nos queda por asumir como propios los errores que pudieran encontrarse en el desarrollo del texto.

Oviedo, 30 de octubre de 2012.

Dra. Montserrat Díaz Fernández
Catedrática de Universidad del Departamento de Economía Cuantitativa
de la Universidad de Oviedo

Dra. María del Mar Llorente Marrón
Profesora Titular de Universidad del Departamento de Economía Cuantitativa
de la Universidad de Oviedo

© Ediciones Pirámide

PARTE PRIMERA
El modelo de regresión

1 Metodología de la investigación econométrica

1. MODELOS ECONOMÉTRICOS

El estudio de la adopción de decisiones por parte de los agentes económicos, además del análisis de sus consecuencias, constituye, en un sentido amplio, el objeto de la ciencia económica.

La Econometría se ocupa de la medición y verificación empírica de las relaciones económicas. Esta disciplina estudia, básicamente, las estructuras que permiten analizar las características y propiedades de una variable económica, utilizando como causas explicativas otras de la misma naturaleza.

Etimológicamente, el término «Econometría» significa «medición económica», si bien su campo de acción es mucho más amplio, como señala Frisch en el editorial del primer número de la revista *Econometrica* al indicar cuáles serían los objetivos de la *Econometric Society:* «*... su primer objeto sería promover estudios que se dirijan a una unificación de la aproximación teórico-cuantitativa y empírico-cuantitativa a los problemas económicos, y que constituyan reflexiones constructivas y rigurosas similares a las que han llegado a dominar las Ciencias Naturales. Existen, sin embargo, varios aspectos dentro del enfoque cuantitativo de la Economía, y ninguno de estos aspectos en sí mismo debe confundirse con la Econometría. Así, la Econometría no es Estadística Económica. Tampoco es lo que llamamos Teoría Económica, aunque una parte considerable de esta teoría tiene, definitivamente, carácter cuantitativo. La Econometría tampoco debe ser considerada como sinónimo de aplicación de las Matemáticas a la Economía. La experiencia ha demostrado que cada uno de estos tres puntos de vista, el de la Estadística, la Teoría Económica y las Matemáticas, es necesario, pero por sí mismo no es condición suficiente para una comprensión real de las relaciones cuantitativas en la vida económica moderna. Es la unión de los tres aspectos lo que constituye una*

© Ediciones Pirámide

herramienta de análisis potente. Es esta unión lo que constituye la Econometría».

Un *modelo* se define como una representación simplificada de la realidad que, para ser operativo, deberá ser *plausible* y *manejable.* En su estructura deberá incorporar los elementos fundamentales de la situación que trata de modelizar, de tal manera que las simplificaciones introducidas en su construcción permitan analizar y obtener sin dificultad conclusiones acerca de aquélla. Aunque el modelo sea una representación simplificada de la realidad, debe permitir la reproducción de los patrones de comportamiento existentes entre las variables económicas.

Un *modelo econométrico* es un modelo económico que incluye las especificaciones necesarias para su aplicación empírica. No obstante, existen ciertas diferencias entre ambos. Los modelos econométricos exigen una especificación más precisa de las variables que los integran, así como una relación funcional definida; tienen, además, carácter aleatorio.

En razón de la finalidad para la que se elaboren, los modelos econométricos se pueden clasificar en modelos explicativos, predictivos y modelos de decisión.

Con los *modelos explicativos* se tratará de captar cuantitativamente el mecanismo de funcionamiento económico, para lo cual será necesaria la estimación de sus parámetros a partir de la información contenida en una muestra. La finalidad inmediata de los *modelos predictivos* será la obtención de pronósticos acerca de la evolución futura de determinadas variables económicas. Si el objetivo del modelo econométrico fuese la *toma de decisiones,* se haría necesario incorporar en su estructura algunas variables sobre las que el decisor pudiese ejercer una influencia directa, denominadas *variables instrumentales*.

La elaboración de un modelo econométrico se puede dividir, básicamente, en las siguientes etapas: *especificación* o formulación de las hipótesis establecidas con las que se intenta medir el fenómeno; *estimación* de sus parámetros, por medio de técnicas econométricas apropiadas; *evaluación* de los resultados, sobre la base de criterios económicos, estadísticos y econométricos, y finalmente valoración de su *capacidad predictiva* (figura 1.1).

Si bien en una primera aproximación estas etapas siguen un orden secuencial, la elaboración de un modelo no constituye un proceso lineal que siga un orden preestablecido. Será necesaria la confrontación continua de aquél con los datos y cualquier otra fuente de información, con objeto de obtener un modelo compatible con los mismos que permita analizar la realidad, ofrezca mejores predicciones y constituya una buena base para la toma de decisiones.

© Ediciones Pirámide

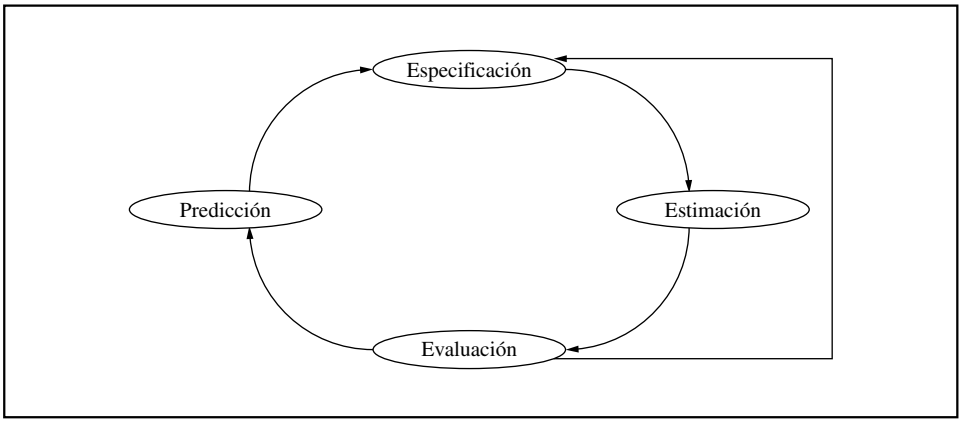


Figura 1.1. Etapas a seguir en la construcción de un modelo econométrico.

2. ESPECIFICACIÓN DEL MODELO

En la elaboración de un modelo econométrico, la *especificación,* o expresión matemática de la relación económica considerada, constituye la primera etapa del proceso. Esta fase requiere la determinación de las variables a incluir en el mismo, el conocimiento de expectativas teóricas acerca del signo y tamaño de los parámetros de la función, así como la elección de la forma matemática de la relación económica (figura 2.1).

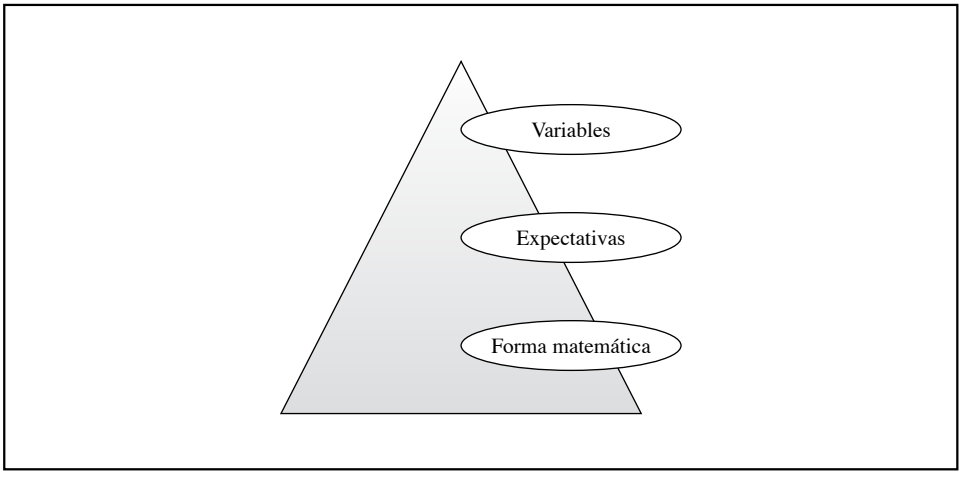


Figura 2.1. Especificación del modelo.

© Ediciones Pirámide

La Teoría Económica constituye, en general, la base que dirige la especificación de todo modelo econométrico. En ella se encuentran orientaciones sobre qué variables pueden ser relevantes para explicar un determinado fenómeno. La información acerca del carácter, positivo o negativo, de la influencia de cada variable explicativa, así como la magnitud o los límites de su variación, son también recogidos por la Teoría Económica. Sin embargo, en pocas ocasiones se establece de forma explícita la forma funcional de la relación considerada, siendo, por tanto, su especificación fruto de un trabajo casi exclusivamente matemático.

2.1. Variables del modelo

A partir de las fuentes de información disponibles, el económetra establecerá, en primer lugar, una relación de variables relevantes para explicar, de forma satisfactoria, el comportamiento de la variable dependiente. La Teoría Económica aportará orientaciones al respecto, en cada caso particular.

Supongamos, por ejemplo, que estamos estudiando la demanda del producto Z. La teoría de la demanda establece como determinantes de ésta el precio del bien considerado, el precio de los demás bienes (únicamente los sustitutivos y complementarios), además del nivel de renta de los consumidores y su estructura de preferencias. La función de demanda se expresaría, por tanto, como una relación de carácter determinista,

$$Q_Z = f(P_Z, P_0, Y, T) \tag{2.1.1}$$

siendo:

Q_Z = cantidad demandada del producto Z.
P_Z = precio del producto Z.
P_0 = precio de otros bienes (sustitutivos y complementarios).
Y = renta de los consumidores.
T = una medida adecuada de los gustos de los consumidores.

Además de lo postulado por la Teoría Económica, estudios empíricos al respecto pueden aportar información adicional acerca de cuáles son las variables relevantes que explican el comportamiento de la variable dependiente. La información acerca de las condiciones individuales en cada caso particular y de la conducta actual de los agentes económicos, consumidores o productores, completará el conocimiento de la teoría e investigación aplicada.

A modo de síntesis, diremos que el número de variables a incluir en el modelo dependerá tanto de la naturaleza del fenómeno a analizar como del ob-

© Ediciones Pirámide

jeto de la investigación. Normalmente, se introducirán explícitamente en la función únicamente aquellas variables que se consideren relevantes. La influencia del resto de factores se tendrá presente mediante la introducción en el modelo de un término de perturbación aleatoria, cuya distribución de probabilidad se determinará después de establecer algunas hipótesis plausibles acerca de su comportamiento. En consecuencia, la relación (2.1.1), de carácter determinista, se expresará en términos econométricos

$$Q_Z = f(P_Z, P_0, Y, T, u) \tag{2.1.2}$$

como una relación aleatoria, donde u denota precisamente dicha variable.

Una primera clasificación de las variables de un modelo econométrico divide a éstas en variables observables y no observables (figura 2.1.1).

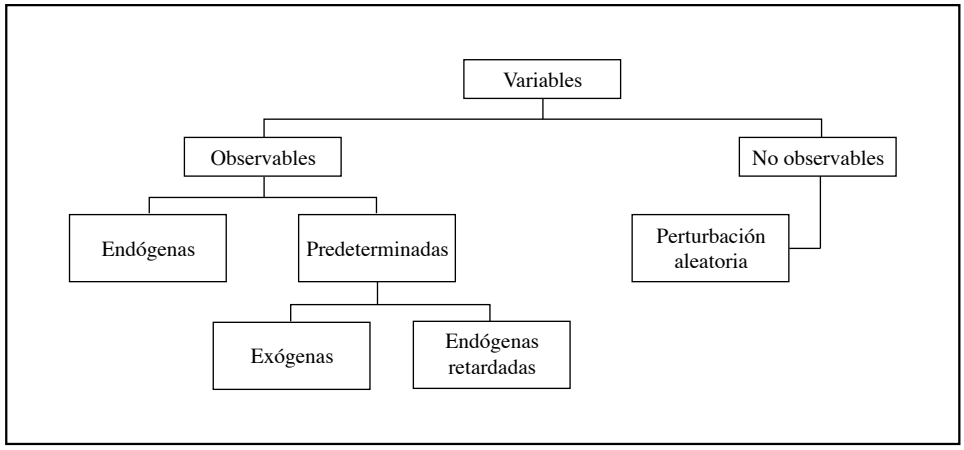


Figura 2.1.1. Variables que intervienen en un modelo econométrico.

Atendiendo al criterio de causalidad, las *variables observables* se pueden clasificar en variables endógenas y predeterminadas. Las primeras serán aquellas que influyen y son a su vez influenciadas por otras variables, determinándose dentro del sistema, mientras que las segundas influyen pero no son influidas por otros factores. Las variables predeterminadas se subdividen a su vez en *exógenas,* cuando se determinan fuera del modelo objeto de estudio, y *endógenas desplazadas,* si lo hacen dentro de éste en unidades temporales anteriores.

El *término de perturbación aleatoria, u,* sustituye a todas aquellas variables que han sido excluidas del modelo, pero que conjuntamente influyen en el comportamiento de la variable dependiente.

© Ediciones Pirámide

Las razones fundamentales que justifican la incorporación del término de perturbación aleatoria al modelo econométrico son, básicamente, las que se exponen a continuación:

1.ª Si existe alguna teoría que determine el comportamiento de la variable dependiente, ésta suele ser incompleta. Por tanto, la variable aleatoria podría ser utilizada como un sustituto de todas las variables excluidas del modelo.
2.ª Aun sabiendo cuáles son las variables relevantes en la especificación del modelo, es muy probable que no haya información disponible acerca de todas ellas. En el análisis empírico es muy común que los datos que todos deseamos tener no siempre sean accesibles.
3.ª Es posible que la influencia individual de variables poco relevantes en el comportamiento de la variable dependiente sea insignificante y no sistemática, y que, por razones de coste, no justifique su introducción explícita en el modelo. Cuando esto suceda, el efecto combinado de dichas variables podría ser tratado como una variable aleatoria.
4.ª Aun suponiendo que en la especificación del modelo se hayan incluido todas las variables relevantes, es preciso tener en cuenta el carácter aleatorio de la conducta humana. La variable aleatoria *u* podría reflejar también dicho comportamiento.
5.ª El modelo econométrico deberá ser sencillo en su estructura, incorporando en su especificación las variables más significativas. Si se puede explicar el comportamiento de la variable dependiente con dos o tres variables, y si, además, la teoría sobre la que se sustenta el estudio no es lo suficientemente sólida, representaremos por *u* a todas las demás.

Por estas y otras razones, el término de perturbación aleatoria, *u*, desempeñará un papel crítico en el análisis de regresión.

2.2. Signos y magnitudes de los parámetros

Las fuentes de información mencionadas anteriormente ofrecerán sugerencias acerca del *signo* de los parámetros y, posiblemente, de su *tamaño* o rango.

Continuando con el ejemplo propuesto, y suponiendo que la relación funcional recogida en (2.1.2) se particulariza en una función de demanda lineal:

$$Q_Z = \beta_0 + \beta_1 P_Z + \beta_2 P_j + \beta_3\ Y + u \qquad (2.2.1)$$

tendremos en cuenta que, de acuerdo con la teoría general de la demanda, β_1 será menor que cero, dado que postula una relación inversa entre la cantidad

© Ediciones Pirámide

demandada y el precio del bien; β_3 será positivo, puesto que la renta y la cantidad demandada estarán directamente relacionadas, excepto en el caso de que los bienes considerados sean inferiores, y β_2 será positivo o negativo si los bienes *j* y *z* son sustitutivos o complementarios, respectivamente.

Obsérvese que, con relación a la magnitud de los parámetros, los diferentes β_i de (2.2.1) representan conceptos económicos tales como elasticidad, propensión marginal u otras magnitudes del mismo carácter derivadas de la Teoría Económica. En una función de demanda lineal, tal como la de nuestro ejemplo, dichos parámetros representan componentes de las elasticidades relevantes.

2.3. Forma matemática del modelo

La Teoría Económica, en algunos casos, indica la forma matemática precisa de las relaciones, o el número de ecuaciones a incluir en el modelo considerado. La teoría de la demanda, por ejemplo, no determina si para un bien particular ésta debe ser estudiada con un modelo de una sola ecuación o con un sistema de ecuaciones simultáneas. Tampoco precisa si la relación funcional es lineal o no, y, sin embargo, contiene información acerca de su forma matemática.

La teoría estática de la demanda se apoya en el supuesto de conducta racional y ausencia de *ilusión monetaria* en el comportamiento del consumidor. Este supuesto significa que si todos los precios y rentas cambiasen en la misma proporción, el consumidor racional no debería alterar sus pautas de consumo. En consecuencia, la función de demanda tendría que adoptar una expresión que tuviese en cuenta su racionalidad, mediante el uso de una función homogénea de grado cero.

En la mayoría de los casos, la Teoría Económica no establece explícitamente la forma matemática de las relaciones económicas. El conocimiento de la función que liga las variables consideradas en el análisis permitirá conocer el tipo de relación existente entre las mismas. Dicho conocimiento constituirá en una fase posterior la base que justifique la adopción de medidas de carácter económico.

La representación gráfica de los datos reales en diagramas de dos dimensiones resulta una opción de gran utilidad en este sentido. Por una parte, se recurre a la teoría establecida para justificar explicaciones de los hechos y, por otra, a técnicas diferentes que permitan completar el análisis.

Normalmente, el económetra experimentará con varias relaciones funcionales, eligiendo de entre los distintos resultados aquellos que considere más satisfactorios.

© Ediciones Pirámide

Continuando con el ejemplo recogido en la función de demanda (2.1.2), supongamos que la variable dependiente, cantidad demandada del bien Z, Q_Z, se explica, en una primera aproximación, únicamente en función del precio de dicho bien, P_Z. La información muestral de tamaño n correspondiente a ambas variables se representará gráficamente, pudiendo adoptar distintas formas (figura 2.3.1). En el primer caso (figura 2.3.1.a) se observa la existencia de una dependencia inversamente proporcional entre ambas variables, dado que un valor bajo de Q_Z se asocia a un valor alto de P_Z, y viceversa. La opción más probable para el económetra a la hora de decidir la relación funcional de (2.1.2) será la lineal, si bien la elección de una función parabólica resultaría, a priori, la más adecuada si la representación gráfica se correspondiese con la figura 2.3.1.c. Cuando la información muestral recoja situaciones como la relativa a la figura 2.3.1.b, donde la nube de puntos refleja una situación de regularidad aparente, el económetra se verá obligado a obtener información adicional con objeto de precisar la relación funcional de (2.1.2).

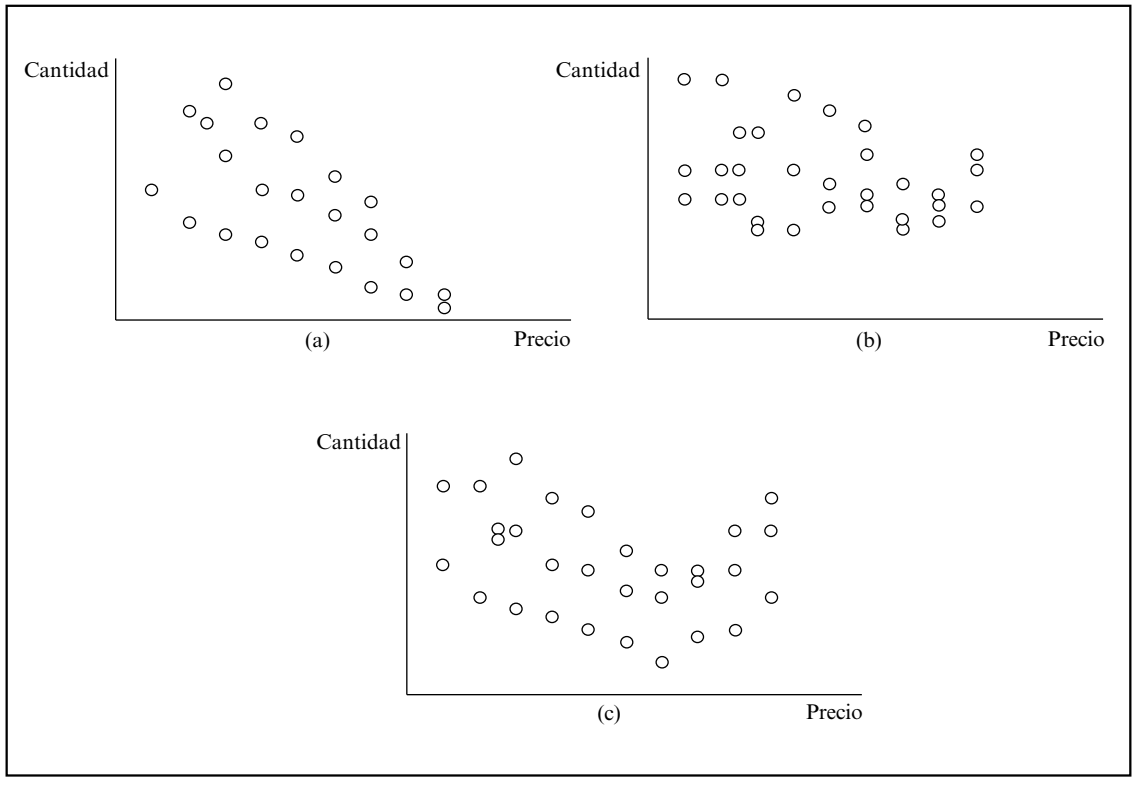


Figura 2.3.1. Información muestral. Diagrama de dispersión.

© Ediciones Pirámide

La especificación constituye la etapa más *importante* y *difícil* de la investigación econométrica, siendo, a menudo, el punto más débil de la mayor parte de las aplicaciones. Algunas de las razones que explican una especificación incorrecta de los modelos económicos son fácilmente detectables. La imperfección y flojedad de las teorías establecidas, la limitación de nuestro conocimiento en relación a los factores que son operativos en cada caso particular y los obstáculos derivados de los requisitos para la obtención de datos en la estimación de grandes modelos, constituyen algunos ejemplos de dicha cuestión. El error más frecuente de la especificación lo constituye la omisión de variables y ecuaciones y la elección incorrecta de la relación funcional. En este sentido, debe señalarse que la mayoría de los métodos econométricos son sensibles a los errores de especificación.

3. ESTIMACIÓN DEL MODELO

Una vez que el modelo ha sido especificado, el económetra deberá proceder a su estimación, con objeto de obtener la estimación numéricas de sus parámetros. La estimación constituye una etapa estrictamente técnica que requiere el conocimiento de las distintas técnicas econométricas, sus hipótesis de funcionamiento e implicaciones económicas.

En su desarrollo, esta etapa incluye, básicamente, los aspectos señalados en la figura 3.1.

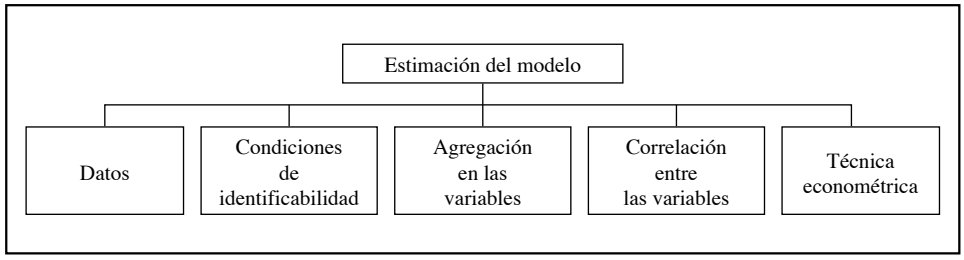


Figura 3.1. Estimación del modelo.

1.º *Recolección de las observaciones estadísticas de las variables incluidas en el modelo*. Los *datos* utilizados para la estimación del modelo considerado pueden presentarse en forma de *serie temporal,* suministrando información acerca de los valores numéricos de las variables a lo largo del tiempo; *cross-section* o *transversales* cuando se refieran a un conjunto de observaciones de una determinada variable, observada en

© Ediciones Pirámide

distintas unidades económicas y una misma unidad temporal; o *panel,* si proporcionan información derivada de análisis repetidos de una muestra simple en diferentes períodos de tiempo.

Las series estadísticas necesarias para la especificación y estimación de un modelo econométrico pueden obtenerse de forma directa, generando la información necesaria, bien mediante la utilización de registros propios o mediciones directas al respecto, o a través de la utilización de estadísticas generadas por otros organismos (Banco de España, Instituto Nacional de Estadística, etc.).

2.º *Examen de las condiciones de identificabilidad de la función objeto de análisis*. La *identificación* de una función permitirá asegurar, mediante el establecimiento de ciertas reglas, que los coeficientes estimados corresponden verdaderamente a la función objeto de estudio.

3.º *Examen de los problemas de agregación entre las variables*. La utilización de variables agregadas, por individuos, mercancías, precios o unidades temporales, en las funciones objeto de estudio, dará lugar a la presencia de problemas de agregación.

4.º *Examen del grado de correlación entre las variables explicativas*. Normalmente, entre las variables económicas existe un cierto grado de asociación lineal, cierta dosis de multicolinealidad en la especificación del modelo. La representación gráfica constituye una opción eficaz, pero nunca definitiva, de cara a obtener un conocimiento intuitivo con relación al comportamiento entre dos variables. Las técnicas utilizadas para poder cuantificar dicha relación, asociación lineal o dependencia, se corresponden con el análisis de correlación y regresión, respectivamente.

5.º *Elección de la técnica econométrica* más apropiada para la estimación de la función, junto con el examen crítico de sus hipótesis e implicaciones económicas. La elección dependerá, entre otras cuestiones, del tipo de modelo utilizado: uniecuacional, constituido por una sola ecuación, o de ecuaciones simultáneas si son varias las ecuaciones que lo integran.

4. EVALUACIÓN DE LAS ESTIMACIONES

Una vez estimado el modelo, el económetra deberá proceder a la *evaluación,* o determinación de la fiabilidad, de los resultados obtenidos, utilizando básicamente criterios de naturaleza económica, estadística y econométrica (figura 4.1).

© Ediciones Pirámide

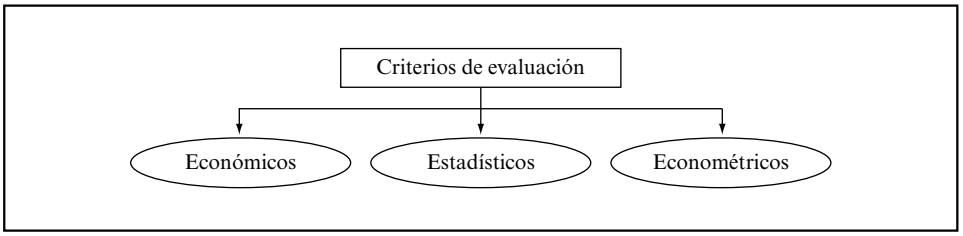


Figura 4.1. Evaluación de las estimaciones.

Los *criterios económicos,* determinados por los principios de la Teoría Económica, se referirán, fundamentalmente, al signo y magnitud de los parámetros de las relaciones económicas. Determinados por la Teoría Estadística, los *criterios estadísticos* se dirigirán a la evaluación de la fiabilidad de las estimaciones de los parámetros del modelo.

Por último, los *criterios de naturaleza econométrica* tendrán como objetivo prioritario la contrastación del cumplimiento de los supuestos en los que se apoyan las técnicas econométricas empleadas.

5. EVALUACIÓN DEL PODER PREDICTIVO DEL MODELO ESTIMADO

El objetivo de cualquier investigación econométrica será la obtención de estimaciones numéricas de los coeficientes de las relaciones económicas, para su utilización en la predicción de los valores de la variable dependiente.

Antes de utilizar la estimación de un modelo para predecir el valor de la variable dependiente, deberá evaluarse, de una forma u otra, su poder predictivo. Es posible que el modelo sea significativo desde un punto de vista económico, estadístico y econométrico, en el período muestral para el que ha sido estimado. No obstante, también puede ocurrir que dicho modelo sea poco fiable para la predicción, debido, por ejemplo, a un comportamiento estructuralmente inestable.

La bondad de un modelo econométrico se juzgará, habitualmente, por el cumplimiento de las denominadas *propiedades deseables,* como son la plausibilidad teórica, capacidad predictiva, fiabilidad de las estimaciones de los parámetros poblacionales y simplicidad.

El modelo deberá ser compatible con los postulados de la Teoría Económica y describir adecuadamente los fenómenos económicos con los que se relaciona. Deberá ser consistente con el comportamiento observado de las variables económicas cuya relación determina, representando las relaciones económicas de la forma más sencilla posible.

© Ediciones Pirámide

2 El modelo de regresión lineal simple

1. INTRODUCCIÓN

Como ya hemos señalado en el capítulo anterior, la Econometría se ocupa de la medición y verificación empírica de las relaciones económicas que, dentro del campo de la Teoría Económica, recogen los denominados «modelos económicos».

Generalmente, la Teoría Económica utiliza relaciones de carácter determinista, esto es, relaciones funcionales de la forma

$$Y = f(X)$$

donde a cada valor de la variable independiente, *X,* le corresponda un único valor de la variable a explicar o variable dependiente, *Y.* Si dichas relaciones fuesen en realidad deterministas, no se plantearía el problema de la contrastación estadística, siendo suficiente la realización de mediciones precisas para obtener las estimaciones de los parámetros desconocidos.

Sin embargo, la Economía es una ciencia de carácter empírico-social, en la que el azar está presente de forma relevante. Por ello, los economistas insistimos reiteradamente en la necesidad de contrastar las distintas teorías económicas.

2. EL MODELO DE REGRESIÓN LINEAL SIMPLE

La Econometría se ocupa exclusivamente del análisis y estudio de las relaciones estocásticas que se producen entre las variables económicas, siendo el *modelo de regresión lineal simple:*

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \forall i = 1, 2, \ldots, n \tag{2.1}$$

© Ediciones Pirámide

donde X e Y denotan las variables independiente y dependiente, respectivamente, β_0 y β_1 los coeficientes de regresión, u el término de perturbación aleatoria, n el número de observaciones, e i el orden de la observación, la relación más sencilla.

Del carácter estocástico del segundo miembro de (2.1) se deriva la aleatoriedad de la variable dependiente del modelo.

Los valores de las variables X e Y son observables, pero no así los relativos a la variable aleatoria u. Las observaciones tanto de X como de Y pueden efectuarse bien a lo largo del tiempo, en cuyo caso diremos que la información se presenta en forma de serie temporal, o para distintos conjuntos de individuos, objetos o unidades territoriales, esto es, en forma de corte transversal.

Normalmente las relaciones agregadas, tales como funciones de consumo o demanda de mercado, se estiman a partir de datos en forma de serie temporal, mientras que para las microrrelaciones, funciones de gasto de las economías domésticas, gastos de producción de las empresas, etc., se utilizan datos en forma de corte transversal.

El origen de las observaciones de las variables X e Y no se tiene en cuenta a la hora de proceder a la estimación de los parámetros β_0 y β_1, si bien influirá, como veremos más adelante, en las propiedades de los estimadores de dichos parámetros.

Es importante señalar que, en la disciplina que nos ocupa, el análisis de regresión constituye la técnica de análisis fundamental, frente al análisis de correlación. La distinción entre ambas es básica. En esta última, el objetivo prioritario se concreta en la medición del grado de asociación lineal entre dos variables, no existiendo en ningún momento distinción entre variable dependiente y explicativa. Sin embargo, el análisis de regresión parte de la existencia de una relación causal entre dos variables, dependiente y explicativa, siendo la primera aleatoria y la segunda fija o no estocástica.

En términos generales, el *análisis de regresión* se define como la técnica que se ocupa de analizar la dependencia entre una variable dependiente y una o más variables explicativas. Su objetivo consiste en estimar y/o predecir el valor medio poblacional de la variable dependiente a partir de valores conocidos y fijos de las variables explicativas, obtenidos mediante un proceso de muestreo repetido.

Cuando la especificación o formulación concreta de la relación entre la variable dependiente y la variable o variables independientes sea lineal en los parámetros, como por ejemplo

$$Y_i = \beta_0 + \beta_1 X_i^2 + u_i \qquad \forall i = 1, 2, \ldots, n$$

© Ediciones Pirámide

hablaremos de *análisis de regresión lineal*[1]. Obsérvese que una relación de la forma

$$Y_i = \beta_0 + \sqrt{\beta_1} X_i + u_i \qquad \forall i = 1, 2, \ldots, n$$

no cumpliría esta condición.

Señalaremos, por último, que el calificativo *simple* en (2.1) se refiere al hecho de considerar una única variable explicativa, frente al *modelo de regresión lineal múltiple*

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i \qquad \forall i = 1, 2, \ldots, n \tag{2.2}$$

que incorpora cualquier número de variables explicativas en su especificación.

Con objeto de simplificar la exposición, dejaremos para un capítulo posterior el estudio del modelo de regresión lineal múltiple, centrándonos ahora en el modelo de regresión lineal simple.

Establecido el significado del término *análisis de regresión lineal simple,* profundizaremos un poco más en la naturaleza aleatoria de dicho análisis. Para ello, introduciremos dos conceptos básicos en la especificación de un modelo econométrico: *función de regresión poblacional* (FRP) y *función de regresión muestral* (FRM).

Supongamos que queremos analizar el comportamiento de las pautas de consumo, gasto en consumo, de una unidad territorial integrada por 100 economías domésticas, en función de la renta familiar disponible y de los ingresos salariales mensuales, medidas ambas variables en euros. Si existiese una relación funcional exacta entre el gasto en consumo e ingreso salarial, para cada nivel de renta disponible se produciría un nivel de gasto determinado, esto es,

$$Y_i = f(X_i) \qquad \forall i = 1, 2, \ldots, n \tag{2.3}$$

siendo f la relación funcional que ligaría el comportamiento entre ambos factores.

Si la relación funcional recogida en (2.3) fuese lineal,

$$Y_i = \beta_0 + \beta_1 X_i \qquad \forall i = 1, 2, \ldots, n \tag{2.4}$$

para el conjunto de observaciones correspondientes a la población considerada, la representación gráfica de (2.4) se ajustaría a la figura 2.1. Sin embargo, las unidades familiares presentan comportamientos heterogéneos porque su estructura de preferencias lo es. Diferencias en el tamaño, composición, estructura de edades de sus miembros, estatus social, etc., dan lugar a comportamientos diferenciados en la realización del gasto. Ello significa que unidades

[1] Nótese que el ejemplo propuesto no satisface el supuesto de linealidad con respecto a las variables.

© Ediciones Pirámide

familiares con el mismo nivel de ingresos podrían materializar niveles de gasto diferentes. Para la población considerada, la representación gráfica de las observaciones relativas al ingreso y consumo de las unidades familiares correspondería al diagrama de dispersión recogido en la figura 2.2, en el que puede observarse cómo unidades familiares con el mismo nivel de ingresos materializan pautas diferentes.

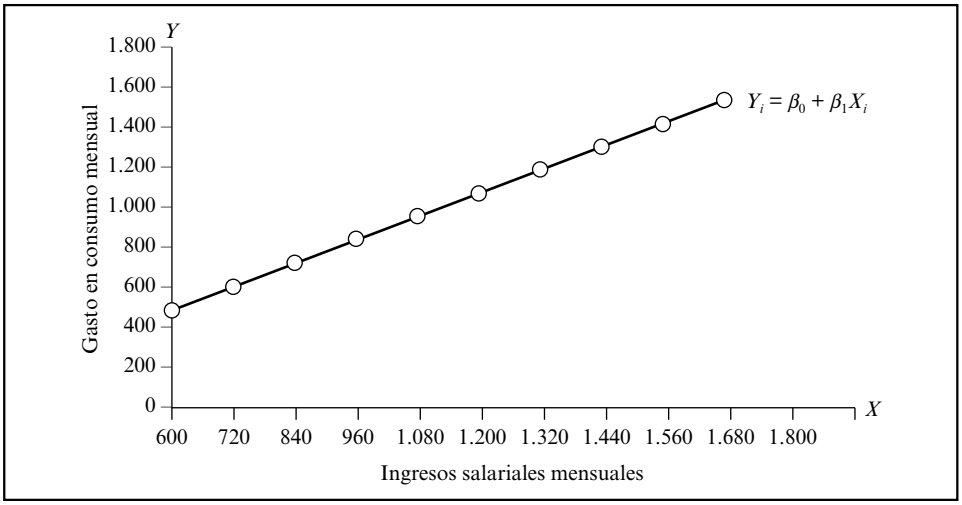


Figura 2.1. Relación determinista entre las variables gasto en consumo y renta familiar disponible.

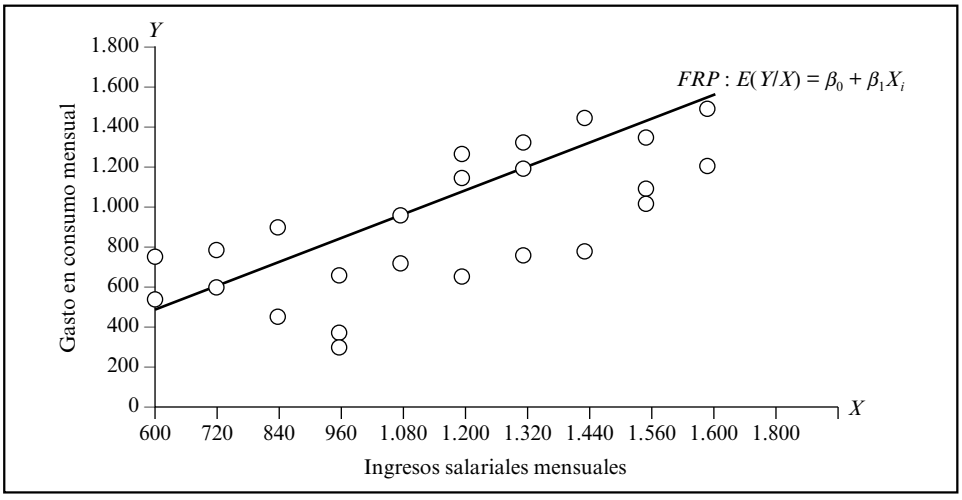


Figura 2.2. Función de regresión poblacional. Relación hipotética entre las variables gasto en consumo y renta familiar disponible.

© Ediciones Pirámide

Esta consideración hace necesaria la modificación de la función de consumo determinista recogida en (2.3), incorporando en su especificación una variable aleatoria,

$$Y_i = f(X_i, u_i) \qquad \forall i = 1, 2, \ldots, n$$

esto es,

$$Y_i = E\left(Y/X = X_i\right) + u_i \qquad \forall i = 1, 2, \ldots, n \tag{2.5}$$

expresión que ofrece la ventaja de mostrar claramente cómo, además de los ingresos salariales, otras variables influyen en la estructura de las pautas de consumo de las unidades familiares consideradas.

Obsérvese que (2.5) expresa la pauta de consumo de la *i*-ésima economía doméstica para un nivel de ingreso igual a X_i, como la suma del gasto promedio de todas las unidades familiares que perciben la misma renta, más una cierta cantidad, positiva o negativa, aleatoria, u_i, esto es,

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad \forall i = 1, 2, \ldots, n \tag{2.6}$$

teniendo en cuenta (2.4).

La FRP, denominada también *línea de regresión poblacional,* se obtiene al ajustar, lo mejor posible, los puntos del diagrama de dispersión recogido en la figura 2.2,

$$E\left(Y/X = X_i\right) = \beta_0 + \beta_1 X_i \qquad \forall i = 1, 2, \ldots, n \tag{2.7}$$

que muestra cómo la media poblacional de la distribución de la variable Y, para valores fijos de la variable X, $E\left(Y/X = X_i\right)$, se relaciona linealmente con la variable explicativa X.

En nuestro ejemplo nos hemos referido a una población integrada por 100 economías domésticas, y en la que, por tanto, es factible un conocimiento exacto de la FRP. Sin embargo, dicha función constituye una construcción idealizada y teórica. En la práctica dispondremos, generalmente, de información muestral relativa a la variable dependiente Y, para valores fijos de la variable o variables explicativas.

Siguiendo con el ejemplo anterior, supongamos que disponemos de una muestra aleatoria de las variables consideradas correspondiente a 10 economías domésticas. De la representación gráfica de las observaciones de la muestra seleccionada, y tratando de ajustar lo mejor posible la nube de puntos,

© Ediciones Pirámide

obtendremos la denominada *línea de regresión muestral,* recogida en la figura 2.3.

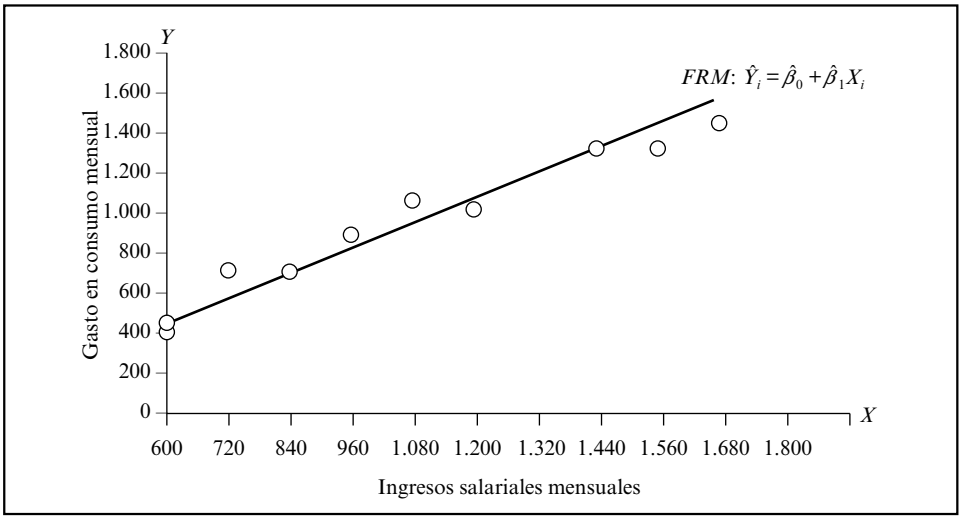


Figura 2.3. Función de regresión muestral. Relación hipotética entre las variables gasto en consumo y renta familiar disponible.

El concepto de FRM que representa la *línea de regresión muestral,* esto es, la contrapartida muestral de (2.7), expresada como

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad \forall i = 1, 2, \ldots, n \tag{2.8}$$

donde $\hat{Y}_i$ representa un estimador de $E\left(Y/X = X_i\right)$ y $\hat{\beta}_0$ y $\hat{\beta}_1$, de los parámetros poblacionales β_0 y β_1, respectivamente, se desarrolla de forma análoga al concepto de FRP.

Alternativamente, y teniendo en cuenta (2.5) y (2.7), los valores de la variable dependiente se pueden expresar como

$$Y_i = \hat{Y}_i + e_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i \qquad \forall i = 1, 2, \ldots, n \tag{2.9}$$

siendo e_i el término residual, esto es,

$$e_i = Y_i - \hat{Y}_i \qquad \forall i = 1, 2, \ldots, n \tag{2.10}$$

conceptualmente análogo a u_i y que puede ser considerado como una estimación del desconocido término de perturbación aleatoria ($e_i = \hat{u}_i$).

© Ediciones Pirámide

3. SUPUESTOS DEL MODELO DE REGRESIÓN

La naturaleza estocástica de (2.1) pone de manifiesto la existencia de una distribución de probabilidad para la variable Y, determinada por los valores de X y la distribución probabilística de u.

La especificación completa del modelo de regresión exigirá, por tanto, la determinación de la forma concreta de (2.5), la distribución de probabilidad y características del término aleatorio, así como el origen de los valores de la variable explicativa.

Esta información adicional se recoge en lo que se conoce como los *supuestos básicos del modelo de regresión.* Pasaremos a analizar seguidamente dichos aspectos, diferenciando entre *supuestos estocásticos,* que hacen referencia a la componente aleatoria de (2.1), y *no estocásticos,* relativos a la componente determinista del mismo.

3.1. Supuestos estocásticos

1.º El valor esperado del término de perturbación aleatoria es cero,

$$E(u_i) = 0 \qquad \forall i = 1, 2, \ldots, n \tag{3.1.1}$$

lo que significa que para cada conjunto de observaciones de la variable dependiente del modelo, correspondiente a un determinado valor de X, el valor promedio de la variable aleatoria es cero.

2.º La variable aleatoria del modelo se distribuye de forma serialmente independiente, esto es, no está autocorrelada,

$$\text{Cov}\,(u_i, u_j) = 0 \qquad \forall i \neq j \tag{3.1.2}$$

lo cual significa que dicha variable no presenta un patrón sistemático de comportamiento. Los valores de la variable aleatoria u para las distintas observaciones son independientes entre sí.

El análisis de la evolución temporal de cualquier macromagnitud constituye un ejemplo claro de la restrictividad de este supuesto.

3.º La varianza del término de perturbación aleatoria es constante u homoscedástica,

$$\text{Var}\,(u_i) = \sigma_u^2 \qquad \forall i = 1, 2, \ldots, n \tag{3.1.3}$$

esto es, las desviaciones de dicha variable con respecto a su valor medio elevadas al cuadrado son las mismas para cualquier observación de la variable explicativa.

© Ediciones Pirámide

Si consideráramos este supuesto en el estudio de las pautas de consumo de las unidades familiares pertenecientes a una determinada unidad territorial, nos encontraríamos ante una situación poco plausible: todas las unidades familiares responderían a una misma pauta de consumo. Ello significaría la negación del carácter aleatorio de la conducta humana.

4.º Las variables aleatoria y explicativa de (2.1) se distribuyen de forma estadísticamente independiente,

$$\text{Cov}(u_i, X_i) = 0 \qquad \forall i = 1, 2, \ldots, n \tag{3.1.4}$$

es decir, presentan efectos separados y de carácter aditivo sobre la variable dependiente, Y.

3.2. Supuestos no estocásticos

1.º La variable explicativa de (2.1) es fija en el muestreo. El carácter no aleatorio de dicha variable constituye un supuesto especialmente fuerte para las ciencias de ámbito social, en las que los datos se obtienen mediante procesos de observación y no de experimentación.

2.º El modelo está correctamente especificado. Si se omiten del modelo variables relevantes, se elige una relación funcional incorrecta o se plantean supuestos erróneos en relación a la estructura de sus componentes, la validez de los resultados sería muy cuestionable.

Obsérvese que los supuestos analizados se refieren a la FRP, si bien, como comprobaremos más adelante, la aplicación del método de estimación mínimo-cuadrático, con objeto de obtener la FRM, conduce a resultados similares en relación al término residual.

Los rasgos que definen la variable dependiente de (2.1) se derivan directamente de los supuestos establecidos.

1.º El valor esperado de la variable Y es igual a

$$E(Y_i) = \beta_0 + \beta_1 X_i \qquad \forall i = 1, 2, \ldots, n \tag{3.2.1}$$

de acuerdo con (2.1) y (3.1.1). En efecto, cada conjunto de observaciones de la variable dependiente Y correspondiente a valores fijos de la variable explicativa X

$$\begin{aligned} E(Y_i) &= E(\beta_0 + \beta_1 X_i + u_i) = E(\beta_0) + E(\beta_1 X_i) + E(u_i) \\ &= \beta_0 + \beta_1 X_i + 0 = \beta_0 + \beta_1 X_i \end{aligned}$$

se distribuye alrededor de su valor medio.

© Ediciones Pirámide

2.º La variable dependiente de (2.1) se distribuye de forma serialmente independiente,

$$\text{Cov}\,(Y_i, Y_j) = 0 \qquad \forall i \neq j \tag{3.2.2}$$

es decir, no está autocorrelada.

A partir de la definición estadística de covarianza entre dos variables,

$$\begin{aligned}\text{Cov}\,(Y_i, Y_j) &= E\{[Y_i - E(Y_i)][Y_j - E(Y_j)]\} \\ &= E\{[Y_i - (\beta_0 + \beta_1 X_i)][Y_j - (\beta_0 + \beta_1 X_j)]\} \\ &= E(u_i\, u_j) = 0\end{aligned}$$

teniendo en cuenta (3.1.2).

3.º La varianza de la variable dependiente de (2.1) es constante e igual a la del término aleatorio del mismo,

$$\text{Var}\,(Y_i) = \sigma_u^2 \qquad \forall i = 1, 2, \ldots, n \tag{3.2.3}$$

dado que

$$\text{Var}\,(Y_i) = E[Y_i - E(Y_i)]^2 = E[Y_i - (\beta_0 + \beta_1 X_i)]^2 = E(u_i)^2 = \sigma_u^2$$

las desviaciones de la variable dependiente, Y, con respecto a su valor medio elevadas al cuadrado, son constantes para cualquier valor de la variable explicativa.

4. OBTENCIÓN DE LOS ESTIMADORES MÍNIMO-CUADRÁTICOS

Dado que (2.1) no es directamente observable, y teniendo en cuenta (2.9), trataremos de estimar la FRM que mejor se adapte a la información muestral disponible,

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + e_i = \hat{Y}_i + e_i \tag{4.1}$$

siendo $\hat{Y}_i$ el valor estimado de $E\left(Y/X = X_i\right)$ y e_i el término residual.

Con este objeto, mencionaremos algunos de los criterios que se podrían adoptar.

© Ediciones Pirámide

1.º *Minimización de la suma de residuos,* $\text{Min} \sum_{i=1}^{n} e_i$. Aunque en principio se trata de un criterio atractivo, esta técnica de estimación presenta algunos problemas. Los residuos obtenidos por (2.10) pueden tomar signo positivo o negativo, pudiendo su suma algebraica alcanzar un valor pequeño en términos absolutos, aunque se distribuyera de forma dispersa alrededor de la FRM elegida.

2.º *Minimización de la suma de los residuos en valor absoluto,* $\text{Min} \sum_{i=1}^{n} |e_i|$. Con este criterio se evitaría el problema planteado anteriormente, si bien se trata de una técnica que conduce a cálculos complicados para obtener las estimaciones de los parámetros β_0 y β_1.

3.º *Minimización de la suma de residuos al cuadrado,* $\text{Min} \sum_{i=1}^{n} e_i^2$. Este criterio evita el problema de las compensaciones con respecto al signo de los residuos, siendo, además, las propiedades de los estimadores obtenidos de su aplicación muy deseables desde un punto de vista estadístico. La elección de este criterio con objeto de obtener la FRM se denomina método de estimación de *Mínimos Cuadrados Ordinarios* (MCO), o regresión mínimo-cuadrática.

La aplicación del criterio mínimo-cuadrático a (2.1) se concreta en un problema de optimización matemático,

$$\text{Min} \sum_{i=1}^{n} e_i^2 = \text{Min} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \text{Min} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \qquad (4.2)$$

en el que la función objetivo no está sometida a ninguna restricción.

Obsérvese que (4.2) denota un problema de optimización libre, a resolver mediante las condiciones necesaria y suficiente que la teoría establece al respecto.

Dado que tratamos de ajustar la FRM que más se aproxima a la FRP, en base a la muestra considerada, elegiremos $\hat{\beta}_0$ y $\hat{\beta}_1$ de manera que (2.8) se aproxime lo mejor posible a (3.2.1). En consecuencia, las variables o incógnitas de (4.2) serán las estimaciones de los parámetros β_0 y β_1, esto es, $\hat{\beta}_0$ y $\hat{\beta}_1$. Las estimaciones numéricas se obtendrán de la resolución del sistema de ecuaciones

$$\begin{cases} \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = 0 \\ \dfrac{\partial \sum_{i=1}^{n} e_i^2}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) X_i = 0 \end{cases}$$

© Ediciones Pirámide

esto es,

$$\begin{cases} \sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \end{cases} \tag{4.3}$$

denominado *sistema de ecuaciones normales,* derivado de la aplicación de la condición necesaria de optimización[2].

Obsérvese que al dividir la primera ecuación de (4.3) por el número de observaciones, *n,*

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X} \tag{4.4}$$

se obtiene la FRM particularizada en el punto de coordenadas correspondientes a los valores medios muestrales de las variables X e Y, respectivamente. En consecuencia, al minimizar la suma de residuos mínimo-cuadráticos al cuadrado se elegirán $\hat{\beta}_0$ y $\hat{\beta}_1$, de tal manera que la FRM pase por el punto $(\bar{X}, \bar{Y})$.

El programa matemático recogido en (4.2) se puede expresar alternativamente, considerando las variables del modelo en desviaciones con respecto a la media. En efecto, bajo dicha consideración, tal y como se refleja en la figura 4.1, podremos expresar los residuos como

$$e_i = Y_i - \hat{Y}_i = (Y_i - \bar{Y}) - (\hat{Y}_i - \bar{Y}) = y_i - \hat{y}_i \qquad \forall i = 1, 2, \ldots, n \tag{4.5}$$

por lo que (4.2) se transformaría en

$$\text{Min} \sum_{i=1}^{n} e_i^2 = \text{Min} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.6}$$

siendo

$$\hat{y}_i = \hat{Y}_i - \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} = \hat{\beta}_1 (X_i - \bar{X}) = \hat{\beta}_1 x_i \qquad \forall i = 1, 2, \ldots, n$$

[2] Matemáticamente, el valor crítico obtenido en la condición necesaria se caracteriza en la condición suficiente estudiando el signo de la segunda derivada. En el contexto que nos ocupa, dado que la variable a optimizar es el residuo mínimo-cuadrático, no tendría sentido que el valor crítico resultante se caracterizara como un máximo.

© Ediciones Pirámide

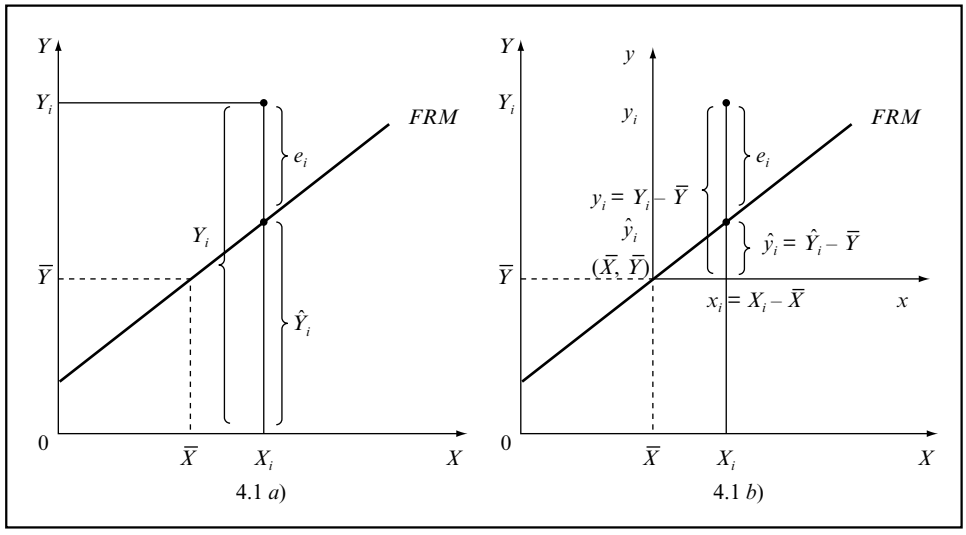


Figura 4.1. Ajuste mínimo-cuadrático. Variables en desviaciones respecto de la media.

y, consecuentemente (4.6), adoptaría

$$\text{Min}\sum_{i=1}^{n} e_i^2 = \text{Min}\sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i)^2 \tag{4.7}$$

una expresión alternativa.

De la aplicación de la condición necesaria de optimización a (4.7) se obtiene:

$$\frac{d\sum_{i=1}^{n} e_i^2}{d\hat{\beta}_1} = -2\sum_{i=1}^{n} (y_i - \hat{\beta}_1 x_i) x_i = 0$$

por lo que:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{S_{XY}}{S_X^2} \tag{4.8}$$

donde S_X^2 denota la varianza muestral de la variable X, y S_{XY} la covarianza de la misma naturaleza entre las variables X e Y. En este caso, la aplicación

© Ediciones Pirámide