

# Manual de Econometría

Alfonso Novales

Abril 29, 2003

## Contents

1	Introducción . . . . .	5
2	Algunos conceptos estadísticos básicos . . . . .	5
2.1	Medidas de posición y medidas de dispersión . . . . .	5
2.1.1	La media muestral como predictor . . . . .	5
2.1.2	La desviación típica como indicador de volatilidad . . . . .	7
2.2	Medidas de asociación . . . . .	9
3	Contrastación de hipótesis estadísticas . . . . .	9
3.1	Contrastes de Normalidad . . . . .	14
3.2	Contrates de asociación . . . . .	14
4	Tratamiento de datos . . . . .	15
4.1	Ajuste estacional . . . . .	15
4.2	Tasas de variación . . . . .	15
4.3	Estimación de componentes . . . . .	15
5	El modelo lineal simple de regresión . . . . .	15
5.1	Descripción del modelo . . . . .	15
5.1.1	Nube de puntos . . . . .	15
5.2	Estimación por mínimos cuadrados (ordinarios) . . . . .	15
5.2.1	Representación gráfica de la recta de regresión estimada . . . . .	15
5.3	Propiedades del estimador de mínimos cuadrados ordinarios . . . . .	15
5.4	Residuos del modelo. Gráficos de residuos . . . . .	16
5.4.1	Estimación de la varianza del término de perturbación . . . . .	17
5.5	Cuando los coeficientes del modelo de regresión cambian a lo largo de la muestra . . . . .	17
5.5.1	Cambio estructural en los coeficientes del modelo de regresión . . . . .	18

5.5.2	Variación gradual en los coeficientes del modelo . . . . .	21
5.6	Algunos modelos de regresión sencillos . . . . .	21
5.6.1	El modelo constante . . . . .	21
5.6.2	El modelo con variables en desviaciones respecto a la media	23
5.6.3	El modelo con tendencia determinista lineal y cuadrática .	24
5.6.4	Modelos no lineales en las variables . . . . .	25
5.7	¿Cómo especificar un modelo de regresión? . . . . .	26
5.7.1	¿Debe incluirse una constante en el modelo de regresión? .	26
5.7.2	¿Debemos estimar en valores originales o en logaritmos de las variables? . . . . .	27
5.7.3	¿Debe estimarse el modelo con variables en niveles o en diferencias? . . . . .	27
5.7.4	La frecuencia de observación de las variables . . . . .	27
6	Contrastes de hipótesis en el modelo de regresión lineal simple . . . . .	30
6.1	Significación estadística versus precisión . . . . .	30
6.1.1	¿Se hace una variable más o menos significativa? . . . . .	32
6.1.2	¿Cómo puede discutirse qué variable es más relevante en una regresión? . . . . .	32
7	Correlación versus causalidad . . . . .	32
8	Variables no estacionarias . . . . .	32
8.1	Características de una variable estacionaria . . . . .	32
8.2	Tendencias deterministas y tendencias estocásticas . . . . .	33
8.3	Regresión espúrea . . . . .	36
8.3.1	Regresión espúrea bajo tendencias deterministas . . . . .	38
8.3.2	Regresión espúrea bajo tendencias estocásticas . . . . .	39
8.4	Tratamiento de tendencias deterministas . . . . .	42
8.5	Ejercicios de simulación . . . . .	44
8.6	Tendencias estocásticas y raíces unitarias . . . . .	44
8.7	Contrastes de raíz unitaria . . . . .	45
8.8	Cointegración . . . . .	45
8.8.1	Contraste de cointegración . . . . .	45
8.8.2	Contraste de hipótesis sobre la relación de cointegración estimada por mínimos cuadrados . . . . .	47
8.8.3	Correlación y cointegración . . . . .	47
8.8.4	Variables cointegradas: un ejemplo . . . . .	48
8.8.5	El modelo de corrección de error . . . . .	49
8.8.6	El contraste de cointegración de Johansen . . . . .	51

8.8.7	Aspectos comunes a varias variables temporales: tendencias comunes, volatilidad común. . . . .	51
8.8.8	¿Qué hacer en presencia de variables con tendencias estocásticas (raíces unitarias)? . . . . .	51
9	Matrices de covarianzas no escalares . . . . .	54
9.1	Detección de la autocorrelación . . . . .	54
9.2	Tratamiento de la autocorrelación. . . . .	54
9.3	El estimador de mínimos cuadrados generalizados . . . . .	54
9.4	Detección de la heteroscedasticidad . . . . .	54
9.5	Contraste de igualdad de varianza entre submuestras . . . . .	54
9.6	Tratamiento de la heteroscedasticidad . . . . .	54
10	El modelo de regresión lineal múltiple . . . . .	54
10.1	Estimación por mínimos cuadrados . . . . .	59
11	Propiedades del estimador de mínimos cuadrados. . . . .	60
12	Bondad de ajuste del modelo . . . . .	68
13	Contrastes de hipótesis . . . . .	68
14	Matrices de covarianzas no escalares . . . . .	74
14.1	Comparación de estimadores de la regresión múltiple y la regresión simple . . . . .	75
14.2	Regresión particionada . . . . .	77
15	Grado de ajuste del modelo de regresión múltiple . . . . .	77
15.1	Coefficientes de correlación parcial y de determinación parcial . . . . .	78
16	Colinealidad entre variables explicativas en un modelo de regresión . . . . .	80
16.1	Efectos de la colinealidad entre variables explicativas . . . . .	81
16.2	Detección de la colinealidad . . . . .	84
16.3	Tratamiento de la colinealidad . . . . .	85
16.3.1	Regresión ortogonalizada . . . . .	85
16.3.2	Otros tratamientos . . . . .	87
17	Predicción . . . . .	87
18	Modelos univariantes de series temporales . . . . .	88
18.1	Primeros conceptos . . . . .	88
18.1.1	Procesos estocásticos . . . . .	88
18.1.2	Funciones de autocorrelación simple y parcial . . . . .	88
18.2	Procesos autoregresivos, AR(p) . . . . .	88
18.2.1	El modelo AR(1) . . . . .	88
18.2.2	El modelo AR(2) . . . . .	88
18.3	Procesos de medias móviles, MA(q) . . . . .	88

18.4	Procesos mixtos, ARMA(p,q)	88
18.5	Procesos integrados ARIMA(p,d,q)	88
18.6	Predicción con modelos univariantes	88
18.6.1	Predicción con modelos AR(p)	88
18.6.2	Predicción con modelos MA(q)	88
18.6.3	Predicción con modelos ARMA(p,q)	88
18.6.4	Predicción con modelos ARIMA(p,d,q)	88
18.7	Estimación de modelos univariantes de series temporales	88
18.7.1	Estimación de modelos autoregresivos	88
18.7.2	Estimación de modelos de medias móviles	88
18.7.3	Estimación de modelos ARMA(p,q) y ARIMA(p,d,q)	88
19	El procedimiento de variables instrumentales	88
19.1	Correlación entre variables explicativas y término de error	88
19.2	Errores de medida	88
20	Modelos dinámicos	88
20.1	Colinealidad entre variables explicativas	88
20.2	Estimación	88
20.2.1	Perturbación sin autocorrelación	88
20.2.2	Perturbación con autocorrelación	88
21	Simultaneidad	88
21.1	Identificación	88
21.2	Estimación de una ecuación del sistema	88

# 1. Introducción

Ha sido tradicional dejar muchas decisiones a los métodos estadísticos.

La discusión importante es si el investigador debe plantear su investigación incorporando sus creencias a priori o, por el contrario, la investigación ha de ser aséptica en ese sentido, no debiendo estar condicionada en ningún aspecto por las creencias iniciales del investigador.

## 2. Algunos conceptos estadísticos básicos

### 2.1. Medidas de posición y medidas de dispersión

Media, media ponderada, mediana, moda, varianza, desviación típica. Medidas alternativas de volatilidad.

La media, muestral o poblacional, es la constante con respecto a la cual el error cuadrático medio de la variable aleatoria es menor. Conveniencia de su utilización.

#### 2.1.1. La media muestral como predictor

La esperanza matemática de una variable tiene una propiedad de gran importancia: es la constante alrededor de la cual la variable aleatoria experimenta fluctuaciones de menor tamaño. Análogamente, dada una determinada muestra, la media muestral es la constante con respecto a la cual la variable experimenta unas menores desviaciones. Es decir, si nos planteamos resolver el problema,

$$\underset{a}{\text{Min}} F(a) \equiv \underset{a}{\text{Min}} E (X - a)^2$$

donde la incógnita es la constante  $a$ , la solución es  $a = \mu$ . El valor mínimo de la función objetivo es:  $F(\mu) = \text{Var}(X)$ . Dada una determinada muestra de tamaño  $n$ , la solución al problema,

$$\underset{a}{\text{Min}} F(a) \equiv \sum_{i=1}^n (X_i - a)^2$$

viene dada por  $a = \bar{x}$ . El valor mínimo de la función objetivo es:  $F(\bar{x}) = \text{Var}(X)$ , varianza muestral de  $X$ .

Quizá sorprendentemente, esta propiedad tiene implicaciones en relación con la predicción: un ejercicio de predicción consiste en anticipar un valor de una

variable  $X_t$  en un instante futuro  $t_0$ ,  $\hat{x}_{t_0}$  a partir de observaciones temporales  $x_1, x_2, \dots, x_T$ ,  $T < t_0$ . Por supuesto que una predicción no se verá exactamente corroborada por los datos futuros: el dato que se reciba en el instante  $t_0$ ,  $x_{t_0}$ , podrá exceder de la previsión  $\hat{x}_{t_0}$  que efectuamos en el instante  $T$ , o ser inferior a la misma. Lo que el investigador quiere es que su método de predicción garantice que el error que espera cometer sea el menor posible. no tendría sentido utilizar un método de predicción que incumpla esta propiedad, pues equivaldría a creer que existe algún método de predicción alternativo con una expectativa de error inferior. Ahora bien, como el error de predicción (definido como la diferencia entre valor realizado y valor anticipado,  $x_{t_0} - \hat{x}_{t_0}$ ) que se materialice en  $t_0$  puede ser positivo o negativo, es razonable buscar un método de predicción que minimice la expresión,

$$\text{Min}_{\hat{x}_{t_0}} E (x_{t_0} - \hat{x}_{t_0})^2 \quad (2.1)$$

Ahora bien, si la variable  $X$  tiene una distribución de probabilidad constante en el tiempo, como ocurriría si estamos considerando una muestra aleatoria simple, y en ausencia de tendencias, cambios estructurales en media, etc., la propiedad anterior nos sugiere que el mejor procedimiento de predicción será,

$$\hat{x}_{t_0} = \bar{x}$$

Por tanto, si hemos de predecir un valor futuro de una variable aleatoria, sin disponer de datos de la misma, su esperanza matemática, si es conocida, minimiza el Error Cuadrático Medio entre todas las predicciones posibles. Si conocemos únicamente la media muestral, pero no disponemos de los datos individuales, dicha media muestral tiene una propiedad análoga a la que acabamos de enunciar.

Tan importante es esta propiedad que la media muestral debe utilizarse como referencia con respecto a la cual tratar de obtener una predicción mejor. Cuando decimos que una variable es impredecible, no nos referimos a que su predicción es cero, o que no puede obtenerse, sino a que propondremos como predicción su media muestral, sin hacer ningún cálculo adicional. Lo importante es observar que, si se conoce la esperanza matemática de la variable, o se dispone de una media muestral, éstas son predicciones sencillas de obtener, y aceptables, al menos en ausencia de información adicional.

Como matiz puede añadirse que, si el criterio a minimizar cuando se calcula una predicción no es el ECM como en (2.1), sino el Error Absoluto Medio de la predicción,

$$\underset{\hat{x}_{t_0}}{\text{Min}} E | x_{t_0} - \hat{x}_{t_0} | \quad (2.2)$$

entonces la predicción debe ser la Mediana poblacional, o la mediana muestral, si se dispone de dicha información,  $\hat{x}_{t_0} = \text{Mediana}(x)$ .

Hay que resaltar, sin embargo, que, con más información que simplemente la media muestral, el investigador puede aspirar a obtener una predicción mejor que la proporcionada por la media muestral; para ello, deberá sustituir el promedio muestral por la esperanza condicional  $E_T x_{t_0}$ . Si por ejemplo, el investigador cree que la variable que pretende predecir obedece una estructura AR(1), entonces la predicción que minimiza el Error Cuadrático Condicional Medio vendrá dada por  $E_T x_{t_0} = \rho^{(t_0-T)} x_T$ , como vimos al examinar este tipo de procesos. La varianza condicional no es nunca superior a la varianza incondicional y es, en la mayoría de los casos, muy inferior. La media muestral minimiza la varianza incondicional, mientras que  $E_T x_{t_0}$  minimiza la varianza condicional, alcanzando un resultado menor de este criterio y, por tanto, preferible.

El modo en que puede utilizarse la información muestral detallada disponible para obtener el valor numérico de la varianza condicional  $E_T x_{t_0}$  es el objeto de XXX.

### 2.1.2. La desviación típica como indicador de volatilidad

Ha sido tradicional en el análisis de datos económicos utilizar la desviación típica como medida de volatilidad de una variable. Esto es especialmente cierto en el análisis de datos financieros, donde, aunque recientemente se han introducido otras medidas de volatilidad, el uso de la desviación típica es todavía habitual. Esta práctica se deriva de la interpretación directa de la desviación típica, como la desviación promedio entre los valores que toma una determinada variable aleatoria, y su valor medio. Sin embargo, existen múltiples situaciones en las que tal caracterización de la volatilidad puede proporcionar una imagen engañosa de lo que el investigador pretende medir.

Tomamos como punto de partida la idea de que, al medir volatilidad, el investigador pretende cuantificar el tamaño medio de las fluctuaciones que experimenta una determinada variable aleatoria. La simple lectura de esta afirmación debería sugerir al lector que, como posible definición de volatilidad, resulta fundamentalmente incompleta. Las fluctuaciones que experimenta una variable aleatoria no pueden estudiarse si no se define previamente el valor que sirve de referencia respecto al cual medir dichas fluctuaciones.

En una primera lectura, podría pensarse que es evidente que la pretensión es la de cuantificar las fluctuaciones que experimenta una variable aleatoria respecto a su nivel medio. Esta es la idea que subyace al uso de la desviación típica como medida de volatilidad; sin embargo, es fácil ver que existen situaciones en que dicha utilización no está totalmente justificada:

- Cambio estructural en media: supongamos una variable  $X$  que es constante a lo largo de cada una de las dos submuestras en que podemos dividir el período muestral. Es decir,  $X = \mu_1$  en la primera parte de la muestra, y  $X = \mu_2$  en la segunda parte de la muestra. En muchos sentidos, podríamos decir que esta variable es constante, y ha experimentado una volatilidad nula a lo largo del intervalo muestral, si bien es verdad que en el instante  $t_0$  se produjo un cambio estructural, de carácter permanente, en el nivel medio de la variable. Si no se tiene en cuenta dicho cambio en la media, la varianza de  $X$  resulta ser no nula, mientras que si tenemos en cuenta el cambio en media, la varianza que calculemos será cero.
- Presencia de una tendencia determinista: supongamos una variable  $X$  que crece a una tasa media de  $\gamma$ , a la vez que experimenta fluctuaciones alrededor de dicha tasa media de crecimiento. En este caso, la varianza muestral de la variables, así como su desviación típica, serán importantes, y tanto más elevadas cuanto mayor sea la tendencia o tasa de crecimiento  $\gamma$ . Esta situación es muy frecuente en Economía en general y en Finanzas en particular, y se utiliza la desviación típica como indicador de volatilidad. Hay dos dificultades que suelen ignorarse: una, que, en presencia de tendencia, carecemos de valor central de referencia. En presencia de una tendencia o crecimiento constante  $\gamma$ , el valor medio tenderá a ser el valor que tomó la variable hacia el período central de la muestra, pero no es representativo de los valores muestrales de la variable: la primera parte de la muestra tenderá a estar por debajo de la media, estando la segunda parte de la muestra por encima de la media muestral; en tal situación la media muestral no es un valor representativo de la variable y, en consecuencia, no tiene mucho sentido calcular el tamaño de las fluctuaciones alrededor de dicha media. En este contexto, el tamaño medio de las fluctuaciones alrededor de la media será, en realidad, un indicativo de la magnitud de  $\gamma$ , la tasa media de crecimiento o tendencia determinista.

Este es un caso donde debemos distinguir entre corto y largo plazo: a largo plazo, el uso de la desviación típica podría estar justificado, si entendemos que la



tendencia media de crecimiento  $\gamma$  será eventualmente sustituida por un descenso en la variable, y la sucesión de un ciclo u oscilación de período amplio, que terminaría justificando el cálculo de un valor central de referencia. La volatilidad a corto plazo será el tamaño medio de las fluctuaciones alrededor de la tendencia determinista. Por tanto, tendría plena justificación extraer de la variable el crecimiento tendencial, que había que estimar previamente, y calcular el tamaño medio de las fluctuaciones que experimente el componente que resulta tras la extracción de la tendencia determinista.

Una vez más, no se trata de discutir cuál es el modo correcto de proceder; más bien, hay que entender que estamos hablando de estimar características diferentes y, alguna de ellas puede no estar estadísticamente justificada. De acuerdo con la discusión anterior, medir la volatilidad mediante la desviación típica de una variable con tendencia determinista, tiene poca justificación. Esto es especialmente cierto en el análisis financiero: en un período en que el precio de un activo experimente un sólido crecimiento, su avrianza muestral resultará relativamente alta, por lo que podría concluirse que ofrece un riesgo importante; en consecuencia, a pesar de su sostenida rentabilidad (variación porcentual en precio), un inversor podría decidir no incorporarlo en su cartera; este análisis sería incorrecto.

## 2.2. Medidas de asociación

Coeficiente de correlación. Coeficiente de correlación parcial.

## 3. Contrastación de hipótesis estadísticas

Uno de los procedimientos básicos de la inferencia estadística es la contrastación de hipótesis, mediante el cual el investigador pretende conocer el grado en que la información muestral de que dispone es compatible con una determinada hipótesis. La hipótesis que se contrasta, denominada hipótesis nula, hace referencia a una determinada característica de la distribución de probabilidad de la que procede la información muestral disponible. Así, la hipótesis nula, denotada como  $H_0$ , puede ser del tipo:  $H_0$  : "La población de la que se extrajo la muestra es Normal" o referirse a valores numéricos para algún parámetro de dicha distribución de probabilidad poblacional, como  $H_0$  : "La esperanza matemática poblacional es igual a 10" :  $H_0 : \mu = 10$ , ó "La varianza poblacional es igual a 25" :  $H_0 : \sigma^2 = 25$ , ó ambas propiedades a la vez:  $H_0 : \mu = 10, \sigma^2 = 25$ . Generalmente, los contrastes anteriores se realizan bajo el supuesto de que el carácter de la distribución de

probabilidad poblacional es conocido, por ejemplo, Normal. Este tipo de contrastes son contrastes paramétricos, puesto que se basan en la estimación de algunos parámetros de la distribución de probabilidad poblacional. El investigador no debe olvidar nunca que las propiedades de estos contrastes dependen de que sea correcta la hipótesis que se haya establecido acerca del tipo de distribución poblacional de la que se extrajo la muestra disponible, así como del carácter de dicha muestra (muestra aleatoria simple). Un contraste que tiene muy buenas propiedades XXXX

Existen contrastes no paramétricos, que no precisan de la estimación de parámetros poblacionales, ni descansan en ningún supuesto acerca de la distribución de probabilidad poblacional, que son de un enorme interés en el análisis de datos económicos, a pesar de ser poco habituales. Son contrastes cuyas propiedades son muy robustas (es decir, continúan siendo válidas total o aproximadamente) con independencia del tipo de distribución poblacional. Una segunda razón que confiere enorme interés a los contrastes no paramétricos es que nos permiten discernir el grado en que son válidas hipótesis que no pueden representarse en términos de valores numéricos para los parámetros de la distribución de probabilidad poblacional. Así, en el primero de los ejemplos antes mencionados, queremos contrastar la hipótesis de que la población de la que se extrajo la muestra sigue una distribución Normal. Por supuesto, que también podría contrastarse la hipótesis nula de que obedece una distribución  $t$  de Student, o chi-cuadrado, o cualquier otra. Asimismo, podemos contrastar la hipótesis de que dos muestras proceden de igual distribución de probabilidad, sin necesidad de especificar de qué tipo es ninguna de ellas.

Un tipo de contraste de gran interés para las cuestiones examinadas en el trabajo específico econométrico estriba en el grado de asociación entre variables. Precisamente la Econometría consiste en el conjunto de métodos estadísticos que permiten asignar valores numéricos a los coeficientes de un modelo que trata de representar la relación existente entre un conjunto de variables económicas. Es, por tanto, un análisis de tipo paramétrico; una vez asignados valores numéricos a los parámetros del modelo, generalmente se llevarán a cabo contrastes paramétricos de hipótesis, utilizando los valores numéricos estimados. Existen, asimismo, contrastes no paramétricos que, sin necesidad de pasar por una fase de estimación, permiten discutir si la evidencia muestral es consistente con la hipótesis de que dos variables determinadas están relacionadas entre sí. Es tan fácil llevar a cabo este tipo de contrastes que deberían formar parte, como paso previo a la estimación de todo modelo econométrico. No sería razonable que las conclusiones de tales

contrastes dictaminen la relación de variables que deben incluirse en un modelo econométrico, pero es sumamente ilustrativo complementar la información proporcionada por ambos tipos de contrastes. En definitiva, como proponíamos en la Introducción a este texto, precisamente por su naturaleza probabilística, los métodos estadísticos no deben utilizarse de manera dogmática. En definitiva, se trata de examinar la cuestión que está siendo objeto de análisis en un determinado estudio, a la luz de la información muestral disponible, desde diversas ópticas, con el objeto de proporcionar distintos tipos de evidencia. Por supuesto que en la generalidad de los casos, esas distintas perspectivas no serán todas consistentes entre sí. Debe esperarse del investigador que proporcione toda la información generada en relación con la cuestión que definía la investigación, para que cada lector pueda extraer sus propias conclusiones. En un contexto probabilístico no existen las verdades absolutas, y un determinado análisis puede conducir a conclusiones diferentes.

En un contraste paramétrico se establece, como confrontación a la hipótesis nula, una hipótesis alternativa. La forma que adopte dicha hipótesis no es irrelevante en cuanto a la resolución del contraste de hipótesis. Como ejemplo, tomando como un hecho cierto que la distribución de probabilidad poblacional es Normal, y que la varianza de dicha distribución es conocida, un investigador puede desear contrastar la hipótesis nula  $H_0 : \mu = 10$ , frente a la hipótesis alternativa  $H_1 : \mu \neq 10$ . En este caso, la hipótesis nula es simple, por cuanto que incluye un único valor posible para la esperanza matemática poblacional, mientras que la hipótesis alternativa es compuesta, por cuanto que incluye todo un rango de valores, todos los distintos del incluido en la hipótesis nula. Este contraste tiene, por tanto, otra característica, y es que el conjunto de valores incluidos en ambas hipótesis cubre todo el espacio paramétrico. Un contraste diferente, con la misma hipótesis nula, sería aquél que considerase como hipótesis alternativa,  $H'_1 : \mu < 10$ . Esta es la hipótesis que debería especificar un investigador que sabe que, dada la naturaleza del problema con el que trata, existen razones teóricas para creer que el valor numérico de  $\mu$  no puede exceder de 10. Cuando se dispone de dicha información, el último contraste descrito, que restringe el rango de valores numéricos en la hipótesis alternativa fijando 10 como cota superior, tiene mejores propiedades que el primero de los contrastes, que no establecía tal relación. En la mayoría de las aplicaciones económicas, se dispone de información de este tipo, por lo que el investigador debe especificar cuidadosamente no sólo la hipótesis nula que contrasta, sino también la hipótesis que considera como alternativa, de modo que su contraste de hipótesis tenga las mejores propiedades posibles.

Lamentablemente, este hecho no suele tenerse en cuenta, estableciéndose con demasiada frecuencia hipótesis del tipo  $H_1 : \mu \neq 10$ . No sólo el contraste se resuelve de distinta manera, según sea ésta o  $H'_1$  la hipótesis alternativa; además, como hemos mencionado, las propiedades del contraste son distintas, y mejoran si introducimos en la definición del mismo, información procedente del modelo teórico que sea, por tanto, incuestionable.

- En un contraste de hipótesis, se rechaza la hipótesis nula cuando la información muestral es: i) significativamente contraria a la hipótesis nula, ii) a la vez que es favorable a la hipótesis alternativa.

Este principio, absolutamente básico en la teoría estadística de contrastación de hipótesis, también suele ignorarse con demasiada frecuencia. Por ejemplo, uno de las actuaciones incorrectas en el análisis estadístico de datos económicos, que se produce con cierta frecuencia, se refiere a ocasiones en que la información muestral es contraria al rango de valores numéricos contenidos en la hipótesis nula, pero aún más contraria al rango considerado bajo la hipótesis alternativa. Este sería el caso si al confrontar  $H_0$  frente a  $H'_1$  obtenemos que la media muestral, un estadístico muestral que es estimador eficiente de la esperanza matemática, resulta ser igual a 22, por ejemplo. Otro ejemplo (crecimiento monetario e inflación)

En tal caso, los métodos estadísticos de contrastación de hipótesis conducirán a no rechazar la hipótesis nula, a pesar de que la información muestral es contraria a la misma. Cuando esto sucede, es muy frecuente comprobar que el investigador concluye que su hipótesis nula es válida (con éste u otro calificativo de similares connotaciones). Sin embargo, estamos en una situación donde la información muestral es contraria a dicha hipótesis nula; lo que el investigador debería hacer en este caso es reconocer este hecho, cuestionar la validez de su hipótesis nula, pero cuestionar asimismo el razonamiento que le llevó a establecer la hipótesis alternativa de su contraste, pues la información muestral ha sido aún más contraria a la misma.

En cuanto a la resolución de los contrastes paramétricos de hipótesis, es preciso llevar a cabo previamente un ejercicio de estimación que proporcione una estimación numérica  $\hat{\theta}(x_1, x_2, \dots, x_n)$  a partir del cálculo del valor numérico que toma un determinado estimador  $\hat{\theta}$  del parámetro o vector de parámetros  $\theta$  mediante el que se define la hipótesis nula. Recordemos que un estimador es una función de la información muestral; evidentemente, no todas las funciones de la información muestral, es decir, todos los estimadores posibles, tienen buenas propiedades estadísticas. Para llevar a cabo el contraste de hipótesis es preciso

que la distribución de probabilidad del estadístico  $\hat{\theta}(x_1, x_2, \dots, x_n)$  dependa del parámetro o parámetros incluidos en la hipótesis nula. Por ejemplo,...

A su vez, el contraste de hipótesis tendrá buenas propiedades si: i) ambas hipótesis están correctamente establecidas, ii) los supuestos sobre los que se condiciona el contraste, que pueden referirse al tipo de distribución poblacional, así como al valor numérico de algunos parámetros que no aparecen explícitamente en el contraste, sean correctos, iii) el estadístico utilizado en la resolución del contraste tenga buenas propiedades estadísticas.

Cuando la hipótesis nula es simple, es decir, incluye un único valor numérico, y la hipótesis alternativa es del tipo  $H'_1$ , la resolución del contraste se lleva a cabo mediante la construcción de un intervalo de confianza alrededor del valor numérico del estimador, utilizando la distribución de probabilidad del mismo. Así, llegamos a una afirmación del tipo,

$$\alpha = P \left[ a \leq g \hat{\theta}(X), \theta \leq b \right] \quad (3.1)$$

donde  $\alpha$ , un número positivo próximo a 1, es el nivel de confianza del contraste.

El contraste se resuelve despejando el valor teórico del parámetro desconocido  $\theta$  dentro de la expresión anterior, para obtener una igualdad del tipo,

$$\alpha = P \left[ h_1(\hat{\theta}(X)) \leq f(\theta) \leq h_2(\hat{\theta}(X)) \right] \quad (3.2)$$

donde  $h_1(\hat{\theta}(X)), h_2(\hat{\theta}(X))$  son números reales que dependen de: i) el valor numérico del estimador  $\hat{\theta}(x_1, x_2, \dots, x_n)$  en la muestra disponible, ii) el supuesto acerca de la distribución poblacional, iii) el nivel de confianza (o el nivel de significación) escogidos para el contraste.

En esta expresión, dada una determinada muestra, el valor numérico del estimador  $\hat{\theta}(X)$  es conocido, por lo que puede comprobarse si se satisfacen o no las desigualdades,

$$h_1(\hat{\theta}(X)) \leq f(\theta^0) \leq h_2(\hat{\theta}(X)) \quad (3.3)$$

donde hemos sustituido el valor desconocido de  $\theta$  por el valor numérico incluido en la hipótesis nula,  $\theta^0$ . S

Bajo el supuesto que se ha hecho acerca de la distribución de probabilidad poblacional, y de los valores numéricos de los parámetros que se han supuesto conocidos (si haya alguno), la probabilidad de que el valor numérico de la función  $f(\theta)$  caiga fuera del intervalo (3.3) es de  $1-\alpha$  y, por tanto, pequeña. Es decir,

éste sería un suceso poco probable; si para  $\theta = \theta_0$ , la función  $f(\theta)$  incumple las cotas definidas en (3.3) diremos que dicho valor del parámetro  $\theta$  es poco verosímil, rechazando, en consecuencia, la hipótesis nula.

A modo de ejemplo, recordemos cómo se lleva a cabo un contraste de hipótesis acerca del valor numérico de la esperanza matemática de una población Normal cuya varianza se supone conocida. El punto de partida es la propiedad de la media muestral de tal población, que sigue una distribución asimismo Normal, con la misma esperanza matemática que la población, y con una varianza igual a la varianza poblacional dividida por el tamaño muestral. Así, si la población es  $X \sim N(\mu, 25)$ , la media muestral es una variable aleatoria con distribución  $\bar{x} \sim N(\mu, \frac{25}{n})$ . Por tanto, si se trabaja a un nivel de confianza del 95%, por ejemplo, tendremos,

$$0,95 = P \left[ -1,96 \leq \frac{\bar{x} - \mu}{5/\sqrt{n}} \leq 1,96 \right]$$

donde hemos utilizado el hecho de que  $\bar{x} - \mu \sim N(0, \frac{25}{n})$  y  $\frac{\bar{x} - \mu}{\sqrt{25/n}} \sim N(0, 1)$ . Esta igualdad es la correspondiente a (3.1), donde el parámetro poblacional sobre el que se establece el contraste es la esperanza matemática,  $\theta = \mu$ ; el estimador  $\hat{\theta}(x_1, x_2, \dots, x_n) = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  es la media muestral,  $\alpha = 0,95$ , y  $a = -1,96$ ;  $b = 1,96$ .

De esta igualdad, obtenemos,

$$0,95 = P \left[ -1,96 \frac{5}{\sqrt{n}} - \bar{x} \leq -\mu \leq 1,96 \frac{5}{\sqrt{n}} - \bar{x} \right] = P \left[ \bar{x} - 1,96 \frac{5}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \frac{5}{\sqrt{n}} \right]$$

que es una igualdad análoga a (3.2), con  $h_1(\bar{x}) = \bar{x} - 1,96 \frac{5}{\sqrt{n}}$ ,  $h_2(\bar{x}) = \bar{x} + 1,96 \frac{5}{\sqrt{n}}$ ,  $f(\mu) = \mu$ .

En definitiva, nos queda comprobar si cuando se introduce en esta última igualdad el valor de  $\mu$  que define la hipótesis nula, la cadena de desigualdades se satisface o no. Supongamos que en la muestra disponible, de tamaño 400, se ha calculado para la variable  $X$  una media muestral  $\bar{x} = 7,5$ ; en consecuencia, el intervalo anterior es:  $0,95 = P [7,0 \leq \mu \leq 8,0]$ .

### 3.1. Contrastes de Normalidad

### 3.2. Contrates de asociación

Antes de proceder a la estimación de un modelo específico que establezca una relación paramétrica entre dos variables, conviene explorar la posible existencia

de una relación entre ellas por los procedimientos estadísticos disponibles. Uno de ellos son los contrastes no paramétricos. Como ya hemos comentado anteriormente, una de las virtudes de este tipo de análisis es que su validez y, en particular, los umbrales críticos que debe sobrepasar el estadístico que define el contraste, no dependen de ningún supuesto acerca de la distribución de probabilidad seguida por las variables cuya relación se trata de caracterizar. Frente a otros procedimientos que forman el núcleo tradicional de la Econometría, esto es una ventaja pues, en el caso de los últimos, la Normalidad de la variable cuyo comportamiento se pretende explicar es clave. Como, además, el supuesto de Normalidad acerca de la distribución de probabilidad de una variable económica es muchas veces rechazado, resulta que las buenas propiedades de las estimaciones de un modelo econométrico quedan muchas veces en cuestión. Por eso es conveniente su uso en combinación con otro tipo de procedimientos estadísticos, especialmente si sus propiedades no precisan tal hipótesis.

## **4. Tratamiento de datos**

### **4.1. Ajuste estacional**

### **4.2. Tasas de variación**

### **4.3. Estimación de componentes**

## **5. El modelo lineal simple de regresión**

### **5.1. Descripción del modelo**

#### **5.1.1. Nube de puntos**

### **5.2. Estimación por mínimos cuadrados (ordinarios)**

#### **5.2.1. Representación gráfica de la recta de regresión estimada**

### **5.3. Propiedades del estimador de mínimos cuadrados ordinarios**

Generalmente, estamos muy interesados en contratar hipótesis de distinto tipo: a) si una variable explicativa contiene información significativa acerca de la variable dependiente, b) si el coeficiente de impacto de una determinada variable es igual a 1, c) si dos variables explicativas tienen el mismo coeficiente, etc...

Sin embargo, aunque los coeficientes del modelo de regresión son constantes, si bien desconocidas, sus estimaciones, por cualquier procedimiento que podamos

utilizar, son aleatorias, pues son función de la muestra que utilicemos, que es aleatoria. Si el modelo que estamos estimando es correcto, como hemos de suponer, la perturbación aleatoria del mismo,  $u_t$ , otorga naturaleza asimismo aleatoria a la variable dependiente,  $y_t$ . Esto significa que si cambiamos por ejemplo el período muestral que utilizamos en la estimación, la realización de dicha perturbación, es decir, sus valores numéricos, serán diferentes, con lo que las observaciones de  $y_t$  también los serán, y la estimación de los parámetros diferirá de la obtenida con otro período muestral. Asimismo, si cambiamos la frecuencia de observación de los datos, de diaria a mensual, por ejemplo tomando el último dato de cada mes, la muestra cambia, y con ella, las estimaciones de los coeficientes de las variables explicativas en el modelo.

Siendo variables aleatorias, nos interesa que los estimadores tengan ciertas propiedades deseables, lo cual dependerá del procedimiento de estimación utilizado, y de las características del modelo que estamos estimando. Las principales propiedades en que podemos estar interesados son: insesgo, eficiencia y consistencia.

El insesgo consiste en que la esperanza matemática del estimador coincida con el verdadero valor numérico del coeficiente que estamos estimando. Un estimador eficiente es un estimador de mínima varianza. El procedimiento de mínimos cuadrados proporciona el estimador lineal de mínima varianza, si bien pueden existir otros estimadores no lineales de varianza todavía menor. Un estimador es consistente si, al aumentar el tamaño muestral, converge en probabilidad al verdadero valor del parámetro desconocido que se está estimando. Se dice entonces que su límite en probabilidad es dicho parámetro. Bien podría ocurrir que el estimador fuese sesgado en muestras pequeñas, pero si es consistente, dicho sesgo irá reduciéndose si ampliamos el tamaño muestral. El estimador de mínimos cuadrados no es siempre consistente. El estimador de máxima verosimilitud lo es, pero siempre que la hipótesis acerca de la distribución de probabilidad en que se basa, sea correcta, sobre lo que no se puede tener seguridad.

#### 5.4. Residuos del modelo. Gráficos de residuos

Los residuos del modelo, a veces denominados los errores del modelo de regresión, son aquél componente de la variable dependiente que no está explicado por los valores que toma la variable independiente o explicativa. En consecuencia, los residuos  $\hat{u}_t$  se calculan a partir de la expresión,

$$y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{u}_t$$



de la que se obtiene,

$$\hat{u}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t$$

en el caso de datos de serie temporal, y

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$$

en el caso de datos de sección cruzada.

Gráficamente, si volvemos a la nube de puntos que representa la posible relación entre  $x$  e  $y$ , y dibujamos sobre ella la recta de regresión, el residuo no es sino la distancia vertical entre la altura (ordenada) de cada punto de la nube, y la altura que le correspondería de acuerdo con la recta de regresión estimada. Dicha altura debe tomarse con signo, de modo que el residuo es positivo cuando el punto de la nube está por encima de la recta de regresión estimada, y negativo cuando el punto queda por debajo de la recta de regresión estimada.

#### 5.4.1. Estimación de la varianza del término de perturbación

### 5.5. Cuando los coeficientes del modelo de regresión cambian a lo largo de la muestra

Supongamos que se dispone de datos de serie temporal, y que el coeficiente que mide la relación entre las variables  $x$  e  $y$ , es decir, la pendiente del modelo, ha variado a lo largo del tiempo. Es claro que un procedimiento de estimación como mínimos cuadrados nos proporciona un único valor numérico de dicho coeficiente y, por tanto, cobra pleno interés únicamente bajo el supuesto de que dicho valor numérico ha permanecido constante a lo largo del período muestral.

Sin embargo, en la mayoría de las aplicaciones económicas que pueden considerarse, tal supuesto parece demasiado restrictivo pues, más bien, el valor numérico de dicha elasticidad habrá variado a lo largo de la muestra. ¿Qué proporciona en tal caso el método de mínimos cuadrados? Lo primero que debemos entender es que el investigador no observa en ningún caso si el coeficiente ha variado en el tiempo o no, dado que no observa el valor numérico de dicho coeficiente. Esta es, precisamente, la razón que le mueve a estimar su valor numérico.

En muchas situaciones, sin embargo, el investigador puede tener fundadas creencias acerca de que se han producido variaciones en el mismo. Por ejemplo, muchas veces se afirma que la capacidad de la política monetaria para controlar la inflación se ha reducido significativamente recientemente; tal afirmación se debe

a la observación, en datos reales, de que una fuerte expansión monetaria solía venir acompañada de un claro repunte inflacionista, mientras que, más recientemente, un robusto crecimiento monetario puede ser compatible con una inflación contenida.

En una situación de tal tipo, un procedimiento de estimación como MCO proporciona como valor numérico del parámetro un promedio de los valores numéricos que ha tomado durante el intervalo de tiempo correspondiente a la muestra de datos disponible. Por tanto, resulta de suma importancia el modo en que el coeficiente ha variado en el tiempo, como vamos a ver en los ejercicios de simulación siguientes.

### 5.5.1. Cambio estructural en los coeficientes del modelo de regresión

Supongamos que el coeficiente ha sido constante en la primera mitad de la muestra e igual a 0,5, mientras que en la segunda mitad de la muestra ha sido asimismo constante, e igual a 1,5. El estimador de mínimos cuadrados proporcionaría entonces una estimación en torno a 1,0. En realidad,  $\hat{\beta}_1 = 1,0$  no es representativo de lo que ha ocurrido en ningún momento de la muestra, como ilustra el gráfico XX.

Este tipo de situaciones puede conducir a impresiones engañosas, como ocurriría si en una parte de la muestra la pendiente ha sido positiva, invirtiéndose el signo de la relación entre  $x$  e  $y$  en la segunda mitad de la muestra. Si en ambas partes el valor numérico de la pendiente ha sido el mismo, cambiando únicamente su signo, la estimación resultante será próxima a  $\hat{\beta}_1 = 0,0$ , sugiriendo la ausencia de relación entre ambas variables. Tal conclusión será bastante errónea, pues habría existido una relación, posiblemente bastante exacta entre  $x$  e  $y$  a lo largo de toda la muestra, pero el signo de la misma habría cambiado de la primera a la segunda submuestra, conduciendo a la equívoca estimación mencionada.

Que la estimación proporcionada por un procedimiento del tipo de MCO sea un promedio de los verdaderos valores numéricos (no observados) de la pendiente, no debe tomarse en el sentido de que es la media aritmética de dichos valores numéricos. Sin embargo, tal intuición es aproximadamente correcta. Así, por ejemplo, si en el primer tercio de la muestra la pendiente hubiese sido  $\beta_1 = -1,0$ , y en las dos terceras partes finales de la muestra la pendiente hubiese sido  $\beta_1 = 1,0$ , la estimación numérica del parámetro no sería muy diferente de  $\hat{\beta}_1 = 0,33$ , como correspondería a la media aritmética de los verdaderos valores, ponderados por el número de observaciones a los que aplica cada uno de ellos.

## Ejercicios de simulación

- Ejercicio 1: Simule 300 observaciones de un camino aleatorio  $N(10,25)$  como datos muestrales de la variable explicativa  $x$ . Utilizando un valor numérico de  $-1,0$  para la pendiente, e ignorando el término de perturbación del modelo, simule los 100 primeros datos para la variable dependiente  $y$ . Luego, utilice un valor igual a  $1,0$  para la pendiente y genere los datos ficticios correspondientes a las observaciones 101 a 300 de la variable dependiente. Estime el modelo de regresión simple con dichos datos.

Comentario: Al ejecutar el programa XXX, el lector comprobará que la estimación de la pendiente del modelo se comporta de acuerdo con lo comentado en la sección previa.

Es recomendable ejecutar el programa varias veces, para obtener así un conjunto de estimaciones numéricas de dicho parámetro. En ningún caso se obtendrá estimaciones próximas al verdadero valor de la pendiente, que es de  $+1,0$  en una parte de la muestra, y de  $-1,0$  en la otra, sino más bien un promedio, ponderado de acuerdo con el número de datos u observaciones en el que la pendiente ha tomado uno u otro valor numérico.

El lector puede comprobar que, si aumenta la longitud muestral, las estimaciones que obtiene al ejecutar repetidas veces el programa, se aproximan aún más al promedio que cabría esperar, dado el porcentaje de datos con pendiente igual a  $+1,0$  ó  $-1,0$ .

Los estadísticos resultantes de la estimación presentan toda la apariencia de proceder de un problema estadístico de relación entre dos variables aleatorias. Sin embargo, es muy importante observar que el problema que acabamos de estimar es, en realidad, puramente determinista. No hay en el mismo ningún componente estocástico o aleatorio pues, no hemos utilizado un elemento de perturbación. En nuestra simulación, la relación entre las variables  $x$  e  $y$  es, en todos los períodos exacta, no faltando ninguna otra variable, ni estando sujeta a ningún elemento de error impredecible.

Es precisamente el hecho de tratar como constante un coeficiente del modelo que no lo es, lo que produce la apariencia de ser un problema de naturaleza estocástica. Considerar tal supuesto (erróneo, aunque no lo sabíamos), es comparable a introducir una perturbación estocástica en el verdadero modelo, que incorporaría la variación temporal en la pendiente.

En dicha estimación, aparentemente estadística, obtenemos un R-cuadrado reducido, junto con una pendiente estimada que resulta estadísticamente significativa, si juzgamos por su estadístico tipo- $t$ .

Existe asimismo evidencia de autocorrelación en los residuos, como sugiere el estadístico Durbin-Watson. La presencia de autocorrelación es, en principio, sorprendente, pues hemos generado una variable  $x$  con estructura de ruido blanco  $y$ , por tanto, sin autocorrelación y la variable  $y$  presenta una dependencia exacta y estrictamente contemporánea con ella. Por tanto, no hay ni estructuras de autocorrelación, ni estructuras dinámicas que pudieran producirla. Los indicios de autocorrelación provienen del cambio estructural que se produce en el valor numérico de la pendiente. En el programa se calculan las funciones de autocorrelación simple de ambas variables del modelo; mientras que la variable independiente (explicativa) no presenta autocorrelación, como corresponde a su estructura de ruido blanco que hemos utilizado en su generación, la función de autocorrelación de la variable dependiente sugiere indicios claros de autocorrelación. En realidad, esta variable, por construcción, carece de autocorrelación, lo que sugiere una llamada de atención al uso indiscriminado de las funciones de autocorrelación para detectar autocorrelación serial en una variable o en los residuos de una regresión.

El gráfico XX que representa las observaciones de la variable dependiente junto con los residuos de la regresión muestra la similitud entre ambas variables, lo que es, evidentemente, sinónimo de una relación estimada pobre, a pesar de la significación estadística de la pendiente del modelo. De hecho, puede observarse que la desviación típica de los residuos es similar a la desviación típica de la variable dependiente, lo que significa, como ya discutimos en la Sección XX, que el modelo no tiene una capacidad explicativa importante. Nuevamente, es importante apreciar que ello ocurre junto con un valor claramente significativo del estadístico  $F$  ( $F = T \cdot R^2$ ) de significación conjunta de la regresión estimada. Esta apariencia de validez del modelo, de acuerdo con los estadísticos habituales es, en este caso, falsa, lo que constituye una importante llamada de atención sobre el uso indiscutido de contrastes de hipótesis.

De modo análogo, el gráfico XX, que presenta los valores observados de la variable dependiente, junto con sus valores ajustados, muestra que el ajuste no se adecúa en modo alguno al comportamiento de la variable que pretendíamos explicar con el modelo, y que ha experimentado un cambio estructural entre la primera y la segunda submuestras. Por el contrario, precisamente porque hemos impuesto una pendiente constante a lo largo de toda la muestra, los valores ajustados, que utilizan dicho valor constante de la pendiente, dibujan un comportamiento constante, promedio, durante toda la muestra, que no representa adecuadamente ninguna de las dos submuestras.

- Simule asimismo con perturbación. Construya un gráfico de varianza de la perturbación y varianza de la distribución de la pendiente estimada para un número de observaciones muestrales dado.

### 5.5.2. Variación gradual en los coeficientes del modelo

## 5.6. Algunos modelos de regresión sencillos

### 5.6.1. El modelo constante

Como representación analítica dle comportamiento de una variable, no cabe duda de que el modelo estadístico más sencillo es,

$$y_t = \beta_0 + u_t$$

que especifica que, excepto por un término de perturbación de naturaleza aleatoria, la variable  $y_t$  es constante. Con este modelo, el investigador declara la imposibilidad de encontrar ninguna variable que pueda explicar el comportamiento de  $y_t$ .

Puesto que la matriz de datos de la (única) variable explicativa es en este caso,

$$X = (1, 1, 1, \dots, 1)$$

un vector de dimensión  $T$ , se tiene  $X = 1_T$ , y recordando que  $1_T'1_T = T$ , y  $1_T'y = \sum_{t=1}^T y_t$  podemos adaptar la expresión general (XX) del estimador de mínimos cuadrados a este caso particular, teniendo,

$$\hat{\beta}_0 = (1_T'1_T)^{-1} 1_T'y = \frac{\sum_{t=1}^T y_t}{T} = \bar{y} \quad (5.1)$$

de modo que la estimación de mínimos cuadrados de la constante es la media muestral de la variable que pretendemos explicar. En consecuencia, el residuo es, en cada período,

$$\hat{u}_t = y_t - \hat{\beta}_0 = y_t - \bar{y}$$

es decir, el dato correspondiente, en la forma de desviación respecto a la media muestral. Por tanto, la Suma Residual, o suma de los residuos al cuadrado es,

$$SR = \sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (y_t - \bar{y})^2 = TS_y^2$$

es decir, igual al producto del tamaño muestral por la varianza muestral de la variable dependiente  $y_t$ . Por otra parte, la Suma Total es, como en cualquier modelo de regresión,

$$ST = \sum_{t=1}^T (y_t - \bar{y})^2 = TS_y^2$$

y, por tanto, es específico de este modelo que  $SR = ST$ . En consecuencia, puesto que el modelo tiene una constante, se tiene:  $SE = ST - SR = 0$ . En este modelo, la Suma Residual coincide con la Suma Total, indicador de las variaciones en  $y_t$  que se pretende explicar, por lo que la Suma Explicada es igual a cero. En consecuencia, el R-cuadrado de la regresión es asimismo cero,

$$R^2 = 1 - \frac{SR}{ST} = \frac{SE}{ST} = 0$$

Esto puede parecer paradójico a primera vista; sin embargo, tiene una interpretación totalmente acorde con la naturaleza del modelo de regresión. Como discutimos en la Sección XX, aunque el modelo de regresión se especifica generalmente para relacionar los valores numéricos observados para variables  $y_t, x_t$ , en realidad, su utilización fundamental en Economía estriba en establecer alguna inferencia acerca de la relación que pueda existir entre las variaciones que experimentan las variables explicativas por un lado, y la variable dependiente por otro. Son las fluctuaciones en ambas variables lo que fundamentalmente pretendemos caracterizar. Como vimos al caracterizar el estimador MCO, una vez que tenemos estimaciones numéricas para los coeficientes asociados a las variables explicativas del modelo, obtenemos la estimación numérica de la constante del modelo mediante,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

que relaciona las medias muestrales de todas las variables del modelo. Por tanto, la relevancia de la constante en el modelo estriba en ajustar las distintas medias muestrales de las variables del mismo, corrigido cada una de ellas por el coeficiente asociado. Pero este es un ajuste generalmente poco relevante para el investigador, que se interesa básicamente en el modo en que variaciones en las variables  $x_t$  generan fluctuaciones en  $y_t$ . Por tanto, en este modelo constante, en que no se explican las fluctuaciones de  $y_t$ , llevándose únicamente a cabo el ajuste de medias muestrales a través de (5.1), es lógico que el indicador de ajuste  $R^2$

resulte igual a cero. Nótese, por último, que no hemos impuesto esta condición, sino que estamos interpretando la propiedad del modelo constante, de tener un  $R^2$  igual a cero.

### 5.6.2. El modelo con variables en desviaciones respecto a la media

Consideremos un modelo lineal,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \quad (5.2)$$

cuya versión estimada es,

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \hat{u}_t \quad (5.3)$$

y calculemos su promedio a través de todas las observaciones muestrales. Tendremos,

$$\sum_{t=1}^T y_t = \sum_{t=1}^T \hat{\beta}_0 + \hat{\beta}_1 \sum_{t=1}^T x_{1t} + \hat{\beta}_2 \sum_{t=1}^T x_{2t} + \sum_{t=1}^T \hat{u}_t \quad (5.4)$$

y teniendo en cuenta la propiedad de los residuos MCO de tener suma igual a cero, se convierte en,

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 \quad (5.5)$$

Restando (5.5) de (5.3) se tiene,

$$\hat{y}_t - \bar{y} = \hat{\beta}_1 (x_{1t} - \bar{x}_1) + \hat{\beta}_2 (x_{2t} - \bar{x}_2) + \hat{u}_t$$

en el que se observa que el modelo (5.2) es consistente con un modelo cuyas variables son las de (5.3) pero medidas cada una de ellas en desviaciones respecto a su promedio muestral; los coeficientes estimados por MCO en este modelo coincidirían con los que se estimarían para el modelo original, excepto por el hecho de que el modelo en desviaciones respecto a la media carece de término constante. Por último, los residuos MCO del modelo en desviaciones son, observación a observación, los mismos que se tendrían para el modelo en las variables originales.

La diferencia entre ambos modelos es que el modelo con las variables en desviaciones respecto a la media no precisa de término constante. Es fácil entender por qué: de acuerdo con  $(XX)$ , la estimación MCO de la constante es la diferencia entre la media muestral de la variable dependiente y las medias muestrales de

las variables explicativas, cada una de ellas corregida por el coeficiente asociado. Pero en el modelo con variables en desviaciones respecto a la media, todas las variables, dependiente y explicativas, tiene media muestral igual a cero. Por tanto, la estimación MCO de una hipotética constante en dicho modelo sería igual a cero. Incluso si incluimos dicha constante, su estimación de mínimos cuadrados será numéricamente igual a cero.

### 5.6.3. El modelo con tendencia determinista lineal y cuadrática

Un modelo sencillo interesante es aquél que incluye una tendencia determinista como única variable explicativa, además del término constante,

$$\ln y_t = \beta_0 + \beta_1 t + u_t \quad (5.6)$$

en el que la estimación del coeficiente  $\beta_1$  nos proporciona una estimación de la tasa de crecimiento muestral de la variable  $y_t$ . En efecto, si consideramos la estructura,

$$y_t = A e^{\beta_1 t} \quad (5.7)$$

tendremos una tasa de crecimiento dada por,

$$\frac{dy_t/dt}{y_t} = \frac{d(\ln y_t)}{dt} = \frac{d(\ln A + \beta_1 t)}{dt} = \beta_1$$

En definitiva, el modelo (5.6) no es sino la versión logarítmica de la ecuación de crecimiento (5.7), con  $\beta_0 = \ln(A)$ , y donde el término de perturbación  $u_t$  puede recoger cualquier fluctuación de corto plazo alrededor de la tasa de crecimiento constante,  $\beta_1$ . Este modelo es, por tanto, apropiado cuando se pretende estimar la tasa de crecimiento media de una variable  $y_t$  a lo largo del período muestral. Hay que tener en cuenta, sin embargo que el parámetro  $\beta_1$  proporciona la tasa de crecimiento, supuesta constante, entre cada dos observaciones consecutivas. Si los datos de que disponemos son de naturaleza anual, entonces habremos estimado el crecimiento anual de la variable. Si los datos son trimestrales, el crecimiento anual se obtendrá a partir de la estimación de  $\beta_1$  mediante  $\gamma = (1 + \beta_1)^4 - 1$ , que habremos de multiplicar por 100 si queremos presentar en términos porcentuales. Si los datos de base son de naturaleza mensual, entonces obtendremos una estimación del crecimiento anual a partir de  $\gamma = (1 + \beta_1)^{12} - 1$ .

Este modelo de tendencia determinista lineal es también muy útil precisamente para extraer de una variable su comportamiento de largo plazo, cuando se supone



que éste está bien representado por el supuesto de una tasa de crecimiento constante. Así, una vez estimado el modelo (5.6), los residuos  $\hat{u}_t$  nos proporcionan el logaritmo de la variable  $y_t$  desprovisto de tendencia, es decir,  $\hat{u}_t = \ln y_t - \hat{\beta}_0 - \hat{\beta}_1 t$ , recogiendo así el comportamiento de la variables según fluctúa alrededor de su tendencia de largo plazo. Diremos también que esta es la representación de la variable corregida de tendencia lineal. Como hemos visto, este ejercicio, efectuado sobre el logaritmo de la variable original, incorpora el supuesto implícito de que la tasa de crecimiento de dicha variable es constante.

#### 5.6.4. Modelos no lineales en las variables

Algunos modelos no lineales pueden tratarse de modo muy sencillo, sin necesidad de desarrollar métodos distintos de los estudiados en los capítulos anteriores. Ello ocurre en muchas situaciones en que el modelo presenta no linealidades exclusivamente en las variables que aparecen en la relación que se pretende estimar, como,

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_t^2 + u_t \quad (5.8)$$

que es un modelo con una única variable explicativa, que aparece tanto en su forma original, como al cuadrado. Es interesante observar que, en este modelo, la derivada parcial, que recoge la magnitud de los cambios inducidos en  $y_t$  por un cambio en el valor numérico de  $x_t$  viene dada por,

$$\frac{dy_t}{dx_t} = \beta_1 + 2\beta_2 x_t$$

A diferencia de la que teníamos en el caso del modelo de regresión lineal simple, que era,

$$\frac{dy_t}{dx_t} = \beta_1$$

y, por tanto, constante, la derivada parcial del modelo (5.8) depende del valor numérico de la variable explicativa. Esto puede ser muy interesante en muchas relaciones económicas, en las que es lógico pensar que el impacto que sobre  $y_t$  tiene una variación unitaria en  $x_t$  depende del valor numérico de  $x_t$  a partir del cual se produce dicha variación unitaria. Así, no tendría el mismo impacto negativo sobre el consumo un incremento de un punto en el tipo impositivo del IVA si éste

se produce cuando dicho tipo es del 3%, que si dicho incremento se produce a partir de un tipo del 15%. Parece lógico que así sea.

Si las variables utilizadas en la regresión son los logaritmos naturales de las variables para las que inicialmente obtuvimos datos, como consumo y renta,  $y_t = \ln(Y_t)$ ,  $x_t = \ln(X_t)$ , estaríamos recogiendo en (5.8) la creencia en una elasticidad no constante, como aparece en muchos modelos teóricos económicos. Si el valor numérico del coeficiente  $\beta_2$  es negativo, tendremos que la variable  $y_t$  crece menos que proporcionalmente con un aumento en  $x_t$ , lo que correspondería a una nube de puntos en la forma de una función cóncava, mientras que si estimamos un valor numérico positivo para  $\beta_2$ , tendremos que la variable  $y_t$  crece más que proporcionalmente con un aumento en  $x_t$ , lo cual correspondería a una nube de puntos convexa. Por contraposición, en el modelo lineal incorporamos a priori la hipótesis de elasticidad constante, es decir, el supuesto de que la variable  $y_t$  crece proporcionalmente con  $x_t$ . Sin duda que, aunque no se considera habitualmente, parece conveniente permitir en un primer análisis la posibilidad de una relación cuadrática contrastando, si se desea, la significatividad estadística del coeficiente  $\beta_2$ , si bien dicho contraste habría que verlo a la luz de la discusión de la Sección XX. Así, en muchas ocasiones (como puede ser si se quiere utilizar el modelo estimado con fines predictivos) puede resultar conveniente mantener un posible término cuadrático incluso si el coeficiente  $\beta_2$  asociado aparece como estadísticamente no significativo, en términos del habitual contraste de la  $t$  de Student.

Queda por discutir cómo obtener estimaciones de mínimos cuadrados para este modelo, a pesar de ser una relación no lineal entre variables. Pero este tipo de modelos es muy sencillo de estimar, pues basta definir una nueva variable  $x_{2t} = x_t^2$  para tener el modelo de regresión,

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{2t} + u_t \quad (5.9)$$

que, como modelo lineal que es en las variables explicativas  $x_t$  y  $x_{2t}$ , estimamos por medio de los procedimientos descritos en la Sección XX, teniendo el estimador MCO de este modelo las propiedades habituales.

## 5.7. ¿Cómo especificar un modelo de regresión?

### 5.7.1. ¿Debe incluirse una constante en el modelo de regresión?

Consideremos el modelo de regresión simple,

5.7.2. ¿Debemos estimar en valores originales o en logaritmos de las variables?

5.7.3. ¿Debe estimarse el modelo con variables en niveles o en diferencias?

5.7.4. La frecuencia de observación de las variables

Muchas veces se alusión al aparente principio de que, en cualquier trabajo estadístico con datos, más información es preferible a menos, por lo que, un mayor número de datos es preferible a un número inferior. En este sentido, si el investigador tiene la posibilidad de trabajar con datos mensuales de las variables dependiente e independiente, debe utilizar estos en preferencia al uso de datos anuales. Esto no es en modo alguno cierto, debido al menos a dos consideraciones.

**Errores de medida** Por un lado, las variables económicas, especialmente si son de carácter financiero, tienen un nivel de volatilidad potencialmente importante, por lo que cada dato frecuente que se publica (digamos que mensual) recoge, no sólo la evolución subyacente o verdadera de la variable que se observa, sino también el componente de volatilidad en la misma. esto quiere decir que, al observar una variable cada mes, los datos tienen un componente de naturaleza errática, es decir, puramente aleatorio o impredecible, que puede venir sucedido por un componente de igual naturaleza, pero signo opuesto, al mes siguiente. En consecuencia, si sólo observáramos el dato trimestral, que se forma bien mediante un promedio de los trimestrales, (si la variable es un stock), o acumulando lo tres datos mensuales correspondientes a dicho trimestre (si la variable es un flujo), una buena parte de los componentes erráticos habrá desaparecido por compensación, y el dato trimestral sería, muy posiblemente, más fiable, que los tres datos mensuales. Algo similar puede decirse en cuanto a observar datos trimestrales o anuales. Esta observación está detrás de la sensación que muchas veces se tiene al asistir a la divulgación en los medios de comunicación de un nuevo dato mensual de expectativas de empresarios, inflación, etc.. No es extraño asistir, en variables tan importantes, a un dato mensual aparentemente positivo sigue que se interpreta como negativo, quizá uno siguiente positivo, etc.. De este modo, la avalancha de datos frecuentes genera una incertidumbre que podría evitarse en cierta medida si, aun recogiendo datos mensualmente, se evaluaran sólo con menos frecuencia.

Esto tiene implicaciones cuando se trata de caracterizar la relación que existe entre variables. Si, por ejemplo, estamos interesados en analizar el modo en que

las expectativas de los empresarios acerca de la demanda futura incide en sus decisiones de inversión, la presencia de los componentes erráticos en los datos que se publican nos llevaría a estimar un modelo,

$$inversión_t^* = \beta_0 + \beta_1 \text{expectativas}_t^*$$

donde,

$$\begin{aligned} inversión_t^* &= inversión_t + \varepsilon_t \\ \text{expectativas}_t^* &= \text{expectativas}_t + \xi_t \end{aligned}$$

siendo  $\varepsilon_t, \xi_t$ , los componentes erráticos a los que nos hemos referido, que suponemos no relacionados,  $Corr(\varepsilon_t, \xi_t) = 0$ . Incluso si  $Corr(inversión_t, \text{expectativas}_t)$  es elevada, la presencia de  $\varepsilon_t, \xi_t$ , hará que, generalmente,  $Corr(inversión_t^*, \text{expectativas}_t^*)$  sea inferior a  $Corr(inversión_t, \text{expectativas}_t)$ , proporcionando un ajuste pero del que habríamos tenido en ausencia de dichos componentes erráticos. Como consecuencia de que sólo podemos utilizar en la regresión las medidas  $inversión_t^*, \text{expectativas}_t^*$ , quizá incluso tengamos una estimación del coeficiente  $\beta_1$  no significativa, debido a la pérdida de precisión en su estimación.

**Ejercicio de simulación** Otra posibilidad sería examinar los datos mensuales a la luz de una referencia temporal adecuada.

**La frecuencia en las relaciones estructurales** Otra razón por la cual no es necesariamente preferible utilizar datos más frecuentes que menos se refiere a la propia naturaleza de la relación que se está tratando de medir. Por ejemplo, consideremos la importante relación existente entre crecimiento monetario e inflación. Ningún economista duda de que tal relación exista; algunos, incluso defienden la idea de que la inflación es puramente un fenómeno monetario, estando completamente determinada, por tanto, por la tasa de crecimiento de la cantidad de dinero. Otros economistas no llegan tan lejos, pero aceptan que hay una relación positiva entre crecimiento monetario e inflación; precisamente contrastar si dicha relación es más o menos estrecha, o si un mayor crecimiento monetario se transmite completamente (es decir, con coeficiente  $\beta_1 = 1$ ) a una mayor inflación, puede ser la motivación para estimar un modelo,

$$\text{inflación}_t = \beta_0 + \beta_1 \text{crecimiento monetario} + u_t$$

Ahora bien, ¿verdaderamente creemos que un mayor crecimiento monetario en marzo, por ejemplo, genera una mayor inflación en dicho mes? Muy pocos economistas suscribirían tal concepto. Mucha mayor uniformidad de pareceres encontraríamos en cuanto a creer que observado año a año, mayor crecimiento monetario viene asociado con mayor inflación, mientras que un año en que se instrumenta una política monetaria más restrictiva, definida por una reducción del crecimiento monetario, es un año de inflación menor. Es decir, la proposición conceptual que relaciona positivamente crecimiento monetario e inflación es una proposición referente al medio o largo plazo.

Incluso con datos anuales surgirá, lógicamente, la cuestión de si un menor crecimiento monetario conduce a una menor inflación ese mismo año, el año siguiente, o en ambos. Esta será una cuestión que puede discutirse mediante los procedimientos econométricos y los contrastes estadísticos adecuados, y que conduce a una investigación siempre interesante. Pero ésta es la perspectiva adecuada: en modo alguno tiene sentido pensar que las fluctuaciones que, mes a mes, experimenta el crecimiento monetario se corresponden con las fluctuaciones mensuales que se observan en la tasa de inflación. Este tipo de investigaciones, con datos mensuales, están condenados al fracaso; incluso si estadísticamente detectásemos una relación positiva, deberíamos estar dispuestos a calificarla de espúrea, en el sentido de no ser la relación estructural que cualquier economista estaría dispuesto a admitir entre crecimiento monetario e inflación.

Este ejemplo no debe sino sugerir que las proposiciones económicas teóricas no dicen nada acerca de cuál es la frecuencia en la que se cumplen, y no debemos en modo alguno inferir que una proposición teórica debe cumplirse en todo tipo de datos. Por el contrario, cuando nos proponemos examinar empíricamente una proposición teórica, hemos de pensar cuidadosamente acerca de la frecuencia de datos en que esperamos que ésta se manifieste, y ello debe determinar el tipo de datos a utilizar. El posible cumplimiento de la proposición puede apreciarse en datos de una frecuencia pero no en datos de frecuencia diferente.

**Estructura dinámica de una relación y frecuencia de observación de los datos** Un aspecto relacionado se refiere a las características dinámicas de la relación que se está tratando de estimar: supongamos que existe una relación entre variables económicas que tarda un mes en manifestarse. Así, con datos mensuales, un modelo apropiado sería,

$$y_t = \beta_0 + \beta_1 x_{t-1} + u_t$$

afirmando en este caso que la relación es dinámica, por cuanto que no se manifiesta contemporáneamente, es decir, durante el mismo mes. De acuerdo con este modelo, un incremento en  $x_t$  tendería a conducir a un valor más elevado de  $y_t$  pero no este mes, sino al mes siguiente. Supongamos ahora que sólo disponemos de datos trimestrales: por ejemplo, el dato del primer trimestre será, para ambas variables, el promedio de los datos de enero, febrero y marzo, teniéndose las relaciones,

$$\begin{aligned} y_{febrero} &= \beta_0 + \beta_1 x_{enero} + u_{febrero} \\ y_{marzo} &= \beta_0 + \beta_1 x_{febrero} + u_{marzo} \end{aligned}$$

Aunque  $y_{enero}$  no está relacionado con  $x_{enero}$ ,  $x_{febrero}$ , o  $x_{marzo}$ , las relaciones existentes entre  $y_{febrero}$ ,  $y_{marzo}$  y  $x_{enero}$ ,  $x_{febrero}$  son suficientes para que los datos del primer trimestre de ambas variables,

$$y_{primer\ trimestre} = \frac{y_{enero} + y_{febrero} + y_{marzo}}{3}, \quad x_{primer\ trimestre} = \frac{x_{enero} + x_{febrero} + x_{marzo}}{3}$$

estén relacionados. Sin embargo, debe apreciarse que la relación será entre los datos de  $x$  e  $y$  correspondientes ambos al primer trimestre. En consecuencia, una relación que con datos mensuales tenía una naturaleza dinámica, pasa a ser estrictamente contemporánea con datos trimestrales.

Así, es importante recordar que algunas de las propiedades de una relación econométrica dependen de la frecuencia de observación de los datos, no siendo, por tanto, propiedades de carácter absoluto de la relación entre  $x$  e  $y$ .

## 6. Contrastes de hipótesis en el modelo de regresión lineal simple

Como se ha comentado anteriormente, tres son las finalidades posibles que se derivan de la estimación de un modelo econométrico: a) la posibilidad de contrastar hipótesis económicas teóricas alternativas, b) la predicción de los valores futuros de la variable dependiente

### 6.1. Significación estadística versus precisión

Uno de los contrastes más habituales tras estimar un modelo de regresión se refiere a la hipótesis del tipo  $H_0 : \beta = 0$ , con la que el investigador se pregunta si la

variable asociada a dicho coeficiente tiene un impacto significativo sobre la variable dependiente, cuyo comportamiento se pretende explicar. No debemos olvidar que con ello, lo que estamos contrastando es si dicho impacto es estadísticamente significativo, y tratando de identificar dicha característica con la existencia de un efecto estructural significativo de la variable explicativa sobre la variables dependiente.

Pues bien, ya hemos visto que la manera de resolver el contraste de dicha hipótesis consiste en utilizar el test  $t$  de Student, que en el caso de la hipótesis de significación adopta la forma,

$$\frac{\hat{\beta}}{DT(\hat{\beta})} \sim t_{T-k}$$

donde  $T - k$  denota el número de grados de libertad del modelo de regresión, definido como la diferencia entre el número de observaciones utilizado en la estimación del mismo y el número de coeficientes estimados. Por tanto, para decidir acerca de la hipótesis nula de significación de un coeficiente  $\beta$ , se construye el cociente entre la estimación numérica de dicho coeficiente y su desviación típica estimada, y se compara con el umbral crítico de la distribución  $t_{T-k}$  al nivel de significación escogido de antemano.

La práctica habitual consiste en concluir que la variable explicativa  $x$  no es relevante para explicar el comportamiento de la variable  $y$  si el valor numérico de su estadístico  $t$  de Student es inferior al nivel crítico proporcionado por las tablas de dicha distribución al nivel de significación deseado. Que la comparación se establezca en términos del valor absoluto del estadístico muestral o no depende de que el contraste sea de una o dos colas, lo cual depende a su vez de la forma que adopte la hipótesis alternativa, según sea ésta  $H_1 : \beta \neq 0$ , o adopte alguna de las formas  $H_1 : \beta < 0$ ,  $H_1 : \beta > 0$ .

La forma que adopta este contraste sugiere de modo bastante evidente que el estadístico  $t$  puede ser inferior al umbral crítico de la distribución de referencia a) bien porque el valor estimado  $\hat{\beta}$  sea pequeño incluso cuando se tiene en cuenta el rango de variación de la variable asociada, o b) porque, aun siendo  $\hat{\beta}$  suficiente como para generar un impacto cuantitativo apreciable de  $x$  sobre  $y$ , dicho valor numérico se estima con poca precisión, es decir, con una desviación típica elevada. Mientras el primer caso corresponde a una situación en la que querríamos concluir que, efectivamente, la variable  $x$  no es relevante para explicar  $y$ , en el segundo caso, tal conclusión sería errónea; lo que está sucediendo en este caso es únicamente que la muestra disponible no nos permite asignar un valor numérico concreto al

coeficiente asociado, a pesar de que la variable  $x$  es un factor explicativo relevante de la variable  $y$ .

A pesar de sus importantes implicaciones para el contraste de hipótesis estadísticas, esta discusión se ignora con demasiada frecuencia en el trabajo empírico. Su importancia deriva de que, como hemos ido revisando con anterioridad, existen distintas razones que pueden implicar una pérdida de precisión en la estimación puntual, con independencia del contenido informativo que la variable  $x$  tiene sobre  $y$ . Por ejemplo, aparece una pérdida de precisión apreciable cuando el valor numérico del coeficiente  $\beta$  ha variado a lo largo del intervalo muestral [Recuérdese el ejercicio de simulación XXX]. en distintos puntos Esta se numéricamente de Con mucha frecuencia se ignora

**6.1.1. ¿Se hace una variable más o menos significativa?**

**6.1.2. ¿Cómo puede discutirse qué variable es más relevante en una regresión?**

mayor valor numérico

mayor estadístico  $t$

## **7. Correlación versus causalidad**

## **8. Variables no estacionarias**

La no estacionariedad de las variables involucradas en una regresión es uno de las situaciones que requiere una consideración más cuidadosa en el análisis de regresión. La ausencia de estacionariedad se produce con mucha frecuencia en variables económicas; además, como vamos a ver, sus implicaciones en la estimación de modelos de regresión pueden ser bastante negativas e importantes. Por último, su detección y tratamiento no son siempre evidentes.

### **8.1. Características de una variable estacionaria**

Una variable estacionaria tiene generalmente varianza finita (salvo que obedezca a una distribución que como la Cauchy, carece de este momento); más precisamente, su varianza no cambia con el paso del tiempo y, desde luego, no tiende a infinito. una perturbación transitoria sobre una variable estacionaria tiene efectos



puramente transitorios; pueden durar varios períodos, pero sus efectos terminan desapareciendo. Los valores sucesivos de su función de autocorrelación convergen rápidamente hacia cero, excepto quizá en los retardos de carácter estacional. La serie temporal correspondiente a una variable estacionaria no deambula durante períodos largos de tiempo a un mismo lado de su media muestral, sino que cruza frecuentemente dicho nivel medio. El número medio de períodos que transcurre entre dos cruces consecutivos del nivel medio muestral es pequeño.

Por el contrario, una perturbación de carácter transitorio sobre una variable no estacionaria tiene efectos permanentes. La función de autocorrelación de una variable no estacionaria converge a cero muy lentamente, y su serie temporal muestra claramente largos períodos de tiempo en que deambula sin cruzar su nivel medio.

## **8.2. Tendencias deterministas y tendencias estocásticas**

La ausencia de estacionariedad en variables económicas puede reflejarse mediante la presencia de tendencias estocásticas o de tendencias deterministas en los precios de mercado, a través de volatilidad cambiante en el tiempo, etc.. Una tendencia estocástica es un componente estocástico cuya varianza tiende a infinito con el paso del tiempo. Una tendencia determinista es una función exacta del tiempo, generalmente lineal o cuadrática, lo que hace que el valor de la variable crezca o disminuya constantemente; si la tendencia es lineal, la variable tenderá a más o menos infinito; si la tendencia es cuadrática o de orden superior, la variable puede estar acotada.

Si una variable presenta una tendencia determinista lineal, su valor esperado tenderá a aumentar o disminuir continuamente, con lo que será imposible mantener el supuesto de que la esperanza matemática de la sucesión de variables aleatorias que configura el proceso estocástico correspondiente a dicha variable, es constante. En consecuencia, tampoco podrá mantenerse que la distribución de probabilidad de dichas variables es la misma a través del tiempo. Sin embargo, si efectuamos una correcta especificación de la estructura de dicha tendencia, podrá estimarse y extraerse del precio, para obtener una variable estacionaria, que no presentaría las dificultades antes mencionadas.

Mayor dificultad entraña el caso en que una variable precio incluye una tendencia estocástica pues, en tal caso, su esperanza y varianza no están definidas. La presencia de una tendencia estocástica requiere transformar la variable, generalmente en primeras diferencias temporales, o tomando las diferencias entre las

observaciones correspondientes a una misma estación cronológica, en el caso de una variable estacional. La transformación mediante diferencias resulta bastante natural en el análisis de datos financieros, por cuanto que la primera diferencia del logaritmo de un precio, en logaritmos, es la rentabilidad del activo, lo que hace que la transformación logarítmica sea utilizada muy frecuentemente.

Como prácticamente ningún precio o índice financiero es estacionario, el uso indiscriminado de un estadístico como la varianza o la desviación típica como indicador de riesgo conduce a medidas de volatilidad sesgadas al alza. Consideremos un modelo muy popular en el análisis de mercados financieros, el camino aleatorio:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots$$

que evoluciona a partir de un valor inicial  $y_0$  dado, donde  $\varepsilon_t$  es un ruido blanco: sucesión de variables aleatorias, independientes, con media constante (que suponemos cero), y varianza asimismo constante  $\sigma_\varepsilon^2$ . Mediante sucesivas sustituciones, este proceso puede escribirse, de modo equivalente:

$$y_t = y_0 + t\mu + \sum_{s=1}^t \varepsilon_s$$

En consecuencia, un camino aleatorio  $y_t$  tiene varianza creciente en el tiempo:

$$Var(y_t) = t\sigma_\varepsilon^2$$

Ello se debe a que el último sumando en la representación anterior es un ejemplo de tendencia estocástica. Cuanto mayor sea el número de observaciones consideradas, mayor será la varianza muestral del camino aleatorio: un camino aleatorio tiene menor varianza a lo largo de una hora que a lo largo de un día, a lo largo de un día que a lo largo de una semana, etc..

Esto es lo que ocurrirá con la inmensa mayoría de los precios cotizados en los mercados financieros. Aunque la presencia de tendencias estocásticas se produce generalmente junto con estructuras más complejas que la de un camino aleatorio, la implicación acerca de una varianza creciente con el tiempo se mantiene cuando se añaden a ésta componentes autoregresivos o de medias móviles para  $y_t$ . Para evitarlo, caracterizamos la volatilidad de un mercado o de un activo analizando el comportamiento de la rentabilidad que ofrece a lo largo del tiempo, no de su precio o cotización.

En el ejemplo anterior, de un camino aleatorio, la tendencia estocástica aparece debido al coeficiente unitario del retardo de  $y_t$  en la ecuación que explica el comportamiento de esta variable, por lo que una tendencia estocástica se conoce asimismo como una raíz unitaria. Con más generalidad, recordemos que por la descomposición de Wald, todo proceso estacionario acepta una representación autoregresiva, quizá de orden infinito,

$$y_t = \alpha_0 + \sum_{s=1}^{\infty} \alpha_s y_{t-s} = \alpha(L) y_t$$

donde  $L$  denota el operador de retardos, definido como  $L^j y_t = y_{t-j}$ . Si obtenemos las raíces de dicho polinomio de retardos, podremos escribir,  $\alpha(L) = \prod_{i=1}^p (1 - a_i L) \prod_{j=1}^q (1 - b_j L - c_j L^2)$ , donde los últimos factores tienen como raíces dos números complejos conjugados. una raíz unitaria es un factor del primer tipo, con  $a_i = 1$ . En el lenguaje estadístico, se dice que el proceso  $y_t$  tiene una raíz unitaria.

Si el proceso  $y_t$  siguiese una estructura dependiente de su pasado, pero del tipo:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \quad t = 1, 2, \dots, \quad -1 < \phi_1 < 1$$

sus propiedades serían bastante distintas, con:

$$y_t = \phi_0 \frac{1 - \phi_1^t}{1 - \phi_1} + \phi_1^t y_0 + \sum_{s=1}^t \phi_1^{t-s} \varepsilon_s$$

y si consideramos que el proceso ha durado infinitos períodos,

$$E(y_t) = \frac{\phi_0}{1 - \phi_1}; \quad Var(y_t) = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

estarían bien definidas, son constantes, y el proceso es estacionario. Este proceso se denomina proceso autoregresivo de primer orden y en él hay que distinguir entre momentos incondicionales, cuyas expresiones analíticas acabamos de calcular en el caso de esperanza matemática y varianza, y momentos condicionales.

donde suponemos que  $u_t$  es un proceso sin autocorrelación (correlación temporal consigo mismo). Es decir,  $Corr(u_t, u_{t-k}) = 0 \quad \forall k$ .

En estas condiciones, si  $u_t$  sigue una distribución Normal  $u_t \sim N(0, \sigma_u^2)$ , entonces  $y_t$  sigue una distribución

$$y_t \sim N\left(\frac{\phi_0}{1 - \phi_1}, \frac{\sigma_u^2}{1 - \phi_1^2}\right)$$

Esta es la distribución marginal o incondicional, de  $y_t$ .

Por otra parte, condicional en la historia pasada de  $y_t$ , sin incluir el dato de fecha  $t$ , la distribución de probabilidad condicional de  $y_t$  es,

$$y_t \sim N(\phi_0 + \phi_1 y_{t-1}, \sigma_u^2)$$

que tiene una menor varianza. De hecho, la varianza incondicional de  $y_t$  es tanto mayor cuanto más se acerque el parámetro  $\phi_1$  a 1, creciendo dicha varianza sin límite. Sin embargo, la varianza condicional es siempre  $\sigma_u^2$ , con independencia del valor numérico del parámetro  $\phi_1$ .

La varianza condicional de  $y_t$  es igual a la varianza de  $u_t$ ,  $\sigma_u^2$ , mientras que la varianza incondicional de  $y_t$  es siempre mayor que  $\sigma_u^2$ .

Además,

$$E(y_t/y_{t-1}) = \phi_0 + \phi_1 y_{t-1}; \quad E(y_t) = \frac{\phi_0}{1 - \phi_1}$$

Como veremos más adelante, el concepto de proceso browniano está bastante ligado al de camino aleatorio. Por tanto, la afirmación anterior es coherente con establecer la hipótesis de que la rentabilidad de un determinado activo sigue un proceso browniano, pero no tanto con efectuar dicha hipótesis sobre su precio.

### 8.3. Regresión espúrea

El problema de la regresión espúrea fue analizado por Granger y Newbold (1974), quienes mostraron la posibilidad de que, en determinadas situaciones, estimaciones mínimocuadráticas de un modelo de regresión lineal que sugieren una estrecha relación entre variable dependiente y variables independientes, están reflejando, en realidad, una relación espúrea o ficticia, que en realidad no existe. Es evidente que tal posibilidad sería extremadamente peligrosa, tanto en la estimación de coeficientes de impacto o elasticidades, como en la contrastación de hipótesis teóricas. Lo que suele ignorarse con demasiada frecuencia es que las condiciones para que una regresión sea espúrea se dan con mucha frecuencia en la investigación aplicada en Economía, en general, y en Finanzas, en particular.

Comenzamos describiendo el tipo de dificultades a que puede dar lugar la ausencia de estacionariedad de las variables dependiente e independiente en un modelo de regresión lineal. Para ello, pensemos en el siguiente ejercicio: comenzamos simulando dos ruidos blancos independientes,  $\varepsilon_{x_t}, \varepsilon_{y_t}, t = 1, 2, \dots, T$ , a partir de distribuciones de probabilidad Normal, con esperanza matemática  $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$  (por ejemplo, iguales a cero) y varianzas  $\sigma_{\varepsilon_x}^2, \sigma_{\varepsilon_y}^2$ ; el coeficiente de correlación muestral entre las series temporales resultantes será, por construcción, muy reducido, si bien no exactamente igual a cero.

Nota: Cuanto mayor sea el tamaño muestral, más probable es que dicha correlación sea igual a cero, debido a que la correlación muestral, es decir, la correlación entre las dos series temporales simuladas es, por la ley de los grandes números, un estimador consistente de su análogo poblacional, que es el coeficiente de correlación teórico entre los dos procesos  $\varepsilon_{x_t}, \varepsilon_{y_t}$ , que es igual a cero. Por tanto, al aumentar  $T$ , la distribución de probabilidad del coeficiente de correlación muestral se concentra alrededor de cero.

El gráfico de ambas variables presentará una pauta oscilando alrededor de su media muestral que, por la misma razón apuntada para el coeficiente de correlación, serán próximas, si bien no iguales, a  $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$ . Observaremos que cada serie temporal cruza repetidamente su nivel medio. Si estimamos una regresión del tipo:

$$\varepsilon_{y_t} = \beta_0 + \beta_1 \varepsilon_{x_t} + u_t, \quad t = 1, 2, \dots, T$$

deberíamos obtener una estimación de  $\beta_1$  no significativamente diferente de cero, y un  $R^2$  prácticamente nulo. En efecto, salvo por el error estadístico, así ocurre cuando llevamos a cabo un ejercicio de simulación de Monte Carlo: al 95% de confianza, el habitual contraste tipo t rechazará la hipótesis nula de ausencia de capacidad explicativa de  $\varepsilon_{x_t}$   $H_0 : \beta_1 = 0$  aproximadamente en un 5% de los casos, y el valor mediana del coeficiente de determinación  $R^2$  para todas las simulaciones es muy reducido. El término constante sólo resultaría significativo si en la generación de las series temporales, hemos utilizado valores diferentes de  $\mu_{\varepsilon_x}, \mu_{\varepsilon_y}$ .

Este resultado no se ve afectado significativamente en ningún otro sentido por la presencia de tales términos constantes, ni tampoco por cambios en el valor de las respectivas varianzas. Al variar el valor relativo de  $\sigma_{\varepsilon_y}^2 / \sigma_{\varepsilon_x}^2$  tan sólo se observa un comportamiento algo errático del tamaño del contraste de significación del parámetro  $\beta_0$ . En definitiva, en esta primera parte del ejercicio tendremos el

resultado que esperaríamos: una regresión no significativa, excepto en lo relativo al nivel escogido para el contraste.

### 8.3.1. Regresión espúrea bajo tendencias deterministas

A continuación, añadimos una tendencia lineal determinista a cada una de ellos,

$$\begin{aligned}y_t^* &= at + \varepsilon_{y_t} \\x_t^* &= bt + \varepsilon_{x_t}\end{aligned}$$

donde  $a$  y  $b$  son constantes arbitrarias y  $t$  es una tendencia determinista, es decir, una variable que aumenta cada período en una cantidad constante,  $\Delta$ .

Si calculamos el coeficiente de correlación muestral entre  $x_t$  e  $y_t$ , apreciaremos que es elevado. Esto es sorprendente porque, como muestran las expresiones anteriores, cada variable es la suma de un componente de naturaleza determinista, que no experimenta ninguna fluctuación aleatoria, y un segundo componente de naturaleza estocástica. El coeficiente de correlación debería indicar la asociación estadística entre ambas variables, que es lo mismo que la asociación entre sus componentes estocásticos, es decir, entre sus innovaciones. Pero dicha correlación debería ser, por construcción, prácticamente igual a cero, en contra del resultado que se obtiene cuando se lleva a cabo este ejercicio de simulación. En todo caso, tal elevada correlación no refleja ninguna relación real entre las variables, por lo que se denomina correlación espúrea.

Como consecuencia de la misma, si se estima una regresión lineal, tomando cualquiera de estas variables como variable dependiente y la otra como independiente,

$$y_t^* = \beta_0 + \beta_1 x_t^* + v_t, \quad t = 1, 2, \dots, T$$

los resultados cambian sustancialmente: se obtiene un R-cuadrado elevado pues, como ya sabemos, es igual al cuadrado del coeficiente de correlación entre ambas variables, a la vez que una pendiente  $\beta_1$  aparentemente significativa, de acuerdo con el criterio habitual de utilizar su estadístico tipo t-Student. Ambas cosas ocurrirán en un elevado porcentaje de las simulaciones que realicemos, para distintas series temporales de  $\varepsilon_{x_t}, \varepsilon_{y_t}, t = 1, 2, \dots, T$ . Por consiguiente, creeríamos que la capacidad explicativa de la variable  $x_t^*$  sobre  $y_t^*$  es muy importante. Este resultado es sorprendente, por cuanto que las variables  $y_t^*, x_t^*$  tienen la misma

estructura estocástica que  $\varepsilon_{x_t}, \varepsilon_{y_t}$ , por lo que ambas relaciones deberían proporcionar resultados análogos. Esta apariencia ficticia de capacidad explicativa es lo que se conoce como regresión espúrea.

El grado de correlación observado entre  $y_t^*, x_t^*$  depende de dos factores: la similitud entre las constantes  $a$  y  $b$ , y la relación entre ambas y las desviaciones típicas de los ruidos blancos originales,  $\varepsilon_{x_t}, \varepsilon_{y_t}$ . Si, por ejemplo, fijamos el valor numérico de  $b$  en  $b = 1$ , y vamos tomando en cada ejercicio de simulación valores:  $a = [0, 1; 0, 5; 0, 9; 1; 3; 10; 100]$  el coeficiente de determinación resultante,  $R^2$  aumenta monótonicamente con el valor de  $a$ . Es decir, la mayor correlación no se obtiene cuando hay a ambos lados de la igualdad la misma pendiente, lo que equivaldría a utilizar  $a = 1$ , sino que dicha correlación aumenta con  $a$ . Esto se debe a que, según aumenta  $a$ , cada vez hay más tendencia determinista en  $y_t^*$ , en el sentido de que ésta predomina sobre el componente estocástico  $\varepsilon_{y_t}$ , y dicha tendencia determinista puede explicarse muy adecuadamente mediante el componente análogo de  $x_t^*$ .

### 8.3.2. Regresión espúrea bajo tendencias estocásticas

En su trabajo pionero, Granger y Newbold (1974) trataron el problema de no estacionariedad producido por la presencia de tendencias estocásticas o raíces unitarias. Para ello, realizaron el siguiente ejercicio: a partir de la simulación de dos ruidos blancos independientes que tendrán, por construcción, como antes, un coeficiente de correlación muestral muy reducido, añadieron una raíz unitaria o tendencia estocástica a cada uno de ellos,

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_{y_t} \\ x_t &= x_{t-1} + \varepsilon_{x_t} \end{aligned}$$

obteniendo que el coeficiente de correlación entre  $x_t$  e  $y_t$  era muy próximo a la unidad. Esto es sorprendente, por cuanto que, a partir de condiciones iniciales conocidas, los valores de ambas variables en cada instante de tiempo pueden escribirse como,

$$y_t = y_0 + \sum_{s=1}^t \varepsilon_{y_s}$$

que indican que la evolución temporal de cada una de las variables se debe a la acumulación temporal de sus innovaciones. Por tanto, la naturaleza estocástica de

cada variable está totalmente determinada por la naturaleza de sus innovaciones. Si  $\varepsilon_{x_t}$  y  $\varepsilon_{y_t}$  son independientes, entonces también deberían serlo  $x_t$  e  $y_t$ , en contra de los valores obtenidos para sus coeficientes de correlación muestrales en repetidas simulaciones. En todo caso, nuevamente, tal elevada correlación no refleja ninguna relación real entre las variables, por lo que se denomina correlación espúrea.

Si estimamos una regresión lineal entre estas variables, en cualquier orden, tendremos de nuevo un  $R$ -cuadrado elevado y una pendiente significativa, de acuerdo con el criterio habitual de utilizar su estadístico tipo t-Student, pero la evidencia de capacidad explicativa proporcionada por esta regresión sería espúrea.

Si las series temporales obtenidas mediante simulación para las innovaciones o ruidos blancos  $\varepsilon_{x_t}$  y  $\varepsilon_{y_t}$  tuviesen correlación distinta de cero, las variables  $x_t$  e  $y_t$  de los ejemplos anteriores mostrarían correlaciones muestrales similares a las que se encuentran en los ejercicios de simulación descritos. En ese caso, los elevados coeficientes de correlación no serían tan engañosos, si bien serían numéricamente más altos de lo que la correlación entre  $x_t$  e  $y_t$  haría esperar.

En un ejercicio de simulación como el descrito, Granger y Newbold encontraron una frecuencia aproximada de rechazos de la hipótesis nula  $H_0 : \beta_1 = 0$  del 76%. La frecuencia de rechazos de la capacidad explicativa global de la regresión se eleva muy significativamente al aumentar el número de variables explicativas independientes con estructura de ruido blanco. Nuevamente los coeficientes de determinación son muy elevados, lo que sorprende, pues realmente,  $x_t$  no explica apenas a  $y_t$ . El estadístico de Durbin-Watson habitualmente utilizado para contrastar ausencia de autocorrelación se reduce hacia cero, por lo que la combinación de este hecho con un elevado  $R^2$  suele utilizarse como indicio de una regresión espúrea.

- Ejercicio de simulación

En `espureo.xls` se han generado dos series temporales correspondientes a una población Normal  $N(0,1)$ . El generador de números aleatorios de Excel produce observaciones independientes entre sí, por lo que ambas series temporales se obtienen asimismo de manera independiente. La correlación poblacional entre ellas es cero, si bien la correlación muestral, al final de ambas variables, es de 0,0278. Tampoco la media y desviación típica muestrales de cada variable son exactamente 0 y 1, como sus valores teóricos, si bien no difieren mucho de ellos. El coeficiente de asimetría teórico, así como el exceso de curtosis (respecto de 3.0, que es la curtosis de toda población Normal), deberían ser ambos igual a cero lo que, nuevamente, ocurre sólo con carácter aproximado.



En `regresion_ori` se presentan los resultados de estimar una regresión entre las variables originalmente obtenidas por simulación, la primera de ellas actuando como variable dependiente, la segunda como variable explicativa. El coeficiente de determinación es el cuadrado de su coeficiente de correlación y, por tanto, muy reducido. La estimación del coeficiente asociado a la variable explicativa aparece como no significativamente diferente de cero, de acuerdo con el estadístico  $t$  habitual. El gráfico que presenta el ajuste de la recta a la nube de puntos muestra una línea sin apenas pendiente, y una nube de puntos bastante circular. Ambos hechos reflejan una escasa correlación: una pendiente no significativa sugiere que la variable explicativa puede cambiar de valor sin que la variable dependiente cambie; una nube de puntos circular muestra que el rango de valores de cada una de las dos variables asociado a un valor determinado de la otra es muy amplio. En consecuencia, un valor numérico de cualquiera de ellas apenas nos informa acerca del valor de la otra variable. Esta es la manifestación de la ausencia de correlación entre ambas.

Lo contrario ocurre al estimar una regresión lineal entre las variables [`regresion_tend`], una vez que se ha añadido una tendencia determinista a cada una de ellas. Para ello, en la hoja Datos se han generado dos nuevas variables, sumando una tendencia lineal a la variable Y, y 0,6 veces la misma tendencia lineal a la variable X. Los estadísticos muestrales que aparecen al pie de dichas variables carecen de justificación estadística, como comentaremos en una sección posterior. Aunque el componente estocástico en ambas variables es el mismo de antes, la nube de puntos entre ambas tiene un perfil totalmente distinto, siendo prácticamente una línea recta. Esto se debe a que el componente tendencial predomina sobre el estocástico; como consecuencia, la regresión estimada entre ambas variables muestra un coeficiente de determinación muy próximo a la unidad. Lo que es quizá más preocupante, es que la pendiente estimada en dicha regresión, que es sustancialmente más elevada que la estimada con las variables originales, aparece como claramente significativa, sugiriendo una importante capacidad explicativa a la variable independiente, contrariamente a lo que detectamos en la primera regresión. Aquí cabe discutir si este es un resultado razonable: podría argumentarse que ambas variables tienen un componente tendencial importante y que, en ese sentido, no es sorprendente que el coeficiente de determinación entre ambas sea elevado. Es cierto, pero sólo refleja la relación entre los componentes deterministas, que no son los que deben concentrar la atención del analista: si hay componentes deterministas en las variables dependiente y explicativas, el analista debería indagar las razones que explican la presencia simultánea de tales elementos

en las variables. Esta sería uno de los elementos del análisis; el segundo consistiría en evaluar la relación entre los componentes estocásticos en variables dependiente y explicativas; éste elemento es importante, pues nos proporciona información acerca del impacto que sobre la variable dependiente puede tener una intervención sobre alguna de las variables explicativas, o una alteración exógena en su valor numérico. La dificultad con la regresión anterior estriba en que si no se lleva a cabo este análisis por componentes, el examen mecánico de los resultados de la regresión sugeriría que la variable dependiente reacciona a fluctuaciones inesperadas en la variable explicativa, cuando no es así; tal conclusión sería un error.

Posteriormente, hemos mantenido los mismos elementos tendenciales de ambas variables, pero hemos incrementado de manera apreciable el componente aleatorio en ellas. Siendo tales componentes variables aleatorias de esperanza matemática igual a cero, su tamaño queda representado por su desviación típica, que era unitaria para ambas en las regresiones anteriores. En la hoja Datos hemos generado otras dos variables con desviaciones típicas 20 y 30; la correlación entre ellas desciende, si bien no de manera dramática, situándose en 0,9349. Por último, hemos mantenido estos componentes estocásticos, pero reduciendo el incremento período a período de la tendencia, que pasa de ser 1,0 a ser ahora 0,10. El coeficiente de correlación entre ambas variables se reduce ahora a 0,2446; la regresión entre ambas variables todavía muestra una pendiente significativamente diferente de cero de acuerdo con el uso habitual del estadístico  $t$ , pero de manera menos evidente que antes; el coeficiente de determinación es igual a  $0,2446^2 = 0,0598$ , bastante reducido. El gráfico que muestra la nube de puntos, junto con la recta ajustada, ilustra la dificultad de precisar la pendiente de la recta que mejor se ajusta a dicha nube de puntos, es decir, la dificultad de estimar con precisión dicha pendiente.

#### 8.4. Tratamiento de tendencias deterministas

De las dos situaciones descritas en el apartado anterior, es algo más sencilla de tratar la presencia de tendencias deterministas, cuando se anticipa correctamente que la presencia de las mismas es la única causa de no estacionariedad de las variables que se pretende relacionar, es decir, cuando las variables tienen estructura,

$$y_t = \alpha_0 + \alpha_1 t + \varepsilon_{y_t} \quad (8.1)$$

$$x_t = \delta_0 + \delta_1 t + \varepsilon_{x_t} \quad (8.2)$$

Para ello, hay dos posibilidades: la primera consiste en incorporar en el modelo de regresión una tendencia determinista lineal como variable explicativa,

$$y_t = \alpha + \gamma t + \beta x_t + u_t \quad (8.3)$$

en la que el coeficiente estimado  $\hat{\beta}$  y su desviación típica serán, aproximadamente, los mismos que habríamos estimado en la regresión,

$$\varepsilon_{y_t} = \eta_0 + \eta_1 \varepsilon_{x_t} \quad (8.4)$$

Esto significa que si ambas innovaciones son independientes, en la regresión (8.3) se tendrá un coeficiente reducido en magnitud, y estadísticamente no significativo, en términos de su estadístico  $t$  de Student.

Esto es distinto del resultado que se obtiene en la estimación de la regresión habitual,

$$y_t = \alpha + \beta x_t + u_t \quad (8.5)$$

cuando las variables tienen estructura (8.2), (8.1), en la que se tendría un R-cuadrado muy elevado, una estimación numérica de  $\beta$  relativamente elevada, y un estadístico  $t$  para dicho coeficiente, claramente por encima de 2,0 en valor absoluto, sugiriendo que la capacidad de  $x_t$  para explicar  $y_t$  es significativa, contrariamente a lo que, en realidad, ocurre.

La dificultad con el procedimiento que hemos sugerido es que todavía mantendrá un R-cuadrado muy elevado, debido a la capacidad explicativa que el término  $\gamma t$  tiene sobre  $y_t$ , debido a la presencia de una tendencia determinista en esta última variable. Este término aparecerá como claramente significativo, con un estadístico  $t$  muy elevado.

La diferenciación elimina asimismo las tendencias deterministas, como fácilmente puede comprobarse algebraicamente. De este modo, si el precio de un determinado activo tiene una tendencia temporal determinista lineal, su primera diferencia estará libre de dicha tendencia. Un proceso con una tendencia determinista cuadrática sigue trayectorias con formas parabólicas, cóncavas o convexas. Su primera diferencia presentará una tendencia lineal, mientras que su segunda diferencia estará libre de tendencia. Un proceso con una tendencia determinista representada por un polinomio de grado tres puede tener ciclos. La primera diferencia de este proceso tendrá una tendencia cuadrática.

Como ejemplo, consideremos:

$$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$$

cuya primera diferencia es:

$$\Delta y_t = y_t - y_{t-1} = (\beta_1 - \beta_2) + 2\beta_2 t + (\varepsilon_t - \varepsilon_{t-1})$$

y su segunda diferencia:

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} = y_t - 2y_{t-1} + y_{t-2} = 2\beta_2 + (\varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2})$$

Por tanto, aparentemente, una solución en el caso en que sospechamos que puede haber tendencias deterministas en las variables que pretendemos relacionar, consistiría en estimar la posible relación entre ellas después de haber tomado diferencias temporales. Sin embargo, con dicha transformación perdemos bastante información acerca de las fluctuaciones de corto plazo en las variables, por lo que los procedimientos anteriormente descritos son más recomendables.

### 8.5. Ejercicios de simulación

- Ejercicio 1: Simule 300 observaciones de dos ruidos blancos independientes, con distribuciones  $N(\mu_x, \sigma_x^2)$ ,  $N(\mu_y, \sigma_y^2)$ , como observaciones muestrales para las innovaciones  $\varepsilon_{x_t}, \varepsilon_{y_t}$ . A continuación, genere observaciones para una tendencia determinista  $t$ . Los valores numéricos para las variables  $x$  e  $y$  se obtienen añadiendo la tendencia  $t$ , multiplicada por sendas constantes  $a, b$ , a las respectivas innovaciones, para reproducir las estructuras (8.2), (8.1).

El ejercicio consiste en comparar el coeficiente de correlación que se obtiene para  $\varepsilon_{x_t}, \varepsilon_{y_t}$ , que será muy reducido, con el que se obtiene entre  $x_t$  e  $y_t$ , que debería ser similar, pero será, sin embargo, muy elevado. En segundo lugar, deben estimarse regresiones análogas a (8.4), (8.5), (8.3).

### 8.6. Tendencias estocásticas y raíces unitarias

De modo análogo, un proceso puede tener asimismo varias raíces unitarias. Los tipos de interés ya son rentabilidades, por lo que tienen, generalmente, un orden de no estacionariedad (es decir, un número de tendencias) menos que las series de índices bursátiles o de precios de derivados, por ejemplo. En ocasiones, sin embargo, algunas series de precios son no estacionarias de orden 2 (tienen 2 raíces unitarias), por lo que incluso las rentabilidades pueden ser no estacionarias, presentando una raíz unitaria.

## 8.7. Contrastes de raíz unitaria

Si utilizamos la teoría de la cointegración, comenzaríamos llevando a cabo contrastes de raíz unitaria para ambas variables, que detectarían en un 95% de las simulaciones que ambas variables son  $I(1)$ .

## 8.8. Cointegración

Un vector  $z$  de variables de naturaleza  $I(1)$  se dicen cointegradas si existe una combinación lineal de las mismas, definida por un vector  $\alpha$  tal que  $\alpha'z$  es una variable aleatoria  $I(0)$ , es decir, estacionaria. Más generalmente, se dice que un vector  $z$  de variables cuyo máximo orden de integración es  $q$  están cointegradas si existe una combinación lineal de las mismas, definida por un vector  $\alpha$  tal que  $\alpha'z$  es una variable aleatoria  $I(p)$ , con  $p < q$ . El vector  $\alpha$  se denomina vector de cointegración.

### 8.8.1. Contraste de cointegración

Si partimos de variables  $y_t, x_t$  de naturaleza  $I(1)$ , sus primeras diferencias,  $\Delta y_t, \Delta x_t$  son estacionarias. Contrastaríamos entonces la cointegración de  $y_t, x_t$  estimando una regresión,

$$y_t = \beta_0 + \beta_1 x_t + v_t, \quad t = 1, 2, \dots, T \quad (8.6)$$

y contrastando la estacionariedad de los residuos, como propusieron Engle y Granger (1987). Sin embargo, los niveles críticos para el contraste de esta hipótesis no son los mismos que para el contraste de raíces unitarias en una variable, pues ahora, el contraste se lleva a cabo después de haber estimado el modelo de regresión (8.6). Esto no es irrelevante: el procedimiento de mínimos cuadrados busca los valores del espacio paramétrico ( $\beta_0$  y  $\beta_1$  en la regresión anterior) que minimizan la varianza del residuo resultante, y éste tiene una varianza infinita para los valores de  $\beta_1$  que no hacen que las variables estén cointegradas. Por tanto, si  $y_t, x_t$  están cointegradas, el método de MCO tenderá a seleccionar el valor de  $\beta_1$  que genera residuos estacionarios, es decir, la constante de cointegración. Aunque esto es lo que pretendemos, ello significa que hay una cierta tendencia a concluir con más frecuencia de la que debiéramos que las variables están cointegradas. En consecuencia, los valores críticos para el contraste de raíz unitaria de los residuos de (8.6) deben ser más elevados en valor absoluto que los utilizados para el contraste de raíz unitaria habitual.

Si los residuos de esta regresión resultan ser estacionarios, decimos que las variables  $y_t, x_t$  están cointegradas, siendo (8.6) la relación de cointegración entre ambas. Esta relación sería el producto  $\alpha'z$  anterior después de normalizar una de las coordenadas del vector  $\alpha$  lo cual, evidentemente, siempre es posible. Se interpreta como la relación de largo plazo entre ellas, alrededor de la cual experimentan desviaciones a corto plazo que revierten posteriormente. Es decir, si en un determinado período,  $y_t$  está por encima del valor numérico de  $\beta_0 + \beta_1 x_t$  para ese mismo período, generalmente  $y_t$  crecerá por encima de  $\beta_0 + \beta_1 \Delta x_t$ , de manera que  $y_{t+1}$  tenderá a acercarse a  $\beta_0 + \beta_1 x_{t+1}$ . En el caso de dos variables  $y_t, x_t$ , decimos que  $\beta_1$  es la constante de cointegración entre ambas.

En el análisis de simulación anterior, en el que generamos ambas series temporales a partir de procesos independientes, este contraste nos sugerirá en una mayoría de simulaciones que  $y_t, x_t$  no están cointegradas, lo que aparecerá en la forma de residuos de naturaleza  $I(1)$  en (8.6). En tal caso, habríamos de estimar un modelo en diferencias de ambas variables,

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + v_t, \quad t = 1, 2, \dots, T$$

que arrojará un coeficiente  $\beta_1$  no significativo y un coeficiente de determinación muy reducido.

Al contrastar cointegración, estamos tratando de detectar la posible existencia de relaciones de largo plazo entre las variables del modelo. En ese sentido, la naturaleza del contraste sugiere el uso de una datos no necesariamente frecuentes, y una muestra temporal suficientemente amplia. De lo contrario, predominará en la muestra, en términos relativos, la información acerca de las fluctuaciones de corto plazo en las variables, frente a la de su evolución tendencial, que es lo que tratamos de detectar. Por tanto, una elección inapropiada de la muestra, ya sea por una frecuencia alta de observación de los datos, o por el uso de un período muestral no muy amplio, sesgará el resultado del contraste hacia la no detección de relaciones de cointegración.

Efectuar un análisis de cointegración significa relacionar los niveles de variables como oferta monetaria y precios, y no sus tasas de variación. Por el contrario, basar la caracterización de la relación entre variables como las citadas utilizando coeficientes de correlación estándar es delicado, pues puede conducir a la detección de regresiones espúreas. El concepto de cointegración generaliza el concepto de correlación en la dirección adecuada. La existencia de una tendencia estocástica común generaría una relación sostenible a largo plazo entre ambas variables, lo que hará que sus diferenciales reviertan a través del tiempo, es decir, que sean mean-

reverting. No tiene sentido analizar las relaciones entre los niveles de variables  $I(1)$  si no están cointegradas.

### 8.8.2. Contraste de hipótesis sobre la relación de cointegración estimada por mínimos cuadrados

Al estimar la relación anterior por mínimos cuadrados hay que tener en cuenta que las propiedades de dicho estimador son válidas únicamente en el caso de variables estacionarias. Cuando las variables están cointegradas, el uso de mínimos cuadrados en la estimación de la regresión está justificado estadísticamente, pero la distribución de probabilidad del estimador MCO no es la habitual. Por tanto, aunque el programa de estimación que utilizemos nos proporcionará las desviaciones típicas estimadas y los ratios tipo-t de cada coeficiente, estos no son válidos en este caso, y no deben utilizarse, por ejemplo, para contrastar hipótesis sobre los coeficientes de la relación de cointegración.

Hay muchos casos, sin embargo, en que el modelo teórico sugiere que las variables  $y_t, x_t$  deben estar relacionadas con un determinado valor numérico del coeficiente, por ejemplo,  $\beta^0 = 1$ , por lo que el investigador estará interesado en contrastar dicha hipótesis. Esto puede hacerse por un procedimiento indirecto, sustituyendo el valor teórico de  $\beta$ ,  $\beta = \beta^0$  en la relación entre ambas variables. Ello significa que construimos la variable auxiliar  $w_t = y_t - \beta^0 x_t$ , y contrastamos la estacionariedad de esta variable.

Cuando se procede de este modo, es importante repetir el contraste para valores de  $\beta^0$  en un entorno de  $\beta^0$ , con el objeto de analizar la precisión con que hemos identificado la constante de cointegración.

### 8.8.3. Correlación y cointegración

Sin embargo, correlación y cointegración no son sinónimos. El problema de correlación espúrea surge entre variables no estacionarias, con independencia de que estén o no cointegradas, luego puede haber alta correlación (de hecho, muy elevada) sin cointegración. Alternativamente, el hecho de que exista una relación de largo plazo entre variables no estacionarias no impide que éstas experimenten desviaciones respecto de la misma que, si son de apreciable magnitud, reducirán la correlación existente entre dichas variables. Un ejemplo sería la evolución temporal de la cotización de un valor en Bolsa, analizada conjuntamente con un índice que lo incluya, ya sea el índice de mercado, un índice de los valores más capitalizados, o un índice sectorial; dado que todo índice es un promedio ponderado

de las cotizaciones de los valores en él incluidos, cabría esperar que ambas series temporales estuvieran correlacionadas. Sin embargo, las fluctuaciones que ambos experimentan a corto plazo pueden ser suficientes para que su coeficiente de correlación sea reducido.

#### 8.8.4. Variables cointegradas: un ejemplo

Un ejemplo típico de variables posiblemente correlacionadas pero habitualmente no cointegradas lo constituye algunos tipos de cambio. A partir de dos variables no estacionarias, pero cointegradas, es sencillo construir dos variables no cointegradas, sin más que añadir en cada período a una de ellas, un incremento no correlacionado temporalmente. Si la varianza de este componente no es muy grande, mantendremos una correlación análoga a la inicial, que podía ser elevada; sin embargo, por construcción, las dos variables no están cointegradas.

Ejemplo de variables cointegradas es,

$$\begin{aligned}x_t &= \alpha_x + \beta_x w_t + \varepsilon_{x_t} \\y_t &= \alpha_y + \beta_y w_t + \varepsilon_{y_t} \\w_t &= w_{t-1} + \varepsilon_t\end{aligned}$$

donde  $z_t$  es la tendencia común a  $x_t$  e  $y_t$ , siendo  $\varepsilon_{x_t}$ ,  $\varepsilon_{y_t}$  variables aleatorias  $N(0, \sigma_x^2)$ ,  $N(0, \sigma_y^2)$ , sin autocorrelación. Las variables  $x_t$  e  $y_t$  están cointegradas, puesto que

$$y_t - \beta_y/\beta_x x_t = \alpha_y - \frac{\beta_y}{\beta_x} \alpha_x + \varepsilon_{y_t} - \frac{\beta_y}{\beta_x} \varepsilon_{x_t}$$

que es una variable estacionaria. El vector  $\begin{pmatrix} 1 \\ -\beta_y/\beta_x \end{pmatrix}$  se denomina vector de cointegración, y la combinación lineal  $\xi_t = y_t - \beta_y/\beta_x x_t$ , que es estacionaria, es la cuantía en la que se incumple la relación de equilibrio a largo plazo en el período  $t$ .

Cuando el vector  $z_t$  consta de  $n$  variables, con  $n > 2$ , pueden existir varias relaciones de cointegración. Esto es lo que sucede, por ejemplo, al considerar un vector de tipos de interés a distinto vencimiento, dentro de un mismo mercado, ya sea el mercado secundario de deuda pública, un mercado de swap en una determinada divisa, etc.. En este caso, el procedimiento de Engle-Granger para estimar vectores de cointegración es problemático, pues estimaremos una combinación lineal



de las posibles relaciones de cointegración existentes entre las variables que componen el vector. De hecho, la estimación resultante dependerá de la normalización de coeficientes utilizada en (8.6), a diferencia de lo que ocurre en el caso de dos variables.

### 8.8.5. El modelo de corrección de error

Teorema de representación de Engle y Granger

Este teorema afirma que si dos variables  $y_t, x_t$  de naturaleza  $I(1)$ , están cointegradas, sus relaciones dinámicas están caracterizadas por el modelo de corrección de error,

$$\begin{aligned} \Delta y_t &= \alpha_y + \sum_{i=1}^n \delta_{1i}^y \Delta x_{t-i} + \sum_{i=1}^n \delta_{2i}^y \Delta y_{t-i} + \gamma_y \xi_{t-1} + \varepsilon_{y_t} \\ \Delta x_t &= \alpha_x + \sum_{i=1}^n \delta_{1i}^x \Delta x_{t-i} + \sum_{i=1}^n \delta_{2i}^x \Delta y_{t-i} + \gamma_x \xi_{t-1} + \varepsilon_{x_t} \end{aligned} \quad (8.7)$$

donde  $\xi_{t-1}$  denota la desviación del período anterior respecto de la relación de equilibrio a largo plazo  $\xi_{t-1} = y_{t-1} - \beta x_{t-1}$ , siendo  $\beta$  el coeficiente de cointegración entre  $y_t$  y  $x_t$ , y  $\Delta$  es el operador de primeras diferencias. En el modelo de corrección de error todas las variables son estacionarias,  $I(0)$ , por lo que las propiedades habituales del estimador MCO en dicho contexto, son válidas. Los términos  $\gamma_y \xi_{t-1}$  y  $\gamma_x \xi_{t-1}$  se denominan términos de corrección de error, y han de aparecer en las ecuaciones anteriores con un determinado signo, que depende del modo en que se haya definido el desequilibrio  $\xi_{t-1}$ . Con nuestra definición, ha de tenerse  $\gamma_y < 0, \gamma_x > 0$ ; un valor negativo de  $\gamma_y$  indicará que períodos en que  $y_t$  es alto, es decir, superior a  $\beta x_t$ , tenderán a venir seguidos de crecimientos relativamente reducidos de dicha variable. Un valor positivo de  $\gamma_x$  indica que siguiendo a períodos en que  $y_t$  es alto,  $x_t$  tenderá a experimentar un crecimiento mayor; la conjunción de ambos efectos hace que  $y_{t+1}$  tienda a aproximarse a  $\beta x_{t+1}$ . Lo dual ocurrirá tras períodos en que  $y_t$  haya sido bajo, es decir, inferior a  $\beta x_t$ . Si hubiéramos normalizado la relación de cointegración de otro modo, habríamos definido el término de desequilibrio como  $\xi_{t-1} = \beta y_{t-1} - x_{t-1}$ , y los signos de los coeficientes  $\gamma_y, \gamma_x$  en (8.7) deberían ser entonces los contrarios a los antes descritos.

Es fácil ver, sin embargo, que esto no es preciso: la aproximación entre ambas variables puede conseguirse asimismo si ambas aumentan o disminuyen simultáneamente, pero  $x_t$  experimenta la mayor variación. Por tanto, si ambos

coeficientes tienen igual signo,  $\gamma_x$  debe ser significativamente mayor que  $\gamma_y$  en valor absoluto. De hecho, podría ocurrir también que sólo uno de los dos coeficientes resulte estadísticamente significativo, lo que podría interpretarse en el sentido de que la variable asociada soporta todo el peso del ajuste hacia la relación de equilibrio a largo plazo.

La cointegración entre variables no lleva añadida ninguna interpretación concreta en términos de causalidad entre dichas variables. De hecho, como la relación de cointegración puede normalizarse de distintas maneras, puede presentarse una apariencia de causalidad en cualquiera de las dos direcciones. El modelo de corrección de error muestra que, en presencia de cointegración, existe importante causalidad entre ambas variables, en principio, con carácter bidireccional. Sólo si algunos de los coeficientes del modelo MCE resultan ser estadísticamente no significativos, podría hablarse de causalidad unidireccional. Si dos variables están cointegradas, al menos una de ellas causa a la otra; sin embargo, ello podría también reflejar el efecto común de una tercera variable, no considerada en el modelo.

Por ejemplo, al trabajar con datos de precios de contado y del futuro sobre un determinado activo financiero, es habitual hallar un mayor número de retardos del precio del futuro en la ecuación del contado, que viceversa, lo que sugiere que los mercados de derivados (en este caso, de futuros), incorporan la nueva información más rápidamente que los mercados de contado, por lo que los últimos parecen responder a fluctuaciones en los primeros. En este tipo de ejemplos, en ocasiones el término de corrección de error resulta no significativo en la ecuación de precios del mercado de contado.

Cuando el vector  $z_t$  incorpora más de dos variables, y existe más de una relación de cointegración entre ellas, el modelo de corrección de error adopta una expresión similar a la antes propuesta. La diferencia estriba en que aparecen retardos de todas las variables, en diferencias, en todas las ecuaciones, y aparecen tantos términos de corrección de error como relaciones de cointegración en cada una de las ecuaciones. Dichos términos serán los valores retardados de dichas relaciones de cointegración; la normalización escogida afecta únicamente a la interpretación de los valores numéricos estimados.

La búsqueda de variables cointegradas abundan en la literatura financiera, donde trata de caracterizarse las posibles relaciones de equilibrio a largo plazo entre precios de activos. Así, se han analizado las posibles relaciones de cointegración entre tipos de cambio, entre tipos de interés dentro de una misma estructura temporal, entre mercados de contado y futuro, entre commodities, valoración

de divisas. También se ha utilizado este tipo de análisis para discutir el grado de integración entre mercados de valores o de deuda, si bien parece existir más evidencia favorable en el primer tipo de mercados. Este análisis tiene asimismo implicaciones para la gestión financiera: en principio, debería ser posible encontrar una cesta reducida de valores cointegrada con el índice, lo que podría utilizarse en la gestión pasiva de carteras. Lo mismo debería ocurrir con un pequeño conjunto de índices sectoriales, etc..

### 8.8.6. El contraste de cointegración de Johansen

Si consideramos un vector autoregresivo  $VAR(p)$ ,

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \dots + A_p y_{t-p} + Bx_t + \epsilon_t$$

donde  $y_t$  es un vector de variables no estacionarias,  $I(1)$ ,  $x_t$  es un vector de variables deterministas, y  $\epsilon_t$  es un vector de innovaciones. El  $VAR(p)$  puede escribirse,

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + Bx_t + \epsilon_t$$

con

$$\Pi = \sum_{i=1}^{p-1} A_i - I, \quad \Gamma_i = \sum_{j=i+1}^p A_j$$

### 8.8.7. Aspectos comunes a varias variables temporales: tendencias comunes, volatilidad común.

### 8.8.8. ¿Qué hacer en presencia de variables con tendencias estocásticas (raíces unitarias)?

De acuerdo con la discusión que hemos llevado a cabo en las secciones anteriores, el procedimiento a seguir en caso de presencia de raíces unitarias en las variables de una regresión lineal simple es claro. Si tanto  $x_t$  como  $y_t$  son variables  $I(1)$ , es decir, tienen una raíz unitaria, entonces el tratamiento que hemos de aplicar depende de si están o no cointegradas. Si lo están, hemos de especificar estimar un modelo de corrección de error. Si no están cointegradas, hemos de estimar un

modelo en diferencias. En el ejercicio de simulación descrito, la estimación de la relación en primeras diferencias,

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_t + v_t, \quad t = 1, 2, \dots, T$$

arrojará un coeficiente  $\beta_1$  no significativo y un coeficiente de determinación muy reducido en la mayoría de las simulaciones.

Esto significa, entre otras cosas, que la recomendación de tratar la no estacionariedad diferenciando las variables, no es correcta. Tal sugerencia es válida cuando, existiendo raíces uniatrías en ambas variables, no están cointegradas. Cuando están cointegradas, el modelo que se estima relaciona las variables en diferencias, pero incorpora asimismo un término de corrección de error.

Aún así, subsisten algunos matices:

- Modelo uniecuacional: como hemos comentado anteriormente, la cointegración entre variables no dice nada acerca de la posible relación de causalidad entre ambas. De hecho, de acuerdo con el teorema de representación de Engle y Granger, el modelo de relación entre ambas variables es un modelo de corrección de error, que es un modelo de dos ecuaciones, una para  $y_t$  en diferencias, y otra para  $x_t$  en diferencias. En ambas aparece el término de corrección de error retardado como una de las variables explicativas, debiendo esperar que tome signo opuesto en cada una de las dos ecuaciones, según como se haya definido dicho término, por las razones antes expuestas. Además de dicho término, aparecerán posiblemente algunos retardos de las diferencias, tanto de  $x_t$  como de  $y_t$ .

Sin embargo, es práctica habitual utilizar tal representación para especificar un modelo de regresión con una única ecuación, como proponen Engle y Granger (19xx). Al actuar así, hemos de interpretar que estamos estimando por separado tan sólo una de las ecuaciones del modelo de corrección de error, lo cual puede hacernos perder eficiencia en la estimación, salvo si: a) las innovaciones en las dos ecuaciones están incorrelacionadas, o b) las dos ecuaciones tuvieran exactamente las mismas variables explicativas.

- ¿Qué diferencias? Ya sabemos que, en caso de cointegración, el modelo a estimar es una relación entre las variables  $x_t$  e  $y_t$  en diferencias. En muchos casos, el investigador dispone de observaciones mensuales o trimestrales de variables como el consumo agregado, el PIB de un país, la inversión, un

agregado monetario, etc. Estas variables tienen, generalmente, una raíz unitaria, por lo que, en caso de querer relacionar dos de ellas, y en presencia de cointegración, deberíamos estimar un modelo de corrección de error.

Sin embargo, no sólo la primera diferencia, es decir, la variación entre meses o trimestres sucesivos,  $y_t - y_{t-1}$ , sino la diferencia anual,  $y_t - y_{t-4}$  en el caso de datos trimestrales, o  $y_t - y_{t-12}$  en el caso de datos anuales, también son variables  $I(0)$ , es decir, estacionarias. Por tanto, el modelo de corrección de error puede especificarse en unas u otras diferencias, siempre que seamos consistentes en tratar tanto  $x_t$  como  $y_t$  de igual manera. Y, sin embargo, las propiedades estadísticas de unas u otras diferencias son bien diferentes; por ejemplo, su volatilidad es muy distinta. Además, es perfectamente concebible que la variación anual (es decir, la tasa interanual) de inflación esté correlacionada con la tasa interanual de crecimiento monetario, a la vez que las tasas de variación intermensuales (es decir, mes a mes) de ambas variables, no muestren una relación significativa. Por consiguiente, no sería lo mismo estimar un modelo de relación,

$$\Delta_{12}y_t \equiv y_t - y_{t-12} = \beta_0 + \beta_1(x_t - x_{t-12}) + \gamma(y_{t-1} - \delta_0 - \delta_1x_{t-1}) + u_t$$

que un modelo,

$$\Delta y_t \equiv y_t - y_{t-1} = \beta_0 + \beta_1(x_t - x_{t-1}) + \gamma(y_{t-1} - \delta_0 - \delta_1x_{t-1}) + u_t$$

de los que no cabe esperar resultados comparables. Tampoco debe pensarse que es ésta una cuestión estadística. Por el contrario, es el propio investigador quien debe decidir si piensa que la relación entre las variables se debe a las fluctuaciones que experimentan en períodos breves de tiempo, como un mes, o en períodos más amplios, como un año.

## **9. Matrices de covarianzas no escalares**

### **9.1. Detección de la autocorrelación**

### **9.2. Tratamiento de la autocorrelación.**

### **9.3. El estimador de mínimos cuadrados generalizados**

### **9.4. Detección de la heteroscedasticidad**

### **9.5. Contraste de igualdad de varianza entre submuestras**

### **9.6. Tratamiento de la heteroscedasticidad**

## **10. El modelo de regresión lineal múltiple**

Aunque hasta ahora hemos considerado únicamente modelos con una sola variable explicativa, no hay ninguna razón para restringirse a tal situación. Además, en la mayoría de las situaciones, el investigador creerá que hay más de una variable que condiciona la evolución del fenómeno que pretende caracterizar. Por ejemplo, es razonable creer que en la determinación de los tipos de interés juega un papel la tasa de crecimiento monetario, pero también la tasa de inflación (o las expectativas de inflación futura), e incluso el nivel de endeudamiento.

Sin embargo, el análisis que hasta ahora hemos presentado es de suma importancia, pues la mayoría de las cuestiones se extienden sin mucha dificultad al caso en que hay varias variables explicativas en el modelo. Hay dos dificultades básicas a que nos enfrentamos al estimar un modelo de regresión múltiple: una es la interpretación de los efectos de una de las variables explicativas separadamente de las demás. Por otro lado, la posibilidad de que la variable endógena se determine simultáneamente con alguna de las variables explicativas es indudablemente mayor cuantas más variables explicativas se incluyen en el modelo; esto sería importante, pues los procedimientos de estimación que hasta ahora hemos examinado no tendrían las propiedades que hemos descrito. Cuando hay determinación simultánea de alguna variable explicativa con la variable dependiente del modelo de regresión uniecuacional, el estimador de mínimos cuadrados es no sólo sesgado, sino inconsistente, y esto aplica a los coeficientes de todas las variables explicativas, no sólo aquella que plantea al problema de determinación simultánea.

La razón por la que trabajamos en muchas ocasiones con modelos de regresión múltiple es sencillamente, porque hay más de una variable con capacidad explicativa significativa sobre la evolución de la variable endógena,  $y_t$ . Si, en tal

situación, sólo explicitamos una de ellas como variables explicativas, las restantes estarán incluidas en el término de error, con lo que éste recogerá, además de otros componentes, la evolución de las variables explicativas omitidas del modelo de regresión. Con ello, su fluctuación será importante, por lo que tendrá una varianza notable. La única manera de reducir dicha varianza del término de error es haciendo explícitas todas las variables potencialmente explicativas.

El investigador nunca tiene convencimiento acerca de la capacidad explicativa de un determinado conjunto de variables. Lo que debe hacer es estimar el modelo con ellas, y proceder a contrastar la significación de cada una de ellas por separado, del modo que describimos en este capítulo. En este proceso incide negativamente la dificultad en estimar por separado el efecto de cada una de las variables explicativas sobre la variable endógena, por lo que no puede sorprender que dediquemos a este asunto una buena parte del capítulo.

Si, queriendo caracterizar la determinación de tipos de interés, estimamos el modelo,

$$r_t = \beta_0 + \beta_1 m_t + \beta_2 \pi_t + u_t \quad (10.1)$$

en el que aparecen el crecimiento monetario y la tasa de inflación como variables explicativas. El coeficiente  $\beta_1$  mide el efecto sobre los tipos de interés de un incremento (o disminución) unitario en la tasa de expansión monetaria, dada una determinada tasa de inflación, es decir, manteniendo la tasa de inflación constante. Aunque indudablemente esta es una evaluación interesante, el hecho de que sólo sea válida en ausencia de variaciones en la tasa de inflación limita algo su uso. En todo caso, es claro que junto con este tipo de estimaciones, nos interesaría disponer asimismo de una estimación del efecto que tendría dicha variación en el crecimiento monetario teniendo en cuenta asimismo el impacto que dicha variación puede tener sobre la tasa de inflación.

Podría pensarse que si lo que se pretende es estimar el impacto que sobre los tipos de interés tiene una variación unitaria en la tasa de crecimiento monetario, podemos estimar el modelo de regresión lineal simple,

$$r_t = \beta_0 + \beta_1 m_t + v_t \quad (10.2)$$

Ahora bien, si el verdadero modelo es (10.1), entonces en el modelo (10.2), la tasa de inflación forma parte del término de error. De hecho, tendríamos la relación entre los términos de error de ambas ecuaciones,

$$v_t = \beta_2 \pi_t + u_t$$

Como consecuencia, en la medida en que crecimiento monetario e inflación no son independientes, la variable explicativa  $m_t$  y el término de error  $v_t$  en el modelo (10.2) estarían correlacionados,

$$\text{Corr}(m_t, v_t) = \beta_2 \text{Corr}(m_t, \pi_t) + \text{Corr}(m_t, u_t) = \beta_2 \text{Corr}(m_t, \pi_t) \neq 0$$

por lo que el estimador de mínimos cuadrados de (10.2) no tendría las propiedades que para él probamos en la sección XX. En particular, sería inconsistente.

En general, un modelo de regresión múltiple incorpora un número  $k$  de variables explicativas, ya sea con datos temporales,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_t, \quad t = 1, 2, \dots, T \quad (10.3)$$

o con datos de sección cruzada,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, 2, \dots, N$$

Es útil que consideremos inicialmente un modelo más simple,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \quad (10.4)$$

que, al igual que hicimos en el modelo de regresión simple, podemos interpretar en diferencias temporales. Para ello, escribimos el modelo en dos instantes de tiempo sucesivos,

$$y_{t-1} = \beta_0 + \beta_1 x_{1t-1} + \beta_2 x_{2t-1} + \beta_3 x_{3t-1} + u_{t-1}$$

y restando, con lo que tenemos,

$$\Delta y_t = \beta_1 \Delta x_{1t} + \beta_2 \Delta x_{2t} + \beta_3 \Delta x_{3t} + \Delta u_t$$

que nos muestra que la fluctuación temporal en  $y_t$  puede explicarse a partir de las variaciones temporales en  $x_{1t}, x_{2t}, x_{3t}$ . Por tanto, una vez más, el modelo es interpretado en términos de variaciones, a pesar de estimarse con datos originales, en niveles de las variables.

Una vez que hayamos estimado el modelo, y disponiendo de datos de las variables incluidas en el mismo, podremos calcular los dos miembros de la igualdad,

$$\Delta y_t = \hat{\beta}_1 \Delta x_{1t} + \hat{\beta}_2 \Delta x_{2t} + \hat{\beta}_3 \Delta x_{3t}$$



que, en realidad, no coincidirán. La diferencia entre la variación en  $y_t$  y la variación que para dicha variable se habría previsto en función de los cambios que han experimentado las variables explicativas,  $\hat{\beta}_1\Delta x_{1t} + \hat{\beta}_2\Delta x_{2t} + \hat{\beta}_3\Delta x_{3t}$ , se debe, por supuesto, a la existencia del error o residuo, que confiamos que, en media, no será muy grande. En todo caso, la diferencia entre los dos miembros de la igualdad anterior será  $\Delta\hat{u}_t$ .

El término  $\hat{\beta}_1\Delta x_{1t}$  mide el efecto que sobre  $y_t$  habría tenido la variación en  $x_{1t}$  si las otras dos variables explicativas no hubiesen cambiado. Sin embargo, este es un supuesto ficticio, pues las tres variables habrán visto alterado su valor numérico.

A pesar de ello, el ejercicio *ceteris paribus* puede tener interés en el diseño de política económica. Por ejemplo, retomando el modelo (10.1), el coeficiente estimado  $\hat{\beta}_2$  nos daría el efecto que sobre los tipos de interés tendrá una variación unitaria (positiva o negativa) en la tasa de inflación, si mantenemos inalterada la tasa de crecimiento monetario. Como ésta es una variable de control de la autoridad monetaria en la puesta en práctica de la política monetaria, el ejercicio que acabamos de describir es razonable. Incluso, una vez realizado, la propia autoridad monetaria podría preguntarse por el impacto que sobre los tipos de interés tendría un incremento de un punto en la tasa de inflación (que tenderá a elevar los tipos de interés), parcialmente compensada con una mayor restricción monetaria (que tenderá a reducir los tipos de interés), por ejemplo, recortando en dos puntos el crecimiento monetario. Dicho impacto sobre los tipos de interés, sería,  $\Delta r_t = 2\hat{\beta}_1 + \hat{\beta}_2$  donde, una vez más, hay que apuntar que, muy probablemente,  $\hat{\beta}_1$  tomará un valor negativo.

Si, por ejemplo, la ecuación estimada es,

$$r_t = 4,25 - 0,42m_t + 0,96\pi_t + \hat{u}_t \quad (10.5)$$

entonces la elevación de un punto en la tasa de inflación tendería a incrementar los tipos de interés en 0,96. Si se reduce en dos puntos el crecimiento monetario el efecto combinado será mucho menor,  $\Delta r_t = 2\hat{\beta}_1 + \hat{\beta}_2 = -0,84 + 0,96 = 0,12$ .

El modelo lineal de regresión especifica una relación del tipo,

$$y_t = \beta_0 + \beta_1x_{1t} + \beta_2x_{2t} + \dots + \beta_kx_{kt} + u_t, \quad t = 1, 2, \dots, T$$

en la que se utilizan  $k$  variables para tratar de explicar el comportamiento de la variable  $y_t$ . Esta última se conoce como variable dependiente, mientras que las variables que aparecen en el miembro derecho del modelo se denominan variables

explicativas. El término  $u_t$  se conoce como la perturbación o término de error del modelo, y es una variable aleatoria. También  $y_t$  se considera que es una variable aleatoria. La variable  $u_t$  es el componente de  $y_t$  que el modelo no puede explicar. Se entiende que la lista de variables explicativas recoge todas las variables que pueden estar relacionadas con  $y_t$ , de modo que  $u_t$  es el componente que el investigador reconoce no poder explicar. En un modelo de regresión interesa que este término sea lo más pequeño posible.

La expresión anterior refleja el supuesto de que disponemos de observaciones de series temporales para cada una de las variables del modelo. Los datos de series temporales recogen información acerca de una determinada unidad económica (un país, un mercado financiero, etc...) en distintos instantes de tiempo, reflejando así la evolución temporal de un conjunto de variables relativas a dicha unidad económica. Un ejemplo sería un modelo que pretende explicar la inflación mensual en la zona euro utilizando como variables explicativas la tasa de crecimiento de la M3, las variaciones en el precio del barril de petróleo, etc.. Otros ejemplos serían: a) tratar de explicar la evolución temporal de las rentabilidades diarias ofrecidas por una cesta IBEX35 utilizando las rentabilidades ofrecidas por una cesta S&P500, b) explicar la evolución de la volatilidad del IBEX35 a partir de la volatilidad del Futuro sobre el IBEX35, c) tratar de explicar la volatilidad implícita en una opción put sobre Telefónica a partir de la volatilidad en la cotización de dicha acción, etc...

En otras ocasiones, los datos disponibles no son de dicho tipo, sino de sección cruzada, es decir, recogen información acerca de distintas unidades estadísticas, en un mismo instante de tiempo,

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, 2, \dots, N$$

aunque ambos modelos se tratan de igual modo.

Estimar el modelo consiste en asignar valores numéricos a los coeficientes  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ . Una vez que se dispone de dichos valores numéricos, puede calcularse el residuo del modelo, mediante la expresión,

$$\hat{u}_t = y_t - \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \dots + \hat{\beta}_k x_{kt} \quad t = 1, 2, \dots, T$$

que tiene la misma naturaleza que los datos utilizados, es decir, será bien una serie temporal o una sección cruzada de datos.

Un buen ajuste consiste en que los residuos sean lo menor posibles, si bien no es evidente cómo medir el tamaño de una variable aleatoria de este tipo.

## 10.1. Estimación por mínimos cuadrados

Los datos disponibles pueden organizarse en la forma de una matriz  $X$ , de dimensión  $T \times k$  o  $N \times k$  para las variables explicativas, y de un vector  $y$ , de dimensión  $T \times 1$  o  $N \times 1$  para la variable dependiente. Asimismo, podemos considerar el vector  $u$ , de dimensión  $T \times 1$  o  $N \times 1$ , que contiene las  $T$  (o  $N$ ) variables aleatorias correspondientes a la perturbación correspondiente a cada observación muestral. Con ellas, el modelo de regresión lineal puede escribirse,

$$y = XB + u$$

El estimador de mínimos cuadrados ordinarios se obtiene mediante la expresión matricial,

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

que, como puede comprobarse, es un vector columna de dimensión  $k \times 1$ . Es, por tanto, una transformación lineal de la variable dependiente  $y$ . En ese sentido, se dice que el estimador MCO es un estimador lineal.

Este estimador proporciona un buen ajuste a los datos, en el sentido de que los residuos que genera proporcionan la menor suma de cuadrados posible. Es conveniente utilizar criterios como el de la suma de los cuadrados de los residuos porque estos pueden ser positivos o negativos, de modo que su suma directa no debe utilizarse como criterio de bondad de ajuste.

En el caso de un modelo de regresión lineal simple,

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, 2, \dots, T$$

este estimador se convierte en,

$$\hat{\beta}_1 = \frac{Cov(x_t, y_t)}{Var(x_t)} = \rho_{xy} \frac{DT(y_t)}{DT(x_t)}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

por lo que el coeficiente de correlación entre  $x_t$  e  $y_t$  está muy relacionado con la estimación de mínimos cuadrados de la pendiente de la recta de regresión. Sin embargo, como puede verse, dicho coeficiente de correlación debe corregirse por la volatilidad relativa de variable dependiente e independiente. La pendiente del modelo mide el efecto que una determinada variación en  $x_t$  tiene sobre  $y_t$ . Si, por ejemplo, la correlación fuese perfecta (supongamos que de signo positivo), pero la variable dependiente fuese el doble de volátil que la variable explicativa, el coeficiente de correlación sería 1, pero el coeficiente del modelo.

## 11. Propiedades del estimador de mínimos cuadrados.

Generalmente, estamos muy interesados en contratar hipótesis de distinto tipo: a) si una variable explicativa contiene información significativa acerca de la variable dependiente, b) si el coeficiente de impacto de una determinada variable es igual a 1, c) si dos variables explicativas tienen el mismo coeficiente, etc...

Sin embargo, aunque los coeficientes del modelo de regresión son constantes, si bien desconocidas, sus estimaciones, por cualquier procedimiento que podamos utilizar, son aleatorias, pues son función de la muestra que utilicemos, que es aleatoria. Si el modelo que estamos estimando es correcto, como hemos de suponer, la perturbación aleatoria del mismo,  $u_t$ , otorga naturaleza asimismo aleatoria a la variable dependiente,  $y_t$ . Esto significa que si cambiamos por ejemplo el período muestral que utilizamos en la estimación, la realización de dicha perturbación, es decir, sus valores numéricos, serán diferentes, con lo que las observaciones de  $y_t$  también los serán, y la estimación de los parámetros diferirá de la obtenida con otro período muestral. Asimismo, si cambiamos la frecuencia de observación de los datos, de diaria a mensual, por ejemplo tomando el último dato de cada mes, la muestra cambia, y con ella, las estimaciones de los coeficientes de las variables explicativas en el modelo.

Siendo variables aleatorias, nos interesa que los estimadores tengan ciertas propiedades deseables, lo cual dependerá del procedimiento de estimación utilizado, y de las características del modelo que estamos estimando. Las principales propiedades en que podemos estar interesados son: insesgo, eficiencia y consistencia.

El insesgo consiste en que la esperanza matemática del estimador coincida con el verdadero valor numérico del coeficiente que estamos estimando. Un estimador eficiente es un estimador de mínima varianza. El procedimiento de mínimos cuadrados proporciona el estimador lineal de mínima varianza, si bien pueden existir otros estimadores no lineales de varianza todavía menor. Un estimador es consistente si, al aumentar el tamaño muestral, converge en probabilidad al verdadero valor del parámetro desconocido que se está estimando. Se dice entonces que su límite en probabilidad es dicho parámetro. Bien podría ocurrir que el estimador fuese sesgado en muestra pequeñas, pero si es consistente, dicho sesgo irá reduciéndose si ampliamos el tamaño muestral.

El estimador de mínimos cuadrados no es siempre consistente. El estimador de máxima verosimilitud lo es, pero siempre que la hipótesis acerca de la distribución de probabilidad en que se basa, sea correcta, sobre lo que no se puede tener

seguridad.

Por construcción, el estimador MCO proporciona aquel conjunto de valores numéricos para los coeficientes del modelo de regresión que generan unos residuos cuya suma de cuadrados es menor. Como este es el criterio seguido para calcular el estimador de mínimos cuadrados, tal propiedad es incuestionable, y caracteriza a dicho estimador.

- Propiedad 1: No puede hallarse otro conjunto de valores numéricos para los coeficientes del modelo que generen unos residuos con una suma de cuadrados inferior a la obtenida a partir del estimador MCO.

Además de esta propiedad el estimador posee otras características, que examinamos a continuación. En todo modelo lineal de regresión, los residuos generados por el estimador de mínimos cuadrados satisfacen las propiedades:

- Propiedad 2: Los residuos de mínimos cuadrados se relacionan con el término de perturbación del modelo mediante,

$$\hat{u} = Mu$$

siendo  $M = I_T - X(X'X)^{-1}X'$ , una matriz cuadrada de orden  $T$ .

- Propiedad 3: La matriz  $M$  es simétrica e idempotente

Demostración.- Inmediata, a partir de la definición de la matriz  $M$ . Por ejemplo,

$$\begin{aligned} M'M &= MM = (I_T - X(X'X)^{-1}X')(I_T - X(X'X)^{-1}X') = \\ &= I_T - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' = I_T - X(X'X)^{-1}X' \end{aligned}$$

Ahora podemos calcular las propiedades estadísticas (esperanza matemática y matriz de covarianzas) del vector de residuos  $\hat{u}$ , de dimensión  $T \times 1$ :

- Propiedad 4: Si el término de perturbación satisface  $E(u) = 0_T$ ,  $Var(u) = \sigma_u^2 I_T$ , se tiene  $E(\hat{u}) = 0$ ,  $Var(\hat{u}) = \sigma_u^2 M$ .

Demostración.- La primera parte es inmediata a partir de la propiedad anterior. Para probar la segunda parte, tenemos,

$$\begin{aligned} \text{Var}(\hat{u}) &= E(\hat{u}\hat{u}') = E(u'MM'u) = E(u'Mu) = \\ &= \text{tr}(u'Mu) = E(uu'M) = E(uu')M = \sigma_u^2 I_T M = \sigma_u^2 M \end{aligned}$$

Por tanto, incluso si el término de perturbación del modelo tiene una estructura de covarianzas sencilla, el vector de residuos tendrá una matriz de covarianzas bastante más compleja. El vector de residuos tiene, al igual que el término de perturbación, esperanza matemática igual a cero para cada observación muestral.

- Propiedad 5: Los residuos *MCO* están incorrelacionados con cada una de las variables explicativas del modelo.

Demostración.- Si denotamos por  $\hat{u}$  el vector de  $T$  residuos del modelo estimado, tenemos

$$X'\hat{u} = X'(y - X\hat{\beta}) = X'(y - X(X'X)^{-1}X'y) = 0$$

Esta propiedad es muy importante si recordamos la interpretación del coeficiente de correlación en el sentido de que una correlación no nula entre variables permite cierta capacidad de predecir el comportamiento de una cualquiera de ellas a partir de los valores observados para la otra. Por tanto, si los residuos del modelo tuvieran correlación positiva o negativa con alguna de las variables explicativas, significaría que la información muestral relativa a dicha variable permite explicar el comportamiento del residuo, la parte de  $y_t$  que hemos dejado sin explicar. Esto definiría la estimación que hemos obtenido como ineficiente, pues no habría hecho uso de toda la información muestral disponible. Por tanto, esta ausencia de correlación es muy deseable.

- Propiedad 6: En todo modelo de regresión que incorpora un término constante, la suma de los residuos generados por el estimador *MCO* es igual a cero. Como consecuencia, su promedio es asimismo nulo.

Demostración.- Es consecuencia de la propiedad anterior, si tenemos en cuenta que el término constante acompaña a una variable explicativa que toma un valor igual a uno en todos los períodos.

- Propiedad 7: El estimador MCO es insesgado

$$\begin{aligned} E(\hat{\beta}) &= E \left[ (X'X)^{-1} X'y \right] = E \left[ (X'X)^{-1} X'(X\beta + u) \right] = \\ &= E \left[ \beta + (X'X)^{-1} X'u \right] = \beta + (X'X)^{-1} X'E(u) = 0 \end{aligned}$$

- Propiedad 8: Si la matriz de covarianzas del término de perturbación es  $Var(u) = \sigma_u^2 I_T$ , la matriz de covarianzas del estimador MCO es:  $Var(\hat{\beta}_{MCO}) = \sigma_u^2 (X'X)^{-1}$ .

$$\begin{aligned} Var(\hat{\beta}) &= E \left[ (\hat{\beta} - E(\hat{\beta})) (\hat{\beta} - E(\hat{\beta}))' \right] = E \left[ (\hat{\beta} - \beta) (\hat{\beta} - \beta)' \right] = \\ &= E \left[ (X'X)^{-1} X'uu'X(X'X)^{-1} \right] = (X'X)^{-1} X'E(uu')X(X'X)^{-1} = \\ &= (X'X)^{-1} X'Var(u)X(X'X)^{-1} = (X'X)^{-1} X'Var(u)X(X'X)^{-1} \end{aligned}$$

- Propiedad 9: Si la matriz de covarianzas del término de perturbación es  $Var(u) = \sigma_u^2 I_T$ , el estimador MCO es el estimador lineal insesgado de menor varianza.

Demostración.- Cualquier otro estimador lineal puede escribirse,  $\tilde{\beta} = Ay$ . Si definimos la diferencia entre las matrices que definen este estimador y el estimador MCO,

$$D = A - (X'X)^{-1} X'$$

tenemos,

$$\tilde{\beta} = D + (X'X)^{-1} X'y = D + (X'X)^{-1} X'(X\beta + u) = DX\beta + \beta + D + (X'X)^{-1} X'u$$

por lo que  $E \tilde{\beta} = DX\beta + \beta$ . El estimador  $\tilde{\beta}$  será insesgado sólo si la matriz  $D$  satisface la propiedad  $DX = 0$ , que equivale a la propiedad  $AX = I_k$ .

Supongamos que se cumple esta propiedad. La matriz de covarianzas del estimador  $\tilde{\beta}$  es entonces,

$$\begin{aligned} \text{Var}(\tilde{\beta}) &= E \begin{bmatrix} \tilde{\beta} - \beta \\ \tilde{\beta} - \beta \end{bmatrix}' = E \begin{bmatrix} D + (X'X)^{-1}X'u \\ (X'X)^{-1}X'u \end{bmatrix}' \\ &= \sigma_u^2(X'X)^{-1} + \sigma_u^2DD' \end{aligned}$$

siendo el segundo sumando una matriz semidefinida positiva. En consecuencia, la diferencia entre las matrices de covarianza del estimador  $\tilde{\beta}$  y el estimador MCO es una matriz definida positiva, lo que prueba que este último es el estimador lineal insegado de mínima varianza.

- Propiedad 10: Si el término de perturbación del modelo de regresión se distribuye:  $u \sim N(0_T, \sigma_u^2 I_T)$ , entonces el estimador MCO del vector de coeficientes  $\beta$  se distribuye,

$$\hat{\beta} \sim N(\beta, \sigma_u^2 (X'X)^{-1})$$

Demostración.- Se basa en el hecho de que el vector que define el estimador de mínimos cuadrados (de dimensión  $k \times 1$ ) es una transformación lineal del vector de perturbaciones (de dimensión  $T \times 1$ )

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

siendo determinista el resto de los elementos que aparecen en esta expresión.

Ello implica que el estimador de cada uno de los coeficientes sigue asimismo una distribución Normal,

$$\hat{\beta}_i \sim N(\beta_i, \sigma_u^2 a^{ii}) \quad 1 \leq i \leq k$$

donde  $a^{ii}$  denota el elemento  $i$ -ésimo en la diagonal de la matriz  $(X'X)^{-1}$ , que tiene dimensión  $k \times k$ .

- Propiedad 12: Si el término de perturbación del modelo de regresión se distribuye:  $u \sim N(0_T, \sigma_u^2 I_T)$ , entonces el estimador MCO del vector de coeficientes  $\beta$  coincide con el estimador de Máxima Verosimilitud de dichos coeficientes.



Demostración.- Recordemos que la función de densidad de una variable Normal multivariante  $\xi \sim N(\mu, \Sigma)$  de dimensión  $T$  es,

$$f(\xi) = \frac{1}{(2\pi)^{T/2}} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} \xi' \Sigma^{-1} \xi \right\}$$

de modo que, si suponemos Normalidad del término de perturbación, tenemos,

$$f(u) = \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma_u^2)^{T/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} u' u \right\}$$

y la verosimilitud del vector de observaciones de la variable dependiente  $y$  resulta,

$$L(y, X/\beta, \sigma_u^2) = \frac{1}{(2\pi)^{T/2}} \frac{1}{(\sigma_u^2)^{T/2}} \exp \left\{ -\frac{1}{2\sigma_u^2} (y - X\beta)' (y - X\beta) \right\}$$

donde hemos utilizado el hecho de que el Jacobiano de la transformación que convierte el vector  $u$  en el vector  $y$  es igual a la matriz identidad de orden  $T$ , por lo que tiene determinante igual a 1.

Por tanto, dada una muestra  $y, X$ , maximizar la función de verosimilitud respecto a los valores paramétricos  $\beta, \sigma_u^2$ , equivale a minimizar la suma de cuadrados de los residuos del modelo.

La única diferencia estriba en que el estimador  $MV$  del parámetro resulta ser,

$$\hat{\sigma}_u^2 = \frac{SR}{T}$$

frente al estimador que suele calcularse en la utilización del procedimiento de mínimos cuadrados, que es,

$$\hat{\sigma}_u^2 = \frac{SR}{T - k}$$

y que suele conocerse como estimador de mínimos cuadrados de  $\sigma_u^2$ .

Ambos se aproximan para tamaños muestrales moderados. El estimador  $MV$  de  $\sigma_u^2$  es sesgado, mientras que el segundo es insesgado. El estimador  $MCO$  del vector de coeficientes  $\beta$  alcanza la cota de Cramer-Rao (umbral inferior para la varianza de todo estimador insesgado), lo que no le sucede al estimador  $MCO$  de  $\sigma_u^2$ , que excede de dicha cota, si bien no existe ningún estimador insesgado de  $\sigma_u^2$ .

que alcance dicha cota. El estimador  $MV$  de  $\sigma_u^2$  es inferior a dicha cota, pero es sesgado, como ya hemos dicho.

Este resultado es importante, por cuanto que implica que si, además de los supuestos más básicos, de tener esperanza cero y matriz de covarianzas escalar, añadimos el supuesto de que el término de perturbación sigue una distribución Normal, entonces el estimador MCO tiene las mismas propiedades que el estimador MV. Ahora bien, sabemos que, bajo condiciones bastante generales, el estimador MV es eficiente; es decir, no puede encontrarse otro estimador, lineal o no lineal, que tenga una matriz de covarianzas menor que la del estimador MV. En términos del concepto de precisión que introdujimos en la Sección XX, no puede encontrarse un estimador de mayor precisión que el estimador MV.

En consecuencia, la utilización del procedimiento de mínimos cuadrados está totalmente justificada bajo el supuesto de Normalidad del término de perturbación, y algo menos cuando tal supuesto no se establece. Evidentemente, no se trata de que el investigador haga o no una hipótesis acerca del tipo de distribución que sigue el término de perturbación, sino de que se preocupe acerca de si el supuesto de Normalidad es aceptable. Por eso es que los contrastes de normalidad deben tener una cierta importancia en el análisis empírico de regresión sería igual a 2.

Si las variables explicativas no son deterministas, sino aleatorias, como cabe esperar, entonces el cálculo de las propiedades del estimador de mínimos cuadrados es algo más complejo, y sus propiedades varían. En general, el estimador será sesgado, pero será consistente, salvo que alguna de las variables explicativas tenga un comportamiento tendencial. Cuando aparecen tendencias, el estimador de mínimos cuadrados es asintóticamente insesgado y eficiente,

$$\lim_{T \rightarrow \infty} E(\hat{\beta}) = \beta; \quad \lim_{T \rightarrow \infty} Var(\hat{\beta}) = 0$$

La fuerte estructura que hemos impuesto sobre el vector de perturbaciones,  $Var(u) = \sigma_u^2 I_T$  no es necesaria para garantizar la consistencia del estimador de mínimos cuadrados.

De hecho, incluso en presencia de variables explicativas estocásticas, el estimador de mínimos cuadrados todavía es insesgado si las variables explicativas están incorrelacionadas con el término de perturbación del modelo de regresión. En realidad, todo lo que necesitamos es que la esperanza del término de perturbación, condicional en la información proporcionada por las variables explicativas, sea cero:  $E(u/X) = 0$ . Esto implica la ausencia de correlación, puesto que,

$$E(X'u) = E[E(X'u/X)] = E[EX'(u/X)] = E[X'E(u/X)] = E(0) = 0$$

En tales condiciones, tomando esperanzas condicionales, tenemos,

$$\begin{aligned} E(\hat{\beta}/X) &= E_{\mathbb{E}}[(X'X)^{-1}X'y/X] = E_{\mathbb{E}}[(X'X)^{-1}X'(X\beta + u)/X] = \\ &= E[\beta + (X'X)^{-1}X'u/X] = \beta + (X'X)^{-1}X'E(u/X) = 0 \end{aligned}$$

Sin embargo, si bien será insesgado, el estimador *MCO* no será ya eficiente.

La consistencia del estimador *MCO* requiere que las variables explicativas estén incorrelacionadas asintóticamente (es decir, en el límite al aumentar el tamaño muestral) con el término de perturbación,

$$p \lim \frac{1}{T} X'u = 0$$

Como

$$p \lim \hat{\beta} = \beta + p \lim \left[ \frac{X'X}{T} \right]^{-1} p \lim \frac{X'u}{T}$$

y si se existe el límite  $p \lim \left[ \frac{X'X}{T} \right]^{-1}$  y la matriz  $X'X$  es invertible, entonces un argumento de continuidad garantiza la existencia de  $p \lim \left[ \frac{X'X}{T} \right]^{-1}$ . En tal caso,  $p \lim \frac{1}{T} X'u = 0$  garantiza la consistencia del estimador *MCO*, puesto que se tiene,

$$p \lim \hat{\beta} = \beta$$

Dado que la matriz  $\frac{X'X}{T}$  se compone de momentos muestrales de orden 2, la existencia del límite de dicha matriz requiere existencia de segundos momentos para el vector de variables explicativas. La presencia de una tendencia, aun siendo determinista, genera términos del tipo,

$$\frac{\sum_{t=1}^T t^2}{T}; \quad \frac{\sum_{t=1}^T tx_{it}}{T}$$

cuyo límite no existirá si las restantes variables  $x_{it}$  tienen un comportamiento estable alrededor de un nivel de referencia.

En presencia de variables explicativas estocásticas, el estimador *MCO* no tiene distribución Normal en muestras finitas incluso si  $u \sim N(0_T, \sigma_u^2 I_T)$ . Por el contrario, bajo ciertas condiciones de regularidad, se tiene,

$$\sqrt{T}(\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma_u^2 (X'X)^{-1})$$

mediante aplicación del teorema central del límite al caso en que el término de perturbación presenta la misma distribución de probabilidad para todas las observaciones, con esperanza nula y varianza finita, y las variables explicativas satisfacen condiciones de existencia de momentos como la que antes discutimos.

## 12. Bondad de ajuste del modelo

Al tener los residuos media cero, es claro que debemos preocuparnos por el modo en que oscilan alrededor de cero. En particular, lo que nos interesa es que las desviaciones que experimentan con respecto a cero sean lo menor posibles. Es decir, que su varianza sea lo menor posible, pero su varianza muestral es, precisamente, proporcional a  $\sum_{t=1}^T \hat{u}_t^2$ . Por otra parte, puesto que el residuo no es sino una componente de la variable dependiente, aquella que no podemos explicar, tiene perfecto sentido comparar la varianza de los residuos con la de la variable dependiente. Ambas son positivas, y su cociente será necesariamente inferior a la unidad, de modo que oscila entre 0 y 1. Pero esto es lo que hacemos con el coeficiente de determinación.

La bondad del ajuste del modelo de regresión se representa por el coeficiente de determinación del modelo,

$$R^2 = 1 - \frac{SR}{ST} = 1 - \frac{\sum_{t=1}^T \hat{u}_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

donde *SR* denota la suma de los cuadrados de los residuos, y *ST* lo que conocemos como Suma Total, la suma de las desviaciones al cuadrado de la variable dependiente respecto de su media muestral. Esta suma es igual a la varianza de  $y_t$  multiplicada por el tamaño muestral.

## 13. Contrastes de hipótesis

Generalmente, estamos muy interesados en contratar hipótesis de distinto tipo: a) si una variable explicativa contiene información significativa acerca de la variable

dependiente, b) si el coeficiente de impacto de una determinada variable es igual a 1, c) si dos variables explicativas tienen el mismo coeficiente, etc...

Para llevar a cabo contrastes de este tipo, necesitamos hacer alguna hipótesis acerca de la distribución de probabilidad del término de error o perturbación del modelo de regresión. Generalmente suponemos que dicho término sigue una distribución Normal, si bien esto debe contrastarse utilizando los tests paramétricos o no paramétricos apropiados. Como hemos visto en la sección anterior, w bajo ciertas condiciones tenemos,

$$\hat{\beta} \sim N(\beta, \sigma_u^2 (X'X)^{-1})$$

en cuyo caso, cada uno de los coeficientes del modelo sigue asimismo una distribución Normal,

$$\hat{\beta}_i \sim N(\beta_i, \sigma_u^2 a^{ii}) \quad 1 \leq i \leq k$$

donde  $a^{ii}$  denota el elemento  $i$ -ésimo en la diagonal de la matriz  $(X'X)^{-1}$ , que tiene dimensión  $k \times k$ .

Si queremos contrastar una determinada hipótesis acerca del valor numérico del coeficiente asociado a una determinada variable  $x_i, 1 \leq i \leq k$ ,

$$H_0 : \beta_i = \beta_i^0$$

podríamos utilizar esta propiedad, pues tendremos que,

$$\frac{\hat{\beta}_i - \beta_i^0}{\sigma_u^2 a^{ii}} \sim N(0, 1)$$

de modo que bastaría fijar un nivel de significación para el contraste, obtener el nivel crítico correspondiente al mismo en la tabla de una Normal(0,1), y comparar el valor numérico del estadístico  $\frac{\hat{\beta}_i - \beta_i^0}{\sqrt{\sigma_u^2 a^{ii}}}$  con dicho umbral crítico. Por supuesto que el nivel crítico que obtengamos en la tabla de la  $N(0, 1)$  debe depender de que el contraste sea de una o de dos colas, es decir, de que la hipótesis alternativa sea del tipo,

$$H_1 : \beta_i \neq \beta_i^0$$

o de alguno de los tipos,

$$H_1 : \beta_i < \beta_i^0$$

ó,

$$H_1 : \beta_i > \beta_i^0$$

La hipótesis de significación de la variable  $x_i$ ,  $1 \leq i \leq k$  consiste en contrastar

$$H_0 : \beta_i = 0$$

frente a una alternativa,

$$H_1 : \beta_i \neq 0$$

aunque también podría ser de una sóla cola, si tenemos información a priori restringiendo el signo de dicho coeficiente.

Sin embargo, el contraste de hipótesis no puede llevarse a cabo de este modo porque desconocemos el valor numérico de la varianza del término de error. Puede sin embargo estimarse, lo cual hacemos mediante la expresión,

$$\hat{\sigma}_u^2 = \frac{SR}{T - k}$$

Ahora bien, si sustituimos su valor teórico por su valor estimado, las propiedades estadísticas del contraste de hipótesis cambian. Concretamente, habremos de utilizar la propiedad,

$$(T - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi_{T-k}$$

independiente de  $\hat{\beta}$ . Como consecuencia, tenemos que,

$$\frac{\frac{\hat{\beta}_i - \beta_i^0}{\sqrt{\sigma_u^2 a^{ii}}}}{(T - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} / (T - k)} \sim t_{T-k}$$

es decir,

$$\frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_u^2 a^{ii}} \sim t_{T-k}$$

con lo que podremos llevar a cabo contrastes como los mencionados.

Específicamente al llevar a cabo contrastes de significación, conviene que distinguir entre las situaciones en que no rechazamos la hipótesis nula de ausencia de significación de una determinada variable porque efectivamente, no tiene capacidad explicativa, de aquellos casos en que no rechazamos la hipótesis nula porque la estimación del coeficiente asociado a dicha variable se lleva a cabo no una precisión reducida. Baja precisión implica una varianza elevada para el estimador de dicho coeficiente, con lo que el valor numérico del ratio que aparece en el estadístico  $t$  de Student será reducido, y posiblemente inferior al nivel crítico proporcionado por las tablas. En definitiva, hay que distinguir entre los casos en que dicho ratio es pequeño porque el numerador es pequeño, de los casos en que el ratio es reducido porque su denominador es muy elevado.

Asimismo, conviene recordar que para rechazar una hipótesis nula, requerimos que la información muestral contenga evidencia significativa en contra de la hipótesis nula y favorable a la hipótesis alternativa. Esta segunda condición suele olvidarse con demasiada frecuencia, pero es importante, especialmente en contrastes de una cola. En el contraste de hipótesis

$$H_0 : \beta_i = 0$$

frente a la alternativa,

$$H_1 : \beta_i > 0$$

si obtenemos una estimación puntual  $\hat{\beta}_i = -3,5$  estaríamos en una situación en que la evidencia muestral es contraria a la hipótesis nula, pero también es contraria a la hipótesis alternativa. En tal caso, los métodos estadísticos habituales para el contraste de hipótesis nos llevarán a no rechazar la hipótesis nula, a pesar de que la estimación numérica del coeficiente  $\beta$  puede considerarse elevada. Ello se debe a que al establecer la hipótesis alternativa, no hemos considerado la posibilidad de que dicho coeficiente tome valores negativos, seguramente por algún conocimiento previo o alguna razón teórica. No es que nosotros queramos o no rechazar  $H_0$ , sino que los procedimientos habituales nos llevarán a no rechazar dicha hipótesis. En una situación así, el investigador debería cuestionar las razones que le han llevado a establecer una hipótesis alternativa como la que hemos presentado. Si continúa pensando que tal alternativa es razonable, deberá desechar la muestra que ha utilizado; por el contrario, el resultado de la estimación podría en algunos casos reconsiderar la hipótesis alternativa, estableciéndola en la forma  $H_1 : \beta_i \neq 0$ , y volver a contrastar la hipótesis nula  $H_0 : \beta_i = 0$  de nuevo.

Para llevar a cabo el contraste de hipótesis de una cola

$$H_0 : \beta_i = \beta_i^0$$

frente a la alternativa,

$$H_1 : \beta_i > \beta_i^0$$

utilizaríamos el hecho de que para toda variable  $t_n$  se tiene,

$$0,95 = P[\xi \leq t_{n,\alpha}]$$

siendo  $t_{n,\alpha}$  el nivel crítico proporcionado por las tablas de la distribución  $t_n$  al nivel de confianza  $\alpha$ , que habremos de fijar previamente.

Por tanto, bajo el supuesto de que la hipótesis nula es cierta

$$0,95 = P \left[ \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_u^2 a^{ii}} \leq t_{n,\alpha} \right] = P \left[ \hat{\beta}_i \leq \beta_i^0 + t_{n,\alpha} \hat{\sigma}_u^2 a^{ii} \right]$$

de modo que los valores admisibles del coeficiente están por debajo de  $\beta_i^0 + t_{n,\alpha} \hat{\sigma}_u^2 a^{ii}$ . Esto delimita una región crítica  $RC \equiv \beta_i^0 + t_{n,\alpha} \hat{\sigma}_u^2 a^{ii}; +\infty$  cuyo umbral inferior es superior al valor teórico,  $\beta_i^0$ . Es decir, rechazamos la hipótesis nula si nuestra estimación, con la muestra disponible, excede de dicho umbral, que es a su vez mayor que el valor teórico  $\beta_i^0$ .

En el caso particular de que contrastemos si un coeficiente es cero, cuando la alternativa contemplada es que tome valores únicamente positivos, rechazaremos la hipótesis nula cuando la estimación puntual de dicho coeficiente exceda de un cierto umbral estrictamente positivo,  $t_{n,\alpha} \hat{\sigma}_u^2 a^{ii}$ . Este tipo de evidencia sería simultáneamente contraria a la hipótesis nula, y favorable a la hipótesis alternativa.

Para contrastar hipótesis más complejas, como

$$H_0 : \beta_1 + \beta_2 = 5$$

frente a una alternativa como  $H_1 : \beta_1 + \beta_2 < 5$ , basta tratar  $\beta_1 + \beta_2 - 5$  como una nueva variable aleatoria  $z$  cuya varianza puede determinarse sin ningún problema a partir de la matriz de varianzas-covarianzas de  $\hat{\beta}$ , y pensar en contrastar  $H_0 : z = 0$ , frente a la alternativa  $H_1 : z < 0$ . Así,



$$\frac{z}{DT(z)} \sim N(0, 1)$$

y si sustituimos en esta expresión el verdadero valor de  $\sigma_u^2$  por su estimación, tendremos  $\frac{z-5}{DT(z)} \sim t_{T-k}$  y podremos proceder a contrastar  $H_0$ .

Lo que estamos haciendo es comparar la magnitud de la holgura o incumplimiento de una restricción, utilizando su desviación típica como unidad de medida, para poder decidir si dicha holgura es grande o pequeña.

Este argumento puede extender al contraste simultáneo de varias restricciones. Calculamos para cada una de ellas su holgura, obteniendo así un vector de dimensión igual al número de restricciones, cuyas coordenadas pueden ser positivas o negativas. Igualmente, podemos pensar en obtener la matriz de covarianzas de las restricciones, una vez que tratamos cada una de ellas como una sola variable aleatoria, como hemos hecho en el ejemplo anterior. Finalmente, decidimos si el vector de holguras es grande o pequeña calculando su tamaño utilizando la matriz de covarianzas como unidad de medida,

$$(\text{Vector de holguras}) [\text{Var}(\text{Vector de holguras})]^{-1} (\text{Vector de holguras})'$$

que puede probarse que sigue una distribución  $F_{q,T-k}$ , siendo  $q$  el número de restricciones que se contrastan.

Por ejemplo, si contrastamos las hipótesis...

Para contrastar un conjunto de hipótesis lineales de la forma,

$$H_0 : R\beta = r$$

frente a la alternativa,

$$H_1 : R\beta \neq r$$

el estadístico,

$$\frac{(\hat{R}\hat{\beta} - r)' [R(X'X)^{-1}R']^{-1} (\hat{R}\hat{\beta} - r) / q}{\hat{u}'\hat{u} / (T - k)} \sim F_{q,T-k}$$

Utilizar este estadístico matricial es equivalente al ejercicio que hemos propuesto arriba. De hecho, no es difícil ver en el corchete central de este estadístico

tipo  $F$  la matriz de covarianzas del vector de holguras, donde el parámetro  $\sigma_u^2$  se ha sustituido por su estimación en el denominador. Estamos, por tanto, calculando el tamaño del vector de holguras, utilizando su desviación típica como unidad de medida.

Por último, en ocasiones pueden sustituirse en el modelo las regresiones que se pretende calcular. En ese caso, otro modo equivalente de calcular el estadístico  $F$  anterior consiste en comparar la suma de cuadrados de los residuos que se obtienen en el modelo sin restringir y el modelo restringido, mediante el estadístico,

$$\frac{(SRR - SSR) / q}{SSR / (T - k)}$$

que es idéntico al anterior y obedece, por tanto, a una distribución de probabilidad  $F_{q, T-k}$ .

Otros estadísticos son,

$$(T - k) \frac{(SRR - SSR)}{SSR} \text{ Wald}$$

$$(T - k + q) \frac{(SRR - SSR)}{SSR} \text{ Multiplicadores de Lagrange}$$

$$T (\ln SRR - \ln SSR) \text{ Razón de verosimilitudes}$$

## 14. Matrices de covarianzas no escalares

Si el término de perturbación satisface las condiciones,  $E(u_T) = 0_T$ ,  $Var(u_T) : \sigma_u^2 \Sigma$ , siendo  $\Sigma$  una matriz simétrica, definida positiva,  $T \times T$ , el estimador  $MCO$  es insesgado, con matriz de covarianzas :  $Var(\hat{\beta}_{MCO}) = \sigma_u^2 (X'X)^{-1} (X'\Sigma X) (X'X)^{-1}$ . La demostración de esta propiedad es totalmente análoga a la de las Propiedades 7 y 8.

Si el término de perturbación del modelo se distribuye  $N(0_T, \sigma_u^2 \Sigma)$  siendo  $\Sigma$  una matriz simétrica, definida positiva,  $T \times T$ , entonces el estimador  $MCO$  del vector de coeficientes  $\beta$  se distribuye,

$$\hat{\beta}_{MCO} \sim N(\beta, \sigma_u^2 (X'X)^{-1} (X'\Sigma X) (X'X)^{-1})$$

- La demostración es análoga a la realizada en la Propiedad 10, basándose el resultado en el hecho de que, cuando las variables explicativas son deterministas, el estimador  $MCO$  es una transformación lineal del vector  $u$ .

En estas circunstancias, el estimador *MCO* ya no es el estimador lineal insesgado de mínima varianza.

#### 14.1. Comparación de estimadores de la regresión múltiple y la regresión simple

Consideremos la estimación por mínimos cuadrados del modelo,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

supongamos que estimamos una regresión auxiliar,

$$x_{2t} = \delta_0 + \delta_1 x_{1t} + \varepsilon_{2t} \quad (14.1)$$

y construimos los residuos,

$$\hat{\varepsilon}_{2t} = x_{2t} - \hat{\delta}_0 - \hat{\delta}_1 x_{1t}, \quad (14.2)$$

que, como vimos en XX, tendrán, entre otras, la propiedad  $Corr(\hat{\varepsilon}_{2t}, x_{1t}) = 0$ .

Si, a continuación, estimamos el modelo de regresión simple de la variable dependiente de interés sobre los residuos de esta regresión auxiliar,

$$y_t = \alpha_0 + \alpha_2 \hat{\varepsilon}_{2t} + v_t$$

tendremos (como vimos en XX),

$$\hat{\alpha}_2 = \frac{\sum_{t=1}^T \hat{\varepsilon}_{2t} y_t}{\sum_{t=1}^T \hat{\varepsilon}_{2t}^2}$$

Resulta sorprendente que, como puede probarse sin mucha dificultad, este estimador coincide con el estimador de mínimos cuadrados ordinarios de  $\beta_2$  en (10.4),

$$\hat{\beta}_2^{MCO} = \hat{\alpha}_2$$

Lo que hemos hecho es extraer de  $x_{2t}$  el efecto de  $x_{1t}$  debido a la correlación que, en general, existirá entre estas variables. Así, el residuo  $\hat{\varepsilon}_{2t}$  mide el componente de  $x_{2t}$  que no tiene nada en común con  $x_{1t}$  y, de hecho, tiene correlación nula con esta variable.

Este resultado se extiende al modelo más general, (10.3), en el que si estimamos la regresión auxiliar,

$$x_{1t} = \alpha_0 + \alpha_2 x_{2t} + \dots + \alpha_k x_{kt} + \varepsilon_{1t}, \quad t = 1, 2, \dots, T \quad (14.3)$$

y utilizamos los residuos de esta regresión

$$\hat{\varepsilon}_{1t} = x_{1t} - \hat{\alpha}_0 - \hat{\alpha}_2 x_{2t} - \dots - \hat{\alpha}_k x_{kt}, \quad t = 1, 2, \dots, T$$

para estimar la regresión simple,

$$y_t = \alpha_0 + \alpha_1 \hat{\varepsilon}_{1t} + v_t \quad (14.4)$$

tedremos que la estimación de mínimos cuadrados de  $\alpha_1$  en (14.4) coincidirá con la estimación de mínimos cuadrados de  $\beta_1$  en (10.3).

Supongamos por un momento que en un determinado análisis, la variable explicativa  $x_{1t}$  tiene correlación cero con el resto de las variables explicativas. Entonces, en (14.3) los coeficientes estimados deberían ser prácticamente cero, al igual que el coeficiente de determinación de dicha ecuación. En tal caso, el residuo  $\hat{\varepsilon}_{1t}$  será prácticamente igual a  $x_{1t}$ ,  $\hat{\varepsilon}_{1t} \simeq x_{1t}$ , por lo que el modelo (14.4) será esencialmente,

$$y_t = \alpha_0 + \alpha_1 x_{1t} + v_t \quad (14.5)$$

Pero hemos dicho que la estimación de mínimos cuadrados de  $\alpha_1$  en esta regresión coincide con la estimación de mínimos cuadrados de  $\beta_1$  en (10.3). Por tanto, cuando una variable explicativa está incorrelacionada con todas las demás, su coeficiente puede estimarse igualmente bien en la regresión múltiple completa, o en la regresión simple de  $y_t$  sobre esta única variable.

Una segunda situación en que se produce este mismo resultado es cuando la correlación muestral entre cada una de las variables explicativas e  $y_t$  es cero. La razón es que en tal caso, las estimaciones teóricas de los coeficientes asociados a cada una de estas variables sería cero, por lo que estimar el modelo (10.3) equivale, al menos teóricamente, a estimar (14.5).

Es interesante saber que neste resultado puede asimismo aplicarse a bloques de variables. Así, supongamos que en el modelo de regresión múltiple, las  $k$  variables explicativas pueden agruparse en dos bloques, con la condición de que ninguna variable de un bloque está correlacionada con ninguna variable del otro bloque. Permitimos, sin embargo, que variables de un mismo bloque estén correlacionadas,

positiva o negativamente entre sí. Pues bien, en tal situación, los coeficientes asociados a las variables en el primer bloque pueden estimarse en una regresión de  $y_t$  únicamente sobre este subgrupo de variables explicativas, y lo mismo puede decirse acerca de los coeficientes asociados a las variables en el segundo bloque.

## 14.2. Regresión particionada

## 15. Grado de ajuste del modelo de regresión múltiple

Una vez estimados los coeficientes del modelo de regresión múltiple, y obtenidos los valores numéricos de los residuos, la varianza residual se obtiene,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

$$\hat{\sigma}_u^2 = \frac{1}{n-3} \sum_{i=1}^n \hat{u}_i^2 = \frac{S_u^2}{n-3} = \frac{S_{y.123}^2}{n-3}$$

donde la notación que hemos introducido en la última igualdad hace referencia a que se trata de la suma residual que resulta al estimar una regresión en que la variable dependiente  $Y$  aparece explicada por  $x_1, x_2, x_3$ . La Suma Residual puede expresarse,

$$S_{y.123}^2 = \sum_{i=1}^n y_i^2 - \hat{\beta}_0 \sum_{i=1}^n y_i - \hat{\beta}_1 \sum_{i=1}^n y_i x_{1i} - \hat{\beta}_2 \sum_{i=1}^n y_i x_{2i} - \hat{\beta}_3 \sum_{i=1}^n y_i x_{3i} \quad (15.1)$$

expresión que puede expresarse matricialmente,

$$S_{y.123}^2 = y'y - \hat{\beta}X'y$$

y que puede descomponerse en la forma,

$$S_y^2 = S_{y.123}^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (15.2)$$

de modo que la Suma Total, o suma de cuadrados de las desviaciones entre las observaciones de la variable dependiente y su media, es igual a la suma de la Suma Residual de la regresión, más la Suma Explicada por la misma.

El coeficiente de determinación múltiple se define,

$$R_{y.123}^2 = 1 - \frac{S_{y.123}^2}{S_y^2}$$

que puede obtenerse sin necesidad de calcular previamente los residuos de la regresión pues, utilizando (15.1) junto con  $S_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2$  se llega a,

$$R_{y.123}^2 = \frac{\hat{\beta}_1 S_{x_1 y} + \hat{\beta}_2 S_{x_2 y} + \hat{\beta}_3 S_{x_3 y}}{S_y^2}$$

Por la descomposición de la Suma Residual (15.2), tenemos que el coeficiente de determinación  $R_{y.123}^2$  es positivo, y no superior a la unidad.

El coeficiente de correlación múltiple es la raíz cuadrada, con signo positivo, del coeficiente de determinación múltiple,

$$\rho_{y.123}^2 = \frac{S_{y.123}^2}{S_y^2} = \frac{S_y^2 - S_{y.123}^2}{S_y^2}$$

### 15.1. Coeficientes de correlación parcial y de determinación parcial

El modelo de regresión múltiple nos permite considerar asimismo la capacidad explicativa que cada una de las variables independientes por separado, tiene sobre la variable dependiente, lo cual es de sumo interés para el investigador. Para ello utilizamos los coeficientes de correlación parcial,

- Definición.- El coeficiente de correlación parcial entre  $Y$  y  $X_1$ , denotado por  $\rho_{y1.2}$ , en el universo de variables  $Y, X_1, X_2, X_3$ , es el coeficiente de correlación simple entre las variables  $Y$  y  $X_1$ , una vez que se ha extraído de ellas la influencia común que puedan tener de la variable  $X_2$ .

El coeficiente de correlación parcial entre  $Y$  y  $X_2$ , denotado por  $\rho_{y2.1}$ , en el universo de variables  $Y, X_1, X_2, X_3$ , se definiría análogamente. Suele interpretarse el coeficiente  $\rho_{y1.2}$  como el grado de correlación existente entre  $Y$  y  $X_1$ , cuando  $X_2$  se mantiene fija, pero consideramos bastante más adecuada la interretación que hemos dado: dicho coeficiente mide la correlación existente entre las variables  $Y$  y  $X_1$  que no es debida a la posible influencia común que ambas variables puedan

experimentar respecto de  $X_2$ . Cuando hay más de dos variables explicativas, las posibilidades de definir coeficientes de correlación parcial se multiplican.

En el caso de dos variables explicativas,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

si denotamos por  $\rho_{y1}, \rho_{y2}, \rho_{12}$  los coeficientes de correlación simple entre cada par de variables, puede probarse que,

$$\rho_{y1.2} = \frac{\rho_{y1} - \rho_{y2}\rho_{12}}{1 - \rho_{y2}^2 (1 - \rho_{12}^2)}$$

$$\rho_{y2.1} = \frac{\rho_{y2} - \rho_{y1}\rho_{12}}{1 - \rho_{y1}^2 (1 - \rho_{12}^2)}$$

Los coeficientes de correlación parcial pueden escribirse asimismo en términos de varianzas muestrales. Por ejemplo, el coeficiente de correlación parcial entre  $Y$  y  $X_2$  es igual a,

$$\rho_{y2.1} = \frac{S_{y.1} - S_{y.12}^2}{S_{y.1}^2} = \frac{S_{y.1}^2 - S_{y.12}^2}{S_{y.1}^2}$$

que depende de la comparación entre dos sumas de cuadrados de residuos: la que procede de la regresión múltiple de  $Y$  sobre  $X_1$  y  $X_2$ , y la correspondiente a la regresión simple de  $Y$  sobre  $X_1$ . El coeficiente de correlación parcial entre  $Y$  y  $X_2$  es igual a la reducción que la varianza del modelo de regresión simple se obtiene cuando se añade al mismo, como variable explicativa adicional, la variable  $X_2$ . Es claro que  $S_{y.1}^2 > S_{y.12}^2$ , pues la suma residual nunca disminuye al añadir una variable explicativa al modelo de regresión. Si, por ejemplo,  $X_2$  no aporta explicación alguna sobre  $Y$  que no esté ya contenida en  $X_1$ , entonces  $S_{y.1}^2 = S_{y.12}^2$  y  $\rho_{y2.1} = 0$ , a pesar de que  $\rho_{y2}$  habrá sido, en general, diferente de cero.

Analogamente, tendríamos, el coeficiente de correlación parcial entre  $Y$  y  $X_1$ ,

$$\rho_{y1.2} = \frac{S_{y.2} - S_{y.12}^2}{S_{y.2}^2} = \frac{S_{y.2}^2 - S_{y.12}^2}{S_{y.2}^2}$$

Sus cuadrados son los coeficientes de determinación parcial,

$$R_{y1.2}^2 = 1 - \frac{S_{y.12}^2}{S_{y.2}^2} = \frac{S_{y.2}^2 - S_{y.12}^2}{S_{y.2}^2}$$

$$R_{y2.1}^2 = 1 - \frac{S_{y.12}^2}{S_{y.1}^2} = \frac{S_{y.1}^2 - S_{y.12}^2}{S_{y.1}^2}$$

## 16. Colinealidad entre variables explicativas en un modelo de regresión

En un modelo de regresión lineal múltiple, la interpretación de los coeficientes estimados no es inmediata. De hecho, lo verdaderamente importante es entender que hay diversas maneras de interpretar los valores numéricos obtenidos en el proceso de estimación, no todas equivalentes.

La lectura más inmediata de la estimación de un modelo de regresión, utilizada con excesiva frecuencia, consiste en interpretar cada coeficiente como el impacto que la variable explicativa asociada al mismo tiene sobre la variable dependiente. Lamentablemente, una interpretación tan directa no es siempre válida. Es fácil ver que el valor numérico de un coeficiente como  $\beta_2$  en la regresión,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

es la variación, positiva o negativa, según su signo, que experimentaría la variable  $y_t$  si la variable  $x_{3t}$  aumentase en una unidad, y supuesto que ninguna otra de las variables explicativas, alterase su valor numérico.

Ahora bien, recordemos la interpretación que llevamos a cabo acerca del significado del coeficiente de correlación en la Sección XX. Si  $x_{3t}$  estuviera altamente correlacionada con  $x_{2t}$ ,  $\rho(x_{2t}, x_{3t}) = -0,85$ , por ejemplo, sería poco realista hacer el supuesto ceteris paribus del párrafo anterior, pues el valor numérico del coeficiente de correlación entre ambas variables indica que cuando una de ellas aumenta, la otra generalmente disminuye. Por tanto, para calcular el efecto que sobre  $y_t$  tendría un aumento de una unidad en  $x_{3t}$ , habría que restar de  $\beta_3$ , que mide el efecto directo, el producto de  $\beta_2$  por el cambio producido en  $x_{2t}$ . Este producto mediría el efecto indirecto.

Para calcular la magnitud de la variación en  $x_{2t}$  que cabría esperar que viniera asociada con un aumento de una unidad en  $x_{3t}$ , hemos de apelar nuevamente al concepto de coeficiente de correlación,



$$\rho(x_{3t}, x_{3t}) = \frac{\sigma_{23}}{\sigma_2\sigma_3}$$

### 16.1. Efectos de la colinealidad entre variables explicativas

No se producen efectos sobre las características globales de la regresión: R-cuadrado, residuos numéricos, desviación típica de la perturbación, estadístico de significación global de la regresión, etc.

Consideremos un modelo de regresión con las variables medidas en desviaciones respecto a su valor medio,

$$\tilde{y}_t = \beta_1 \tilde{x}_{1t} + \beta_2 \tilde{x}_{2t} + u_t$$

A partir de la matriz de covarianzas muestral entre las variables explicativas del modelo de regresión,

$$Var(X) = \begin{pmatrix} var(x_1) & cov(x_1, x_2) \\ cov(x_1, x_2) & var(x_2) \end{pmatrix} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \quad a, c > 0, \quad b \in \mathbb{R}$$

obtenemos la matriz de covarianzas de los estimadores MCO del modelo de regresión múltiple,

$$Var(\hat{\beta}_{MCO}) = \sigma^2 \begin{pmatrix} a & b \\ b & c \end{pmatrix}^{-1} = \sigma^2 \begin{pmatrix} \frac{c}{ac-b^2} & -\frac{b}{ac-b^2} \\ -\frac{b}{ac-b^2} & \frac{a}{ac-b^2} \end{pmatrix} = \sigma^2 \frac{1}{ac-b^2} \begin{pmatrix} c & -b \\ -b & a \end{pmatrix}$$

1. Que las variables explicativas del modelo de regresión estén altamente correlacionadas significa que, en valor absoluto, su coeficiente de correlación (covarianza dividido entre producto de desviaciones típicas) sea próximo a la unidad o, lo que es equivalente, que  $\frac{b^2}{ac}$  esté próximo a la unidad. Esto implica, evidentemente, que  $b^2$  es aproximadamente igual al producto  $ac$ .

Una primera consecuencia es que el determinante de la matriz de covarianzas de las variables explicativas,  $ac - b^2$ , será próximo a cero, es decir, dicha matriz de covarianzas es próxima a ser singular. Como el inverso de dicho determinante aparece como un factor común en  $Var(\hat{\beta}_{MCO})$ , se tiene que tanto las varianzas de los estimadores MCO como sus covarianzas, son numéricamente elevadas. Por ejemplo, con una matriz de covarianzas entre variables,

$$Var(X) = \begin{pmatrix} 4 & 6.2 \\ 6.2 & 10 \end{pmatrix}, \quad Corr(x_1, x_2) = \frac{6.2}{\sqrt{(4)(10)}} = .98031$$

tenemos una matriz de covarianzas entre estimadores MCO,

$$Var(\hat{\beta}_{MCO}) = 3.0 \begin{pmatrix} 4 & 6.2 \\ 6.2 & 10 \end{pmatrix}^{-1} = \begin{pmatrix} 19.231 & -11.923 \\ -11.923 & 7.6923 \end{pmatrix}$$

con,

$$DT(\hat{\beta}_1) = \sqrt{19.231} = 4.3853, \quad DT(\hat{\beta}_2) = \sqrt{7.6923} = 2.7735$$

1. Por tanto, se produce una pérdida de precisión en las estimaciones, al aumentar las desviaciones típicas de los estimadores MCO. En este caso, por tanto, estimamos con una baja precisión no porque existan variaciones en el valor de los coeficientes, sino porque las variables asociadas están altamente correlacionadas, positiva o negativamente.
2. Como discutimos en la Sección XX al hablar de los contrastes de hipótesis estadísticas, la pérdida de precisión tiene implicaciones importantes en cuanto a efectuar contrastes acerca de valores numéricos con dichos coeficientes, por cuanto que la menor precisión, que es equivalente a una mayor varianza estimada termina plasmándose en intervalos de confianza más amplios. Esto hace mucho más fácil que el valor numérico que se contrasta en una hipótesis nula simple caiga dentro de dicho intervalo, no rechazándose, en consecuencia, dicha hipótesis nula. En este caso, que una hipótesis nula no se rechace no se debe necesariamente a que el valor hipotético del coeficiente sea muy similar al valor estimado, sino a que la desviación típica asociada a este último es muy grande, lo que puede ampliar el intervalo de confianza asociado de manera muy importante. En definitiva, como vimos en su momento, la menor precisión en la estimación viene asociada a una pérdida de potencia en la contrastación de hipótesis estadísticas.
3. Otra consecuencia de la correlación entre variables explicativas es que la correlación entre coeficientes estimados,  $-\frac{b}{\sqrt{ac}}$  tendrá signo contrario al de la correlación entre las variables explicativas  $x_1, x_2$ , y será cercano a 1 en valor absoluto. Si las variables explicativas están positivamente correlacionadas, entonces los estimadores MCO de sus coeficientes asociados están negativamente correlacionados. En el ejemplo anterior,

$$Corr \hat{\beta}_1, \hat{\beta}_2 = \frac{-11.923}{(19.231)(7.6923)} = -.98029$$

Ello quiere decir que, mientras que, a lo largo de la muestra, las dos variables tienden a estar simultáneamente por encima o por debajo de sus respectivas medias muestrales, las estimaciones numéricas de sus coeficientes estarán una por encima de su media, y otra por debajo. Esto significa que el procedimiento tiende a sobre-estimar uno de dichos coeficientes y a sub-estimar el otro, sin que, por supuesto, sepamos cuál de ellos está en cada situación. Pero además, que su correlación sea elevada implica que tendremos una seria dificultad en distinguir entre una colección de posibles pares de estimaciones  $\hat{\beta}_1, \hat{\beta}_2$ , que se distinguen unas de otras porque las que tienen un valor mayor de  $\hat{\beta}_1$ , tienen un valor inferior de  $\hat{\beta}_2$ . Este es un problema que se conoce como de falta de identificación en la estimación del modelo econométrico.

Como hemos visto, el efecto de la colinealidad es aumentar las varianzas de los estimadores de mínimos cuadrados, así como sus covarianzas. En general, las correlaciones entre coeficientes son asimismo elevadas. Esto significa no sólo que exista una tendencia a que cada coeficiente individual tienda a aparecer como estadísticamente no significativo, sino que, además, es difícil estimar numéricamente cada coeficiente por separado. Es lógico esperar que así sea: si dos variables  $x_2$  y  $x_3$  están positivamente correlacionadas, entonces las dos tienden a desviarse en igual dirección respecto a sus medias muestrales. Por tanto, ambas tienden a estar simultáneamente por encima o por debajo de sus respectivas medias; por consiguiente, la misma capacidad explicativa genera la combinación  $\beta_2 x_2 + \beta_3 x_3$  que la combinación  $(\beta_2 + \theta)x_2 + (\beta_3 - \phi)x_3$ . Es decir, puesto que ambas variables se mueven generalmente en igual sentido, existen proporciones  $\theta$  y  $\phi$  de ambas variables que toman valores numéricos muy similares, compensándose entre sí. Por consiguiente, ambas combinaciones tomarán aproximadamente los mismos valores numéricos. Como esta sustitución entre las variables  $x_2$  y  $x_3$  puede llevarse a cabo en cualquier cuantía, siempre que se respete la proporción dada por  $\theta/\phi$ , identificar con precisión por separado los valores numéricos de los parámetros  $\beta_2$  y  $\beta_3$  resulta muy difícil.

## 16.2. Detección de la colinealidad

Puesto que la colinealidad se refiere a la presencia de coeficientes de correlación de elevados en magnitud entre las variables explicativas, nada mejor para detectar esta situación que examinar dichos coeficientes de correlación en la muestra. Aunque parezca sorprendente, hay distintos procedimientos que deben seguirse, pues son todos ellos de una enorme sencillez.

1. En primer lugar, el cálculo de dichos coeficientes de correlación entre todos los pares de variables explicativas; en el modelo del ejemplo anterior, hay tan sólo uno de tales pares.
2. En segundo lugar, deben examinarse las nubes de puntos entre tales pares de variables; como dijimos en su momento, un gráfico tan simple como una nube de puntos nos proporciona una perspectiva muestral completa, a diferencia del único valor numérico proporcionado por un coeficiente de correlación muestral. Por ejemplo, en una nube de puntos podemos percibir que lo que aparece como un coeficiente de correlación muestral sólo moderado, se debe a la existencia de una submuestra, de reducida magnitud, en la que las dos variables en cuestión se desconectan, existiendo en el resto de la muestra una relación estrecha entre ambas.
3. En tercer lugar, pueden calcularse los valores propios de la matriz de covarianzas de las variables observadas. Como hemos visto en el ejemplo anterior, en presencia de colinealidad, dicha matriz estará cercana a la singularidad, lo que se ha de reflejar en que el menor de los valores propios debe ser muy inferior al mayor de los mismos. Ha de tenerse en cuenta que todos los valores propios de dicha matriz serán no negativos, como corresponde a una matriz de covarianzas, que es semi-definida positiva. Por otra parte, el determinante de una matriz cuadrada es igual al producto de sus valores propios, y ha de ser cercano a cero si la matriz es casi singular. Para que el producto de los valores propios sea cercano a cero, es necesario que se cumpla la relación citada entre el menor y el mayor de todos ellos.
4. Por último, pueden estimarse regresiones parciales entre pares de variables explicativas. El R-cuadrado de dichas regresiones será el coeficiente de correlación entre ambas, al cuadrado, por lo que la estimación de estas regresiones auxiliares engloban como caso particular a la primera de nuestras sugerencias. De este modo, obtenemos información adicional útil, especialmente

a través de los residuos de cada regresión, cuyo examen nos proporciona el componente de una de las variables explicativas que no está explicado por la utilizada como variable explicativa en la regresión auxiliar.

5. Sensibilidad a variaciones en un número reducido de observaciones muestrales.

### 16.3. Tratamiento de la colinealidad

Como hemos visto, uno de los dos problemas producidos por la colinealidad entre las variables explicativas del modelo de regresión estriba en la dificultad de interpretar separadamente cada uno de los coeficientes estimados. Concretamente, no podemos interpretar el valor numérico de un coeficiente estimado como el efecto de la variable explicativa asociada sobre la variable dependiente.

#### 16.3.1. Regresión ortogonalizada

Puesto que una de las dificultades generadas por la colinealidad entre variables explicativas estriba en la dificultad de interpretar los valores numéricos estimados para los coeficientes del modelo de regresión, uno de los posibles tratamientos de la colinealidad consiste en modificar las variables explicativas del modelo, de modo que las nuevas variables sean incorrelacionadas entre sí.

Para ello, el investigador debe comenzar estableciendo un ranking de importancia entre las variables explicativas originales, entendiendo por ello, su creencia acerca de la capacidad explicativa individual de cada una de ellas respecto a la variable dependiente. La primera de dichas variables en importancia se conserva inalterada. A continuación, se estima una regresión lineal de la variable que ocupa el segundo puesto en el ranking, sobre la primera,

$$x_{2t} = \delta_0 + \delta_1 x_{1t} + \varepsilon_{2t} \quad (16.1)$$

los residuos de dicha regresión van a sustituir a la segunda variable. Dichos residuos son el componente de  $x_{2t}$  que no está explicada por  $x_{1t}$ ; además, por construcción, están incorrelacionados con  $x_{1t}$ ,  $\rho(\varepsilon_{2t}, x_{1t}) = 0$ . A continuación, se estima una regresión de la tercera variable sobre las dos primeras; por construcción, los residuos de mínimos cuadrados de dicha regresión tienen correlación nula con  $x_{1t}$  y con  $x_{2t}$ , y se procede de este modo hasta llegar a la variable colocada en el último lugar del ranking.

Para establecer el ranking inicial, el investigador puede servirse de regresiones parciales de la variable dependiente sobre cada una de las variables explicativas originales. El  $R^2$  de dichas regresiones, las desviaciones típicas residuales, así como un examen gráfico de los residuos de cada una de ellas, será generalmente suficiente para establecer dicho ranking. La única dificultad en este análisis surge si hay duda entre dos variables acerca de cuál de ellas asignar alguno de los primeros lugares del ranking; esto sería importante, pues si dos variables tienen una capacidad explicativa similar, elevada en ambos casos, entonces la variable que resultase menos favorecida en la selección sería sustituida por los residuos en una regresión de dicha variable por la que se seleccionó como prioritaria en el ranking, y dichos residuos tendrán una capacidad explicativa muy reducida, debido a que estas dos variables eran originalmente muy similares.

En otras palabras, si dudamos entre seleccionar  $x_{1t}$  o  $x_{2t}$  para ocupar el primer lugar en el ranking.....

De este modo, alteramos la regresión inicial por otra,

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 \varepsilon_{2t} + \beta_3 \varepsilon_{3t} + u_t \quad (16.2)$$

donde  $\varepsilon_{2t}$  y  $\varepsilon_{3t}$  denotan los residuos de las regresiones auxiliares, (16.1) y

$$x_{3t} = \varphi_0 + \varphi_1 x_{1t} + \varphi_2 x_{2t} + \varepsilon_{3t}, \quad (16.3)$$

$$\hat{\varepsilon}_{2t} = x_{2t} - \hat{\delta}_0 + \hat{\delta}_1 x_{1t}, \quad (16.4)$$

$$\hat{\varepsilon}_{3t} = x_{3t} - \hat{\varphi}_0 - \hat{\varphi}_1 x_{1t} + \hat{\varphi}_2 x_{2t}, \quad (16.5)$$

En esta regresión, el coeficiente de correlación entre dos variables explicativas cualesquiera es cero. Por ejemplo, si consideramos  $\varepsilon_{2t}$  y  $\varepsilon_{3t}$ , tenemos que  $\varepsilon_{3t}$  tiene correlación nula con  $x_{1t}$  y con  $x_{2t}$ , mientras que  $\varepsilon_{2t}$  es una combinación lineal de  $x_{1t}$  y  $x_{2t}$ ; en consecuencia, el coeficiente de correlación entre  $\varepsilon_{2t}$  y  $\varepsilon_{3t}$  es igual a cero. Por tanto, a diferencia de lo que ocurría en la regresión original, los coeficientes estimados en (16.2) pueden interpretarse como el impacto, positivo o negativo, que se tendría sobre  $y_t$  si la variable asociada, digamos que  $\varepsilon_{3t}$ , aumenta en una unidad.

La estrategia propuesta habría resuelto, por tanto, el problema de interpretación de los coeficientes individuales estimados; sin embargo, surge otra dificultad evidente, y es que el investigador no está interesado en estimar el impacto que sobre  $y_t$  puede tener una variación unitaria en  $\varepsilon_{3t}$ , sino en la variable original,  $x_{3t}$ .

Sin embargo, no todo está perdido: supongamos que, una vez estimada (16.2) el investigador dispone de datos  $y_t, x_{1t}, x_{2t}$  y  $x_{3t}$ . Sustituiría,

$$x_{2t} = \delta_0 + \delta_1 x_{1t} + \varepsilon_{2t} \quad (16.6)$$

$$\varepsilon_{3t} = x_{3t} - (\hat{\varphi}_0 + \hat{\varphi}_1 x_{1t} + \hat{\varphi}_2 x_{2t}) \quad (16.7)$$

en (16.2), obteniendo XXX.

### 16.3.2. Otros tratamientos

Regresión cresta

Componentes principales

## 17. Predicción

Como mínimo, una variable puede predecirse a partir de su valor medio muestral.

## **18. Modelos univariantes de series temporales**

### **18.1. Primeros conceptos**

#### **18.1.1. Procesos estocásticos**

#### **18.1.2. Funciones de autocorrelación simple y parcial**

### **18.2. Procesos autoregresivos, $AR(p)$**

#### **18.2.1. El modelo $AR(1)$**

#### **18.2.2. El modelo $AR(2)$**

### **18.3. Procesos de medias móviles, $MA(q)$**

### **18.4. Procesos mixtos, $ARMA(p,q)$**

### **18.5. Procesos integrados $ARIMA(p,d,q)$**

### **18.6. Predicción con modelos univariantes**

#### **18.6.1. Predicción con modelos $AR(p)$**

#### **18.6.2. Predicción con modelos $MA(q)$**

#### **18.6.3. Predicción con modelos $ARMA(p,q)$**

#### **18.6.4. Predicción con modelos $ARIMA(p,d,q)$**

### **18.7. Estimación de modelos univariantes de series temporales**

#### **18.7.1. Estimación de modelos autoregresivos**

#### **18.7.2. Estimación de modelos de medias móviles**

#### **18.7.3. Estimación de modelos $ARMA(p,q)$ y $ARIMA(p,d,q)$**

## **19. El procedimiento de variables instrumentales**

### **19.1. Correlación entre variables explicativas y término de error**

### **19.2. Errores de medida**

## **20. Modelos dinámicos**

### **20.1. Colinealidad entre variables explicativas**

### **20.2. Estimación**

88

#### **20.2.1. Perturbación sin autocorrelación**

#### **20.2.2. Perturbación con autocorrelación**

## **21. Simultaneidad**

### **21.1. Identificación**

### **21.2. Estimación de una ecuación del sistema**