

**TEMARIO QUINTO EJERCICIO**

**PARTE B:**

**ECONOMÍA CUANTITATIVA**

**[versión 12/2023 – Convocatoria OEP 2022]**

Víctor Gutiérrez Marcos

Parte B) Economía cuantitativa

Tema 1. Métodos de regresión: lineal, por mínimos cuadrados ordinarios y método de máxima verosimilitud.

Tema 2. El modelo de regresión con regresores estocásticos: estimación, verificación y predicción.

Tema 3. Autocorrelación, multicolinealidad y heterocedasticidad: consecuencias, origen, detección y soluciones.

Tema 4. Modelos de ecuaciones simultáneas: identificación, estimación, verificación y predicción.

Tema 5. Modelos uniecuacionales de regresión no lineal: estimación, verificación y predicción.

Tema 6. Modelos ARIMA: Identificación, estimación, verificación y predicción.

Tema 7. El modelo de regresión dinámica uniecuacional: identificación, estimación, verificación y predicción.

Tema 8. Modelos de componentes no observables: tendencia, ciclo y estacionalidad. Aplicaciones en el análisis de coyuntura, estimación, verificación y predicción.



## 5.B.1 : MÉTODOS DE REGRESIÓN: LINEAL, POR MÍNIMOS CUADRADOS ORDINARIOS Y MÉTODO DE MÁXIMA VEROSIMILITUD.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del *método deductivo contrastable*<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

▪ **Contextualización:**

- Desde un punto de vista histórico, desde las primeras aportaciones de GAUSS a principios del s. XIX, el modelo más utilizado y estudiado por los econométricos ha sido el modelo lineal general (o modelo de regresión clásico).
  - Este modelo ha sido aplicado con gran éxito en el estudio de un elevado número de temas económicos. Además, sirve de base para el desarrollo de modelos más complejos que levantan los supuestos clásicos, que pueden ser restrictivos en determinadas aplicaciones (series temporales, relaciones no lineales...).
  - El MLG puede estimarse a través de varios métodos, entre los que destacaremos los 2 métodos más importantes:
    - El método de mínimos cuadrados ordinarios (MCO); y
    - El método de máxima verosimilitud (MV).

▪ **Problemática (Preguntas clave):**

—

▪ **Estructura:**

**0. ASPECTOS GENERALES SOBRE LOS MODELOS ECONOMÉTRICOS**

0.1. *Modelo*

0.2. *Muestra*

**1. MODELO LINEAL GENERAL**

1.1. *Supuestos clásicos del modelo lineal general*

- 1.1.1. Supuesto 1: Linealidad del modelo
- 1.1.2. Supuesto 2: Rango completo (no multicolinealidad perfecta)
- 1.1.3. Supuesto 3: Esperanza condicionada nula de las perturbaciones
- 1.1.4. Supuesto 4: Perturbaciones esféricas
- 1.1.5. Supuesto 5: Regresores deterministas
- 1.1.6. Supuesto 6: Distribución normal de los errores

1.2. *Métodos de regresión a partir de la información muestral*

**2. MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS**

2.1. *Método de estimación (problema de optimización)*

2.2. *Propiedades estadísticas del estimador de MCO*

- 2.2.1. Propiedades para muestras finitas
  - Esperanza del estimador: insesgadez
  - Varianza del estimador: eficiencia ( $T^{\text{ma}}$  de Gauss-Markov)
- 2.2.2. Normalidad y distribución de  $\beta$
- 2.2.3. Propiedades asintóticas: consistencia y distribución

**3. MÉTODO DE MÁXIMA VEROSIMILITUD**

3.1. *Método de estimación (problema de optimización)*

3.2. *Propiedades estadísticas del estimador de MV*



## 0. ASPECTOS GENERALES SOBRE LOS MODELOS ECONÓMÉTRICOS

### 0.1. Modelo

- Los **modelos econométricos** quedan expresados genéricamente del siguiente modo:

$$y = f(x_1, x_2, \dots, x_k, \varepsilon)$$

donde,

- $y$ : Variable endógena, dependiente o explicada.
- $x_1, \dots, x_k$ : Variables explicativas, independientes o regresores.
- $\varepsilon$ : Término de error del modelo, también conocido como perturbación aleatoria (dado que “perturba” lo que en su ausencia sería una relación determinística estable).
  - Las perturbaciones surgen por motivos diferentes. El más importante, porque no podemos esperar capturar todas las influencias sobre una variable económica en un modelo. El efecto de estos factores omitidos queda capturado por el error.

- Esta ecuación trata de **explicar el comportamiento de la variable económica  $y$**  a través de la información proporcionada por un conjunto de  $k$  variables explicativas  $x$  con un claro significado económico, así como por una variable aleatoria  $\varepsilon$  (no observable  $y$ , por consiguiente, sin significado conceptual económico).

### 0.2. Muestra

- Con el objetivo de evaluar empíricamente el modelo, se recoge información muestral, es decir, una lista de valores numéricos de las variables  $(y, x_1, x_2, \dots, x_k)$ . En Econometría aplicada, se trabaja con **3 tipos de datos muestrales**:
  - Datos de sección cruzada: Conjunto de datos recogidos observando diversas unidades económicas (familias, empresas, ciudades...) en un mismo instante de tiempo o haciendo abstracción de la dimensión temporal.
  - Datos de series temporales: Conjunto ordenado de datos sobre una unidad económica a lo largo del tiempo.
  - Datos de panel (combinación de los 2 anteriores): Datos de sección cruzada observados a lo largo de varios periodos de tiempo.
- A lo largo de esta exposición, por sencillez expositiva, nos centraremos en el caso de datos de sección cruzada, por lo que denotaremos las observaciones de cada individuo por el subíndice  $i$ . Dispondremos pues de una serie de  $N$  relaciones como las siguientes:

$$y_i = f(x_{1i}, \dots, x_{ki}, \varepsilon_i) \quad \forall i = 1, \dots, N$$

## 1. MODELO LINEAL GENERAL

### 1.1. Supuestos clásicos del modelo lineal general

- El **Modelo Lineal General (MLG)** se basa en un conjunto de supuestos sobre cómo los datos muestrales serán generados en un proceso subyacente de generación. Como veremos, bajo estos supuestos del MLG, los estimadores cumplen una serie de “buenas propiedades” si se cumplen los supuestos clásicos enunciados originalmente por GAUSS a principios del s. XIX.
  - Se utilizará como referencia de “*Econometric Analysis*” realizada por GREENE para la explicación.

#### 1.1.1. Supuesto 1: Linealidad del modelo

- El modelo que relaciona la variable endógena con las variables explicativas es **lineal en los coeficientes  $\beta$** .
  - Por tanto, el modelo se expresa del siguiente modo:
 
$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ki}, \varepsilon_i) = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_k \cdot x_{ki} + \varepsilon_i \quad \forall i = 1, \dots, N$$
 donde,
    - $\beta_0$  es el término independiente o *intercepto*. El término independiente puede interpretarse como la esperanza de la variable dependiente cuando el resto de las variables es nulo o

puede interpretarse aquel que acompaña a una primera variable explicativa  $x_{0i}$  cuyo valor es siempre igual a 1.

- $\beta_1, \dots, \beta_k$  son los coeficientes o *pendientes* de la regresión. Recogen la magnitud del impacto de cada una de las variables explicativas sobre la variable endógena.

– Utilizando notación matricial:

- Para cada individuo:  $y_i = x_{i1 \times K} \cdot \beta_{K \times 1} + \varepsilon_i \quad \forall i = 1, \dots, N$
- Para la totalidad de la muestra:  $y_{N \times 1} = X_{N \times K} \cdot \beta_{K \times 1} + \varepsilon_{N \times 1}$

▪ **Limitaciones del supuesto:** Suponer que las relaciones económicas tienen un carácter lineal puede parecer a priori muy limitante, pues por la teoría económica, sabemos que existen numerosas relaciones entre variables económicas que no son lineales. No obstante, este supuesto no es tan restrictivo como cabría pensar en un primer momento debido a:

- *Transformaciones de las funciones:* Un gran número de relaciones que, en principio, no son lineales, pueden llegar a serlo si se aplican simples transformaciones. Por ejemplo, la función de producción Cobb-Douglas que relaciona producción y factores productivos, puede ser linealizada tomando logaritmos en ambos lados de la ecuación:

$$Y = A \cdot K^\alpha \cdot L^\beta \cdot e^\gamma \rightarrow \ln(Y) = \alpha \cdot \ln(K) + \beta \cdot \ln(L) + \underbrace{\varepsilon}_{\varepsilon = \gamma = \gamma \cdot \ln(e)}_{=1}$$

- *Aproximaciones de Taylor:* Las series de Taylor permiten realizar aproximaciones de funciones no lineales a través del sumatorio de una serie de potencias denominados términos de la serie. Cada uno de los términos se calcula a partir de las derivadas de distinto orden de la función correspondiente en un punto determinado. De esta forma, conforme la serie contenga mayor número de términos, más precisa será la aproximación realizada.

$$f(x) \approx f(x_0) + \frac{f'(x_0)}{1!} \cdot (x - x_0) + \frac{f''(x_0)}{2!} \cdot (x - x_0)^2 + \dots + \frac{f^{(N)}(x_0)}{N!} \cdot (x - x_0)^N$$

$$f(x) \approx \sum_{n=0}^N \frac{f^{(n)}(x_0)}{n!} \cdot (x - x_0)^n$$

A la serie de Taylor centrada sobre el punto cero ( $x_0 = 0$ ) se la denomina serie de Maclaurin. Se constata entonces, que en el caso de que en un modelo econométrico se quiera explicar una variable  $y$  a través de una variable  $x$  con la que a priori no se puede establecer una relación lineal en los coeficientes de regresión, mediante una aproximación de Taylor, se puede conseguir tal relación lineal, siempre y cuando el punto elegido sea suficientemente derivable, y siempre y cuando las relaciones que se vayan a estudiar sean marginales en torno al punto de aproximación  $x_0$ , donde con el modelo econométrico anteriormente expresado:

- $\frac{f^{(n)}(x_0)}{n!}$  representará el coeficiente de regresión.
- $(x - x_0)^n$  **representa la variable explicativa**. Nótese que en los modelos de regresión lineal, la clave es la linealidad de los coeficientes  $\beta$  y no de la variable explicativa (que puede estar elevada a cualquier orden).

El uso de series de Taylor como aproximación lineal cuenta con 3 ventajas importantes:

- La derivación e integración de estas series se puede realizar término a término, lo que resultan operaciones triviales.
- Se puede utilizar para calcular valores aproximados de funciones.
- Es posible calcular la optimalidad de la aproximación.

### 1.1.2. Supuesto 2: Rango completo (no multicolinealidad perfecta)

- $X$  es una matriz de dimensión  $N \times K$  con rango  $K$ , por lo que es una matriz con rango completo. Esto requiere que se cumplan 2 condiciones:
  - Las columnas de  $X$  son linealmente independientes, lo que implica que las variables explicativas son linealmente independientes, es decir, ninguna variable explicativa es combinación lineal de otra.
  - Condición de identificación: Disponemos de al menos  $K$  observaciones ( $N \geq K$ )<sup>3</sup>.

### 1.1.3. Supuesto 3: Esperanza condicionada nula de las perturbaciones

- Para todo el conjunto de observaciones, se asume que los términos de error tienen esperanza condicionada a  $X$  nula:

$$E[\varepsilon|X] = \begin{bmatrix} E[\varepsilon_1|X] \\ E[\varepsilon_2|X] \\ \vdots \\ E[\varepsilon_N|X] \end{bmatrix} = 0$$

- Ello implica que la esperanza de cada error, condicionado a la información contenida en el vector  $x$  es nula, lo que significa que ninguna observación de  $x$  provee información sobre el valor esperado del error (son ortogonales).
- Por la ley de expectativas iteradas se obtiene que  $E[\varepsilon_i|X] = E_X[E[\varepsilon_i|X]] = E_X[0] = 0 \forall i$  esto implica<sup>4</sup> que la media no condicionada del término error también es nula. Por otro lado, el supuesto también implica lo siguiente:  $Cov[X_i, \varepsilon_i] = 0 \quad \forall i = 1, \dots, N$
- **Limitación del supuesto:** Este último supuesto “débil”, conocido como exogeneidad de las variables explicativas, garantiza la consistencia del método MCO. Sin embargo, su incumplimiento en casos como la presencia de variables omitidas, errores de medida o presencia de retardos de la variable endógena como variable explicativa, puede generar problemas de endogeneidad y requiere un tratamiento especial que queda fuera del alcance de este tema [ver tema 5.B.4].

### 1.1.4. Supuesto 4: Perturbaciones esféricas

- El supuesto de perturbaciones esféricas queda resumido en la siguiente forma de la matriz de varianzas-covarianzas de los términos de error (matriz escalar):

$$E[\varepsilon \cdot \varepsilon'|X] = \sigma^2 \cdot I = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{bmatrix}$$

- Este supuesto contiene, por tanto, 2 sub-supuestos:
  1. Homocedasticidad: Todos los términos de error tienen la misma varianza:
$$Var[\varepsilon_i|X] = \sigma^2 \quad \forall i = 1, \dots, N$$
  2. No autocorrelación: No existe correlación serial entre las perturbaciones:
$$Cov[\varepsilon_i, \varepsilon_j|X] = 0 \quad \forall i \neq j$$

- Todo ello se resume en el siguiente supuesto:

$$E[\varepsilon \cdot \varepsilon'|X] = \sigma^2 \cdot I \rightarrow Var[\varepsilon] = E[Var[\varepsilon|X]] + Var[E[\varepsilon|X]] = \sigma^2 \cdot I$$

<sup>3</sup> Si el nº de observaciones fuera menor que  $K$ ,  $X$  no podría tener rango completo. En principio, basta con que  $N = K$  para que  $X$  pueda llegar a tener rango  $K$ . Sin embargo, cuanto mayor sea  $N$ , más precisas tenderán a ser nuestras estimaciones. La diferencia  $(N - K)$  se conoce como “grados de libertad” de la regresión.

<sup>4</sup> La relación no es cierta en el sentido contrario.

### 1.1.5. Supuesto 5: Regresores deterministas

- Un supuesto habitual del MLG es que las variables explicativas  $x$  son deterministas o no estocásticas.
  - Es decir, se supone que las observaciones de las variables explicativas son fijas en muestras repetidas, como sucedería, por ejemplo, en una situación experimental, en la que el científico elige los valores de las variables explicativas y luego observa  $y_i$ .
    - Nótese, por tanto, que el supuesto no incluye a la variable endógena  $y$  (que es aleatoria por ser función del término de error).
  - Sintetizando, el supuesto establece que  $X$  es una matriz  $N \times K$  de constantes conocidas.
  - Este supuesto garantiza, en coherencia con el S3, que la esperanza condicionada a  $x$  de  $\varepsilon$  es nula, lo cual, como veremos después, es crucial para que los estimadores habituales sean insesgados. Por tanto, podríamos relajar el supuesto de regresores deterministas a un bajo coste puesto que, en definitiva, el supuesto crucial es el S3.
- Limitación del supuesto:** En la práctica, los científicos sociales muy rara vez disponen de datos experimentales<sup>5</sup>. Por eso, el supuesto de regresores deterministas es, en realidad, una “conveniencia matemática” que permite tratar los regresores como constantes dentro de la distribución de  $y$ . Esto permite simplificar la notación anterior y evitar condicionar en  $X$  [ver tema 5.B.2].

#### 2.3.5 DATA GENERATING PROCESS FOR THE REGRESSORS

It is common to assume that  $\mathbf{x}_i$  is nonstochastic, as it would be in an experimental situation. Here the analyst chooses the values of the regressors and then observes  $y_i$ . This process might apply, for example, in an agricultural experiment in which  $y_i$  is yield and  $\mathbf{x}_i$  is fertilizer concentration and water applied. The assumption of **nonstochastic regressors** at this point would be a mathematical convenience. With it, we could use the results of elementary statistics to obtain our results by treating the vector  $\mathbf{x}_i$  simply as a known constant in the probability distribution of  $y_i$ . With this simplification, Assumptions A3 and A4 would be made unconditional and the counterparts would now simply state that the probability distribution of  $\varepsilon_i$  involves none of the constants in  $\mathbf{X}$ .

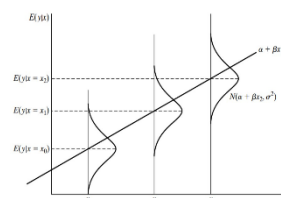
Social scientists are almost never able to analyze experimental data, and relatively few of their models are built around nonrandom regressors. Clearly, for example, in any model of the macroeconomy, it would be difficult to defend such an asymmetric treatment of aggregate data. Realistically, we have to allow the data on  $\mathbf{x}_i$  to be random the same as  $y_i$ . So an alternative formulation is to assume that  $\mathbf{x}_i$  is a random vector and our formal assumption concerns the nature of the random process that produces  $\mathbf{x}_i$ . If  $\mathbf{x}_i$  is taken to be a random vector, then Assumptions A1 through A4 become a statement about the joint distribution of  $y_i$  and  $\mathbf{x}_i$ . The precise nature of the regressor and how we view the sampling process will be a major determinant of our derivation of the statistical properties of our estimators and test statistics. In the end, the crucial assumption is A3, the uncorrelatedness of  $\mathbf{X}$  and  $\varepsilon$ . Now, we do note that this alternative is not completely satisfactory either, because  $\mathbf{X}$  may well contain nonstochastic elements, including a constant, a time trend, and dummy variables that mark specific episodes in time. This makes for an ambiguous conclusion, but there is a straightforward and economically useful way out of it. We will allow  $\mathbf{X}$  to be any mixture of constants and random variables, and the mean and variance of  $\varepsilon_i$  are both independent of all elements of  $\mathbf{X}$ .

ASSUMPTION A5:  $\mathbf{X}$  may be fixed or random.

(2-10)

### 1.1.6. Supuesto 6: Distribución normal de los errores

- Es conveniente añadir el supuesto de que los términos de error **se distribuyen como una normal**, con *media cero* y *varianza constante*, de tal manera que:  $\varepsilon|\mathbf{X} \sim N(0, \sigma^2 \cdot I)$ .
  - Este supuesto es razonable si tenemos en cuenta que las condiciones para la aplicación del Teorema Central del Límite son generalmente aplicables en este caso. Es decir, aunque los errores tuvieran una distribución distinta, este teorema asegura que, en muestras grandes, su distribución tenderá a la Normal. Ello implica que los errores son estadísticamente independientes e incorrelados.
  - Aunque no es un supuesto imprescindible, *es útil para el posterior contraste de hipótesis*.



<sup>5</sup> Este supuesto es más realista en ciencias naturales, en las que se puede realizar un experimento y recrear unas condiciones determinadas para observar un resultado. Por ejemplo, se podría tener en cuenta un supuesto en el que medimos el tamaño de una planta (como variable  $y$ ) en función del tiempo desde que se sembró, las condiciones de humedad y temperatura, etc.

Sin embargo, en ciencias sociales es menos habitual este tipo de experimentos. Por ejemplo, al analizar cualquier modelo macroeconómico sería difícil defender un tratamiento tan asimétrico de los datos. Es más realista pensar que tanto la variable  $y_i$  como las variables explicativas  $\mathbf{x}_i$  son aleatorias. En este caso, los supuestos anteriores deben ser entendidos como supuestos acerca de la distribución conjunta de  $y_i$  y  $\mathbf{x}_i$ .

- En **resumen**, los **supuestos del modelo lineal general** son los siguientes<sup>6</sup>:

- S1.  $y = X\beta + \varepsilon$
- S2.  $X$  tiene una dimensión  $N \times K$  y tiene rango  $K$ .
- S3.  $E[\varepsilon|X] = 0$
- S4.  $E[\varepsilon \cdot \varepsilon' | X] = \sigma^2 \cdot I$
- S5.  $X$  es una matriz no estocástica.
- S6.  $\varepsilon|X \sim N(0, \sigma^2 \cdot I)$

## 1.2. Métodos de regresión a partir de la información muestral

- Los parámetros poblacionales  $\beta$  no son observables, por lo que uno de los principales problemas a los que se enfrenta la econometría consiste en encontrar el estimador de  $\beta$  que logre el “mejor ajuste” entre las variables explicativas y la variable endógena, partiendo de la información muestral.
- En principio, existirán tantos estimadores como funciones entre el espacio muestral y el espacio paramétrico. Por tanto, es preciso introducir algún **criterio de “mejor ajuste”** que permita seleccionar un estimador entre el conjunto de todos los posibles<sup>7</sup>.
  - Por motivos prácticos y teóricos, los dos más utilizados son el método de *mínimos cuadrados ordinarios* (MCO) y el método de *máxima verosimilitud* (MV).
    - Ambos métodos permiten encontrar las estimaciones  $\hat{\beta}$  del vector de parámetros poblacionales  $\beta$ .
    - También nos permitirán encontrar una estimación  $\hat{y}$  de la variable endógena  $y$ .
    - En general, siempre habrá una discrepancia entre el verdadero valor ( $y$ ) y su estimación ( $\hat{y}$ ) denominada residuo del modelo ( $\hat{\varepsilon}$ ):  $\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta}$

## 2. MÉTODO DE MÍNIMOS CUADRADOS ORDINARIOS

### 2.1. Método de estimación (problema de optimización)

- **Concepto:** Constituye, sin duda, el método más habitual para la estimación del modelo de regresión lineal. La idea principal del método de MCO es obtener un estimador tal que la magnitud de los residuos (generados a partir del estimador en cuestión) sea mínima.
  - En concreto, el estimador de MCO es aquel que minimiza la norma euclídea del vector  $\varepsilon$ , es decir, de la suma de residuos al cuadrado (en adelante, SR). En notación matricial<sup>8</sup>:
 
$$SR(\hat{\beta}) = \hat{\varepsilon}' \cdot \hat{\varepsilon} = (y - X \cdot \hat{\beta})'(y - X \cdot \hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$
- **Justificación:** ¿Por qué el criterio de MCO escoge minimizar la suma de residuos al cuadrado y no una simple suma de los residuos?
  1. Los errores obtenidos a partir de  $\hat{\beta}$  y de las observaciones muestrales serán positivos para algunas observaciones y negativos para otras. Consecuentemente, sería un error minimizar la suma de éstos, dado que se tendría un problema de cancelación de residuos: éstos podrían ser muy altos en términos absolutos, pero anularse mutuamente. Además, el minimizar la suma, incentivaría la elección de  $\hat{\beta}$  de gran tamaño en términos absolutos pero negativos, lo cual no parece adecuado.

<sup>6</sup> Autores como A. Novales incluyen algunos supuestos o características adicionales del modelo:

- Especificación correcta del modelo de regresión (omitir una variable relevante puede suponer un sesgo en los estimadores obtenidos).
- Los valores que componen el vector  $\beta$  son constantes a lo largo del tiempo.
- Existe una relación causal desde las variables explicativas hacia la variable endógena.

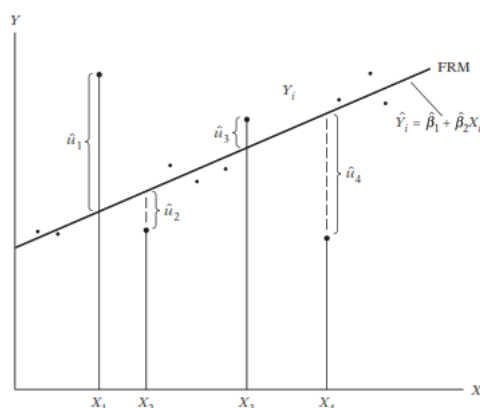
<sup>7</sup> Debe quedar claro que criterios diferentes conducirán a estimadores diferentes, incluso a partir de una misma muestra. De modo similar, un determinado estimador producirá una estimación a partir de una muestra dada, pero generaría una estimación diferente si se dispusiese de una muestra distinta.

<sup>8</sup> En notación normal:  $SR(\hat{\beta}) = \sum_{i=1}^N \hat{\varepsilon}_i^2$ .



2. Minimizar la suma de los valores absolutos de los residuos también plantearía dificultades técnicas, debido a que introduciría discontinuidades en el problema de optimización.

IMAGEN 1.– Criterio de mínimos cuadrados



Fuente: Gujarati, D. & Porter, D. (2010) *Econometría*. 5ª ed. McGrawHill. Pág. 56

- **Estimador:** El problema de optimización de MCO consiste, por tanto, en elegir el vector de estimaciones  $\hat{\beta}$  que minimice la suma de residuos al cuadrado:

– *Problema de Optimización:*

$$\min_{\hat{\beta}} SR(\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

– CPO: Derivando con respecto al vector  $\hat{\beta}$ :

$$\frac{\partial SR(\hat{\beta})}{\partial(\hat{\beta})} = -2X'y + 2X'X\hat{\beta} = 0$$

- *Estimador:* Despejando, se obtiene la siguiente expresión del estimador de MCO (vector columna de dimensión K):

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

- Para garantizar que puede encontrarse una solución única del estimador MCO, se requiere que exista la inversa de  $(X'X)$ , para lo cual es imprescindible que se cumpla el supuesto de rango completo de  $X$  (S2)<sup>9</sup>.
- CSO: Para asegurar que efectivamente el estimador es el resultado de una minimización de los residuos se exige que la matriz de segundas derivadas o matriz hessiana debe ser definida positiva. En este caso:

$$\frac{\partial^2 SR^2(\hat{\beta})}{\partial(\hat{\beta})\partial(\hat{\beta})'} = 2X'X$$

- Dicha condición se cumple pues  $X'X$  es siempre una matriz definida positiva.

- **Bondad de Ajuste:** El MCO ofrece una manera de ajustar el modelo de regresión a los datos. No obstante, no sólo nos interesa conocer los estimadores, sino cómo de bien se ajustan nuestras estimaciones a los datos. Para ello introducimos los siguientes conceptos:

- **Suma de los cuadrados totales (SST):** Es una métrica que refleja cuánto varía la variable dependiente y respecto a su media:

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

- La variable  $y$  puede ser expresada como:  $y = X\beta + \varepsilon = \hat{y} + \varepsilon$
- En consecuencia:  $y_i - \bar{y} = \hat{y}_i - \bar{y} + \varepsilon_i = (X_i - \bar{X})'\beta + \varepsilon_i$
- Teniendo en cuenta que  $M^0$  es una matriz  $N \times N$  que transforma las observaciones en desviaciones con respecto a la media muestral, se puede expresar en términos matriciales:

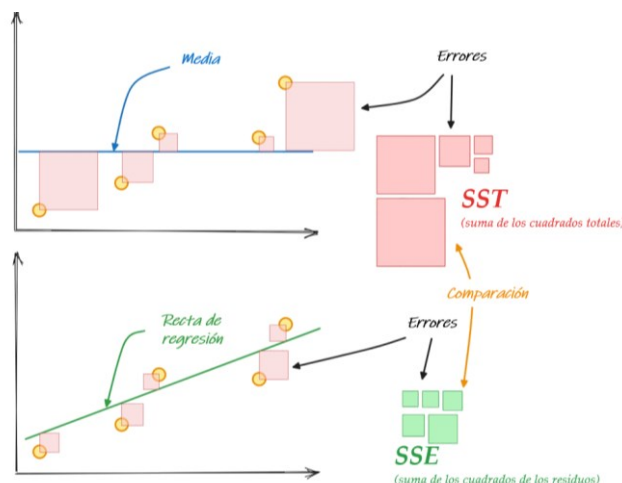
$$M^0 y = M^0 X\beta + M^0 \varepsilon$$

- Ya que los residuos tienen una media de 0:  $M^0 \varepsilon = \varepsilon$

<sup>9</sup> Si  $X'X$  fuera una matriz singular, el sistema de ecuaciones normales tendría infinitas soluciones.

- Por tanto se puede expresar la ecuación previa como:  $y'M^0y = \beta'X'M^0X\beta + \varepsilon'\varepsilon$ . Esto es, la suma de los cuadrados totales es igual a la suma de los cuadrados de la regresión más la suma de los cuadrados de los residuos:  $SST = SSR + SSE$
- **Coefficiente de determinación o  $R^2$** : Mide la proporción total de la variación de la variable endógena  $y$  explicada por la variación de los regresores. El coeficiente de regresión está contenido entre 0 y 1.
  - Si toma valor igual a 1, implica que el ajuste es perfecto, es decir, que la variación de la variable endógena es explicada perfectamente por la variación de los regresores de tal forma que todos los valores de  $y$  se encontrarán en el hiperplano.
  - Mientras que si toma un valor igual a 0 la regresión alcanzada es una recta horizontal.

$$R^2 = \frac{SSR}{SST} = \frac{\beta'X'M^0X\beta}{y'M^0y} = 1 - \frac{\varepsilon'\varepsilon}{y'M^0y}$$



$$R^2 = 1 - \frac{SSE}{SST}$$

$$SST = \sum_i (y_i - \bar{y})^2$$

$$SSE = \sum_i (y_i - f_i)^2$$

Fuente: Elaborado a partir de <https://twitter.com/levikul09/status/1741040229983896041/photo/1>

## 2.2. Propiedades estadísticas del estimador de MCO

- Dado que el estimador MCO depende del vector de observaciones de la variable endógena  $y$  (y, por tanto, también del vector de perturbaciones  $\varepsilon$ ),  $\hat{\beta}$  será un vector aleatorio. Por tanto, cabe estudiar:
  - 1) Su esperanza matemática;
  - 2) Su matriz de covarianzas; y
  - 3) Sus propiedades de estimación.

### 2.2.1. Propiedades para muestras finitas

#### Esperanza del estimador: insesgadez

- Partiendo de la ecuación anterior:

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

- Tomando esperanzas en la expresión anterior (y recordando que se asume que  $X$  es determinista y, por tanto, con covarianza nula con  $\varepsilon$ ), puede demostrarse que el estimador MCO es insesgado:

$$E[\hat{\beta}] = E[\beta + (X'X)^{-1}X'\varepsilon] = E[\beta] + (X'X)^{-1}X' \cdot E[\varepsilon] = \beta + (X'X)^{-1}X'0 = \beta$$

#### Varianza del estimador: eficiencia (T<sup>ma</sup> de Gauss-Markov)

- La matriz de varianzas-covarianzas del estimador de mínimos cuadrados es importante porque:
  - Permite juzgar la exactitud con que la estimación obtenida se aproxima a su esperanza matemática,  $\beta$ .
  - Es necesaria para poder hacer contrastes de hipótesis acerca de  $\beta$  o para establecer intervalos de confianza.



- Considerando la propiedad de insesgadez del estimador MCO, puede obtenerse su matriz de varianzas-covarianzas de la siguiente manera:
  - Por definición:  $VCov(\hat{\beta}) = E[(\hat{\beta} - E[\hat{\beta}]) \cdot (\hat{\beta} - E[\hat{\beta}])']$ .
  - Por la insesgadez del estimador:  $VCov(\hat{\beta}) = E[(\hat{\beta} - \beta) \cdot (\hat{\beta} - \beta)']$ .
  - Por la fórmula anterior:  $VCov(\hat{\beta}) = E[(X'X)^{-1}X'\varepsilon \cdot \varepsilon'X(X'X)^{-1}]$ .
  - Por el supuesto 5 de regresores deterministas:  $VCov(\hat{\beta}) = (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1}$ .
  - Por el supuesto 4 de perturbaciones esféricas  $VCov(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2I)X(X'X)^{-1}$ .
  - $VCov(\hat{\beta}) = \sigma^2(X'X)^{-1}$ .
- El problema reside en que el parámetro poblacional  $\sigma^2$  es desconocido, por lo que la matriz de varianzas covarianzas es también desconocida. Por ello, debemos recurrir al estimador de mínimos cuadrados de la varianza del término de error (calculado a partir de los residuos del modelo):

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N - K}$$

- **Teorema de Gauss-Márkov:** Bajo los supuestos del MLG, el estimador  $\hat{\beta}$  de MCO es el estimador lineal insesgado óptimo (ELIO) (Best Linear Unbiased Estimator, BLUE), es decir, tiene una varianza menor o igual a cualquier otro estimador lineal e insesgado. En otras palabras, el estimador MCO es eficiente bajo estos supuestos [para la demostración ver anexo A.1].
  - Como corolario, podemos señalar que obtener un estimador lineal de menor varianza que el MCO implicaría introducir un sesgo en las estimaciones (existe un *trade-off*).

### 2.2.2. Normalidad y distribución de $\hat{\beta}$

- Asumiendo que  $\varepsilon$  se distribuye como una normal multivariante (S6), cada elemento de  $\hat{\beta}|X$  se distribuirá normalmente (por el teorema del mapeo continuo de Mann-Wald, pues es una combinación lineal de elementos distribuidos normalmente):

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

### 2.2.3. Propiedades asintóticas: consistencia y distribución

- También es posible demostrar la consistencia del estimador, es decir: éste tiende probabilísticamente al verdadero valor poblacional a medida que la muestra crece hacia infinito.

$$plim \hat{\beta} = \beta$$

- Prueba: Cuando tenemos regresores deterministas, la prueba es más sencilla:

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

$$plim(\hat{\beta} - \beta) = plim\left(\left(\frac{X'X}{N}\right)^{-1} \frac{X'\varepsilon}{N}\right) \quad \text{multiplicando y dividiendo por } N$$

$$plim(\hat{\beta} - \beta) = plim\left(\frac{X'X}{N}\right)^{-1} \cdot plim\left(\frac{X'\varepsilon}{N}\right) \quad \text{por el Tma de Slutsky}$$

$$plim(\hat{\beta} - \beta) = \Sigma(xx)^{-1} \cdot E[x_i \cdot \varepsilon_i] \quad \text{por Ley de los Grandes Números (de KHINCHINE)}$$

$$plim(\hat{\beta} - \beta) = \Sigma(xx)^{-1} \cdot 0 \quad \text{por el S2 (2º momento finito) y S3 de exogeneidad}$$

$$plim(\hat{\beta} - \beta) = 0$$

- Además, por el Teorema Central del Límite (de Lindeberg-Lévy), podemos demostrar que:

$$\sqrt{N} \cdot (\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 \cdot \Sigma_{xx}^{-1})$$

### 3. MÉTODO DE MÁXIMA VEROSIMILITUD

#### 3.1. Método de estimación (problema de optimización)

- **Concepto:** A diferencia del método de MCO, el método de MV descansa directamente en un supuesto sobre la distribución del vector de errores, que en la mayoría de los casos se asume normal (S6). Si el supuesto es correcto, se ganará en eficiencia utilizando tal información adicional.
  - La idea de este método es elegir los estimadores de  $\beta$  y  $\sigma^2$  que maximizan la probabilidad de que el verdadero proceso generador de datos haya generado la muestra efectivamente observada<sup>10</sup>.
- **Función de Verosimilitud:** Partimos de la función de la densidad de probabilidad o pdf para una variable aleatoria  $y$ , condicionada en un conjunto de parámetros  $\theta$  y denotada como  $f(y|\theta)$ . Esta función identifica el proceso generador de datos que subyace a los datos observados, a la vez que nos ofrece una descripción matemática de los datos del proceso que generará.
  - La densidad conjunta de  $N$  observaciones independientes e idénticamente distribuidas (*iid*) puede ser expresada como el producto de los valores de la función de densidad para cada una de las observaciones:

$$f(y_1, y_2, \dots, y_N | \theta) = \prod_{i=1}^N f(y_i | \theta) = L(\theta | y)$$

- La densidad conjunta de la función de verosimilitud, definida como la función del vector de parámetros desconocido  $\theta$ , siendo  $y$  la colección muestral de datos. Esta función puede ser expresada en términos aditivos.

$$\ln(L(\theta | y)) = \sum_{i=1}^N \ln(f(y_i | \theta))$$

- Cuando los errores se encuentran normalmente distribuidos, entonces  $y_i$  condicionada sobre  $X$ , se distribuye normalmente de manera normal, con una media  $\mu_i = X_i'\beta$  y una varianza  $\sigma^2$  y la función de densidad puede ser expresada como:

$$\ln(L(\theta | y, X)) = \sum_{i=1}^N \ln(f(y_i | \theta)) = -\frac{1}{2} \sum_{i=1}^N \left( \ln(\sigma^2) + \ln(2\pi) + \ln\left(\frac{y_i - x_i'\beta}{\sigma^2}\right) \right)$$

- **Estimador:** El principio de máxima verosimilitud nos ofrece un medio para elegir un estimador asintóticamente eficiente para el parámetro o el conjunto de parámetros deseados. Para ello se derivará la función de verosimilitud con respecto del parámetro de tal manera que se estime aquel que tiene la máxima verosimilitud de corresponderse con el parámetro poblacional.
  - A partir<sup>11</sup> de la función de verosimilitud anteriormente expresada, podemos derivar las condiciones de maximización:

$$\begin{aligned} \frac{\partial \ln(L)}{\partial \beta} &= \frac{1}{\sigma^2} X'(Y - X\beta) = 0 \\ \frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{N}{2\sigma^2} + \frac{\varepsilon'\varepsilon}{2\sigma^4} = 0 \end{aligned}$$

- De donde se deduce que:

$$\begin{aligned} \hat{\beta}_{MV} &= (X'X)^{-1} X'y \\ \hat{\sigma}_{MV}^2 &= \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N} \end{aligned}$$

<sup>10</sup> Intuitivamente, se acepta que, de varios sucesos posibles, ocurrirá el más probable. Por ejemplo, si en una bolsa hay más bolas blancas que negras y extraemos una al azar, lo más probable es que ésta sea blanca. Con el método de máxima verosimilitud, si no sabemos qué tipo de bolas son las más numerosas y extraemos una bola blanca (muestra), entonces inferiremos que el número de bolas blancas es superior al de negras en la bolsa (población).

<sup>11</sup> La expresión anterior ilustra claramente la propiedad de que, con una distribución Normal del término de error, si el parámetro  $\sigma^2$  no depende de ninguno de los parámetros  $\beta$ , entonces escoger el vector de parámetros  $\hat{\beta}$  que maximice la función de verosimilitud (o su logaritmo) es equivalente a escoger el vector  $\hat{\beta}$  que minimice la suma residual (numerador del último término de la ecuación). En consecuencia, los estimadores de MV y de MCO coinciden.

### 3.2. Propiedades estadísticas del estimador de MV

- El estimador de máxima verosimilitud posee las siguientes **propiedades**:
  - 1) Estimador insesgado: Cuando los errores se distribuyen de forma normal, el estimador de MV del vector  $\beta$  coincide con el estimador MCO. Tal conclusión implica que  $\hat{\beta}_{MV}$  es un estimador insesgado de  $\beta$ .
  - 2) Consistencia: Bajo ciertas condiciones habituales el estimador MV es consistente: si el número de observaciones  $N$  tiende a infinito, el estimador  $\hat{\theta}$  converge en probabilidad a su valor verdadero  $plim \hat{\theta} = \theta$ .
  - 3) Normalidad asintótica: Las estimaciones se encontrarán normalmente distribuidas  $\hat{\theta} \sim N[\theta_0, \{I(\theta_0)\}^{-1}]$ .
  - 4) Estimador de la varianza-covarianza sesgado: El estimador de la varianza-covarianza ( $\hat{\sigma}_{MV}^2$ ) difiere del estimador MCO, dado que, en el denominador, figura “ $N$ ” en lugar de “ $N - K$ ”. Dado que MCO era insesgado, el estimador de máxima verosimilitud de  $\sigma^2$  está sesgado hacia cero (es más bajo)<sup>12</sup>, si bien es consistente (cuando  $N$  tiende a infinito el sesgo del estimador máximo verosímil de  $\sigma^2$  se hace cero). Es decir,  $\hat{\sigma}_{MV}^2$  cumple todas las propiedades asintóticas deseables.
- Podemos concluir que no existe estimador insesgado de  $\sigma^2$  que alcance la **cota de Cramer-Rao**. Con lo que tenemos un trade-off:
  - Por un lado,  $\hat{\sigma}_{MCO}^2$  es el de menor varianza dentro de todos los insesgados.
  - Por otro lado,  $\hat{\sigma}_{MV}^2$  es sesgado, pero tiene una varianza menor que el de MCO.

### CONCLUSIÓN

- El MLG es el modelo de referencia. Se basa en supuestos que pueden ser razonables en algunos casos, pero no en otros. El levantamiento de alguno de los 6 supuestos enunciados puede producir consecuencias más o menos graves sobre las propiedades de los estimadores. En la misma línea, el levantamiento de estos supuestos requerirá de métodos más o menos complejos para resolver estas consecuencias.
- Algunos ejemplos de levantamiento de supuestos y algunas soluciones, tratados en otra parte del temario:
  - **S1** [ver tema 5.B.5]:
    - Modelos no lineales en variables. Soluciones: transformación de variables.
    - Modelos no lineales en coeficientes. Soluciones: MCNL, transformación Box-Cox.
  - **S2** [ver tema 5.B.3]: Multicolinealidad.
    - Perfecta. Solución: revisar la especificación del modelo.
    - Aproximada. Solución más compleja: exclusión de variables.
  - **S3** [ver tema 5.B.2 y 5.B.4]: Variables explicativas endógenas (variable omitida, error de medición...). Solución: estimación por variables instrumentales.
  - **S4** [ver tema 5.B.3]: Perturbaciones no esféricas (heterocedasticidad y autocorrelación). Soluciones: método de mínimos cuadrados generalizado, estimadores robustos de WHITE y NEWKEY WEST, modelos de series temporales ARIMA...
  - **S5** [ver tema 5.B.2 y 5.B.4]:
    - Regresores estocásticos exógenos. No hay grandes problemas.
    - Regresores estocásticos endógenos. Incumplimiento del S3 [ver tema 5.B.2 y 5.B.4].

<sup>12</sup>  $E[\hat{\sigma}_{MV}^2] = \frac{N-K}{N} \cdot \sigma^2 = \left(1 - \frac{K}{N}\right) \cdot \sigma^2 < \sigma^2$

## Bibliografía

Tema de la promoción LXX

Econometric Analysis, William H. Greene

## Anexos

### A.1. Anexo 1: Demostración del teorema Gauss-Márkov

- Supongamos que  $\tilde{\beta} = Cy$  es un estimador lineal e insesgado de  $\beta$  distinto a MCO, donde  $C$  es una matriz  $K \times N$ .

- Como este estimador es insesgado, tenemos que:

$$E \begin{bmatrix} \tilde{\beta} \\ \tilde{Cy} \end{bmatrix} = E[CY\beta + C\epsilon] = E[CY\beta] + \underbrace{E[C\epsilon]}_{=0} = \beta$$

- Para que la anterior ecuación sea cierta, deberá cumplirse que  $CX = I$ . Esto implica que hay muchísimos estimadores lineales e insesgados de  $\beta$ . Podríamos remplazar  $(X'X)^{-1}X'$  por  $C$  y obtener la matriz de covarianzas de  $\tilde{\beta}$  tal que:

$$CX = I \xRightarrow{\text{podría ser}} C = (X'X)^{-1}X' \Rightarrow C'C = \underbrace{(X'X)^{-1}X'X}_{I}(X'X)^{-1} = (X'X)^{-1} \left. \vphantom{C} \right\} \Rightarrow Var(\tilde{\beta}) = \sigma^2 \cdot C'C$$

$$Var(\tilde{\beta}) = \sigma^2 \cdot (X'X)^{-1} \text{ (a través de insesgadez y supuestos 4 y 5 ver pág. 9)}$$

- Considerando que  $D = C - (X'X)^{-1}X'$ , esto es,  $Dy = \tilde{\beta} - \hat{\beta}$ , podemos expresar la matriz de covarianzas de  $\tilde{\beta}$  como sigue:

$$Var(\tilde{\beta}) = \sigma^2 [(D + \underbrace{(X'X)^{-1}X'}_I)(D + (X'X)^{-1}X')']$$

- Sabiendo que  $CX = I = DX + \underbrace{(X'X)^{-1}X'X}_I$  es fácil deducir que  $DX = 0$ , así pues:

$$Var(\tilde{\beta}) = \underbrace{\sigma^2(X'X)^{-1}}_{=Var(\hat{\beta})} + \sigma^2 DD' = Var(\hat{\beta}) + \sigma^2 DD'$$

- Puesto que  $\sigma^2 DD'$  es una matriz definida no negativa, la matriz de covarianzas de  $\tilde{\beta}$  en ningún caso será menor que la de  $\hat{\beta}$  (estimador MCO).

- Es decir, se demuestra que  $\hat{\beta}_{MCO}$  es el estimador lineal óptimo insesgado. ■



## 5.B.2 : EL MODELO DE REGRESIÓN CON REGRESORES ESTOCÁSTICOS: ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

### ▪ Contextualización:

- Desde un punto de vista histórico, desde las primeras aportaciones de GAUSS a principios del s. XIX, el modelo más utilizado y estudiado por los econométricos ha sido el modelo lineal general (o modelo de regresión clásico).
  - Este modelo ha sido aplicado con gran éxito en el estudio de un elevado nº de temas económicos. Además, sirve de base para el desarrollo de modelos más complejos que levantan los supuestos clásicos, que pueden ser restrictivos en determinadas aplicaciones (series temporales, relaciones no lineales...).
- En este sentido, *cuando se incumplen algunos de los supuestos básicos en los que se basa el MLG, los métodos de estimación habituales (p.ej. MCO o MV), pueden dejar de cumplir algunas de sus propiedades de insesgadez, consistencia asintótica o eficiencia. Por ello, será siempre necesario evaluar con espíritu crítico, caso por caso, la validez de los supuestos del MLG antes de realizar cualquier ejercicio econométrico. En caso de detectar algún problema, deberemos evaluar sus consecuencias y, de ser graves, aplicar algún método econométrico para solucionarlo.*
  - En esta exposición, nos vamos a centrar en una desviación de los supuestos clásicos: modelo de regresores estocásticos.

### ▪ Problemática (Preguntas clave):

–

### ▪ Estructura:

#### 0. SUPUESTOS CLÁSICOS DEL MODELO LINEAL GENERAL (MLG)

- Supuesto 1: Linealidad del modelo
- Supuesto 2: Rango completo (no multicolinealidad perfecta)
- Supuesto 3: Esperanza condicionada nula de las perturbaciones
- Supuesto 4: Perturbaciones esféricas
- Supuesto 6: Distribución normal de los errores
- Supuesto 5: Regresores deterministas

#### 1. ESTIMACIÓN CON REGRESORES ESTOCÁSTICOS

- 1.1. Idea
- 1.2. Estimación por MCO con regresores estocásticos exógenos
  - 1.2.1. Método
  - 1.2.2. Propiedades
- 1.3. MCO con regresores estocásticos endógenos
  - 1.3.1. Causas y consecuencias
    - Consecuencias: inconsistencia del estimador
    - Origen: Causas habituales de endogeneidad
  - 1.3.2. Estimación con endogeneidad
    - Método de variables instrumentales
    - Sobreidentificación: mínimos cuadrados bietápicos (MC2E)

#### 2. VERIFICACIÓN

- 2.1. Test de exogeneidad de HAUSMAN y WU
- 2.2. Contratación de hipótesis y construcción de intervalos de confianza
  - Idea
  - Contrastes de hipótesis
  - Intervalos y regiones de confianza
  - Consideraciones específicas para estimadores MC2E

#### 3. PREDICCIÓN

- 3.1. Cálculo de predicciones
- 3.2. Error de predicción
- 3.3. Varianza del error de predicción e intervalos de confianza



## 0. SUPUESTOS CLÁSICOS DEL MODELO LINEAL GENERAL (MLG)

- El **Modelo Lineal General (MLG)** se basa en un conjunto de supuestos sobre cómo los datos muestrales serán generados en un proceso subyacente de generación. Como veremos, bajo estos supuestos del MLG, los estimadores cumplen una serie de “buenas propiedades” si se cumplen los supuestos clásicos enunciados originalmente por GAUSS a principios del s. XIX.
  - Se utilizará como referencia de “*Econometric Analysis*” realizada por GREENE para la explicación.

### Supuesto 1: Linealidad del modelo

- El modelo que relaciona la variable endógena con las variables explicativas es **lineal en los coeficientes  $\beta$** .
  - Por tanto, el modelo se expresa del siguiente modo:
 
$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ki}, \varepsilon_i) = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_k \cdot x_{ki} + \varepsilon_i \quad \forall i = 1, \dots, N$$
 donde,
    - $\beta_0$  es el término independiente o *intercepto*. El término independiente puede interpretarse como la esperanza de la variable dependiente cuando el resto de las variables es nulo o puede interpretarse aquel que acompaña a una primera variable explicativa  $x_{0i}$  cuyo valor es siempre igual a 1.
    - $\beta_1, \dots, \beta_k$  son los coeficientes o *pendientes* de la regresión. Recogen la magnitud del impacto de cada una de las variables explicativas sobre la variable endógena.

### Supuesto 2: Rango completo (no multicolinealidad perfecta)

- $X$  es una matriz de dimensión  $N \times K$  con rango  $K$ , por lo que es una matriz con rango completo. Esto requiere que se cumplan 2 condiciones:
  - Las columnas de  $X$  son linealmente independientes, lo que implica que las variables explicativas son linealmente independientes, es decir, ninguna variable explicativa es combinación lineal de otra.
  - Condición de identificación: Disponemos de al menos  $K$  observaciones ( $N \geq K$ )<sup>3</sup>.

### Supuesto 3: Esperanza condicionada nula de las perturbaciones

- Para todo el conjunto de observaciones, se asume que los términos de error tienen esperanza condicionada a  $X$  nula:
  - Ello implica que la esperanza de cada error, condicionado a la información contenida en el vector  $X$  es nula, lo que significa que ninguna observación de  $x$  provee información sobre el valor esperado del error (son ortogonales)
  - Por la ley de expectativas iteradas se obtiene que  $E[\varepsilon_i|X] = E_X[E[\varepsilon_i|X]] = E_X[0] = 0 \quad \forall i$  esto implica<sup>4</sup> que la media no condicionada del término error también es nula. Por otro lado, el supuesto también implica lo siguiente:  $Cov[X_i, \varepsilon_i] = 0 \quad \forall i = 1, \dots, N$
  - Este último supuesto “débil”, conocido como exogeneidad de las variables explicativas, garantiza la consistencia del método MCO. Sin embargo, su incumplimiento en casos como la presencia de variables omitidas, errores de medida o presencia de retardos de la variable endógena como variable explicativa, puede generar problemas de endogeneidad y requiere un tratamiento especial de los datos.

<sup>3</sup> Si el número de observaciones fuera menor que  $K$ ,  $X$  no podría tener rango completo. En principio, basta con que  $N = K$  para que  $X$  pueda llegar a tener rango  $K$ . Sin embargo, cuanto mayor sea  $N$ , más precisas tenderán a ser nuestras estimaciones. La diferencia  $(N - K)$  se conoce como “grados de libertad” de la regresión.

<sup>4</sup> La relación no es cierta en el sentido contrario.

### Supuesto 4: Perturbaciones esféricas

- El supuesto de perturbaciones esféricas queda resumido en la siguiente forma de la matriz de varianzas-covarianzas de los términos de error (matriz escalar):

$$E[\varepsilon \cdot \varepsilon' | \mathbf{X}] = \sigma^2 \cdot \mathbf{I} = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{bmatrix}$$

- Este supuesto contiene, por tanto, 2 sub-supuestos:
  - Homocedasticidad: Todos los términos de error tienen la misma varianza:

$$\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma^2 \quad \forall i = 1, \dots, N$$

- No autocorrelación: No existe correlación serial entre las perturbaciones:

$$\text{Cov}[\varepsilon_i, \varepsilon_j | \mathbf{X}] = 0 \quad \forall i \neq j$$

- Todo ello se resume en el siguiente supuesto:

$$E[\varepsilon \cdot \varepsilon' | \mathbf{X}] = \sigma^2 \cdot \mathbf{I} \rightarrow \text{Var}[\varepsilon] = E[\text{Var}[\varepsilon | \mathbf{X}]] + \text{Var}[E[\varepsilon | \mathbf{X}]] = \sigma^2 \cdot \mathbf{I}$$

### Supuesto 6: Distribución normal de los errores

- Es conveniente añadir el supuesto de que los términos de error se distribuyen como una normal, con media cero y varianza constante, de tal manera que:  $\varepsilon | \mathbf{X} \sim N(0, \sigma^2 \cdot \mathbf{I})$ .
  - Este supuesto es razonable si tenemos en cuenta que las condiciones para la aplicación del Teorema Central del Límite son generalmente aplicables en este caso. Es decir, aunque los errores tuvieran una distribución distinta, este teorema asegura que, en muestras grandes, su distribución tenderá a la normal. Ello implica que los errores son estadísticamente independientes e incorrelados.
  - Aunque no es un supuesto imprescindible, es útil para el posterior contraste de hipótesis.

### Supuesto 5: Regresores deterministas

- Un supuesto habitual del MLG es que las variables explicativas  $\mathbf{x}$  son deterministas o no estocásticas.
  - Es decir, se supone que las observaciones de las variables explicativas son fijas en muestras repetidas, como sucedería, por ejemplo, en una situación experimental, en la que el científico elige los valores de las variables explicativas y luego observa  $y_i$ .
    - Nótese, por tanto, que el supuesto no incluye a la variable endógena  $y$  (que es aleatoria por ser función del término de error).
  - Sintetizando, el supuesto establece que  $\mathbf{X}$  es una matriz  $N \times K$  de constantes conocidas.
  - Este supuesto garantiza, en coherencia con el S3, que la esperanza condicionada a  $\mathbf{X}$  de  $\varepsilon$  es nula, lo cual, como veremos después, es crucial para que los estimadores habituales sean insesgados. Por tanto, podríamos relajar el supuesto de regresores deterministas a un bajo coste puesto que, en definitiva, el supuesto crucial es el S3.
- Sin embargo, en la práctica, los científicos sociales muy rara vez disponen de datos experimentales<sup>5</sup>. Por eso, el supuesto de regresores deterministas es, en realidad, una “conveniencia matemática” que permite tratar los regresores como constantes dentro de la distribución de  $y$ . Esto permite simplificar la notación anterior y evitar condicionar en  $\mathbf{X}$ .
  - Por ello, en lo que resta de exposición levantaremos este supuesto e introduciremos *regresores estocásticos*.

<sup>5</sup> Este supuesto es más realista en ciencias naturales, en las que se puede realizar un experimento y recrear unas condiciones determinadas para observar un resultado. Por ejemplo, se podría tener en cuenta un supuesto en el que medimos el tamaño de una planta (como variable  $y$ ) en función del tiempo desde que se sembró, las condiciones de humedad y temperatura, etc.

Sin embargo, en ciencias sociales es menos habitual este tipo de experimentos. Por ejemplo, al analizar cualquier modelo macroeconómico sería difícil defender un tratamiento tan asimétrico de los datos. Es más realista pensar que tanto la variable  $y_i$  como las variables explicativas  $\mathbf{x}_i$  son aleatorias. En este caso, los supuestos anteriores deben ser entendidos como supuestos acerca de la distribución conjunta de  $y_i$  y  $\mathbf{x}_i$ .

## 1. ESTIMACIÓN CON REGRESORES ESTOCÁSTICOS

### 1.1. Idea

- Una vez introducimos regresores estocásticos, los supuestos anteriores deben entenderse como supuestos sobre la distribución conjunta de  $y$  y  $X$  (o, lo que es lo mismo, de  $\varepsilon$  y  $X$ ). En concreto, podemos distinguir 3 tipos de supuesto sobre la distribución conjunta de  $X$  y  $\varepsilon$ :
  - Supuesto estricto:  $X$  se compone de variables estocásticas plenamente independientes de  $\varepsilon$ .
  - Supuesto medio:  $X$  se compone de variables estocásticas media-independientes de  $\varepsilon$ . Es decir,  $E[\varepsilon|X] = k = 0$  (por S3).
  - Supuesto débil:  $X$  se compone de variables no correlacionadas en la misma fila con  $\varepsilon$ . Es decir,  $E[\varepsilon_i \cdot x_{k_i}] = 0 \forall i = 1, \dots, N$ . Este supuesto implica covarianza nula.
- En presencia de regresores estocásticos, la elección de los métodos de estimación más adecuados dependerá de los supuestos que se puedan realizar sobre la distribución conjunta del término de error y las variables explicativas. En concreto:
  - Cuando se cumple el supuesto estricto o medio, puede demostrarse que la estimación a través de MCO, cumple todas las buenas condiciones: insesgadez, consistencia...
  - Cuando se cumple el supuesto débil, el estimador MCO deja de ser insesgado, pero preserva sus propiedades asintóticas, es decir, es consistente. Por tanto, en muestras grandes, este método también puede ser aconsejable.
  - En caso de que se incumplan todos estos supuestos y estemos en presencia de variables explicativas endógenas, el método MCO conducirá a resultados sesgados e inconsistentes. Será, por tanto, recomendable introducir *nuevos métodos de estimación*, entre los que destacaremos el de variables instrumentales.

### 1.2. Estimación por MCO con regresores estocásticos exógenos

- En tal caso, se siguen cumpliendo las *propiedades deseables de MCO*, por lo que repasaremos éstas junto con el método de estimación:

#### 1.2.1. Método

- El estimador de MCO es aquel que minimiza la norma euclídea del vector  $\varepsilon$ , es decir, de la suma de residuos al cuadrado (en adelante, *SR*).
  - En notación matricial<sup>6</sup>:  $\min_{\beta} SR(\beta) = \hat{\varepsilon}'\hat{\varepsilon}$
  - Resolviendo el anterior problema, se encuentra la expresión clave para el estimador de MCO (vector columna de dimensión  $K$ ):  $\hat{\beta} = (X'X)^{-1}X'y$

#### 1.2.2. Propiedades

- Dados los supuestos del MLG, el estimador MCO será:
  - Insesgado: Dado que  $y$  es una variable estocástica afectada por un componente error inobservable  $\varepsilon$ , la estimación se verá afectada. No obstante, tomando esperanzas sobre el  $\hat{\beta}$ , se constata que éste coincide con el  $\beta$  poblacional.
  - Eficiente: Bajo los supuestos del MLG, el estimado de MCO es el estimador lineal insesgado óptimo. Es decir, tienen una varianza menor o igual a cualquier otro estimador lineal o insesgado. En otras palabras, el estimador MCO es eficiente.
  - Distribuido normalmente: Asumiendo que los errores se encuentran normalmente distribuidos, cada elemento  $\hat{\beta}|X$  se distribuirá normalmente (por el teorema del mapeo continuo de Mann-Wald, pues es una combinación lineal de elementos distribuidos normalmente):  $\hat{\beta}|X \sim N(\beta, \sigma^2(X'X)^{-1})$ .
  - Asintóticamente consistente: A medida que el tamaño muestral aumenta hasta el infinito, la  $\hat{\beta}$  converge con la  $\beta$  poblacional.

<sup>6</sup> En notación normal:  $SR(\beta) = \sum_{i=1}^n \hat{\varepsilon}_i^2$ .

### 1.3. MCO con regresores estocásticos endógenos

#### 1.3.1. Causas y consecuencias

#### Consecuencias: inconsistencia del estimador

- Si una variable explicativa es endógena, entonces el estimador  $\hat{\beta}$  de MCO es, no solamente sesgado, sino también inconsistente.

- Supongamos que el último regresor,  $x_K$ , es endógeno, es decir  $E[x_K \cdot \varepsilon] \neq 0$ , mientras que todos los demás son exógenos. Entonces, podemos re-examinar la prueba de consistencia del estimador MCO y comprobar que:

$$\begin{aligned} \text{plim}(\hat{\beta} - \beta) &= \text{plim}\left(\frac{X'X}{N}\right)^{-1} \cdot \text{plim}\left(\frac{X'\varepsilon}{N}\right) \\ \text{plim}(\hat{\beta} - \beta) &= \sum xx^{-1} \cdot \begin{pmatrix} 0 \\ \vdots \\ E[x_K \cdot \varepsilon] \end{pmatrix} \end{aligned}$$

- Por tanto, en general, todos los estimadores del vector  $\hat{\beta}$  serán inconsistentes (salvo cuando la matriz  $\sum xx^{-1}$  tenga elementos 0 en la columna  $k$ ).
  - Es decir, en general, la endogeneidad de un sólo regresor, provoca que el estimador MCO sea inconsistente no sólo para el coeficiente del regresor endógeno, sino para todos los  $K$  coeficientes.
  - La gravedad de este fenómeno dependerá de la correlación entre el regresor endógeno y los otros regresores.

#### Origen: Causas habituales de endogeneidad

- En la mayoría de modelos econométricos, la endogeneidad surge por alguno de los siguientes 4 problemas:
  - a) Variable relevante omitida.
  - b) Error de medición de las variables explicativas.
  - c) Endógena retardada y autocorrelación en el error.
  - d) Simultaneidad.

##### a) Variable omitida

- **Caracterización del problema:**

- Supongamos que el modelo correctamente especificado es  $y_i = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \varepsilon_i$ . No obstante, el investigador omite la variable  $x_2$  porque no dispone de información o porque la variable es intrínsecamente imposible de medir.

- Por ejemplo, supongamos que un investigador está interesado en medir el efecto de la educación sobre el salario. En el modelo anterior, la variable endógena  $y$  es el salario de un individuo  $i$ , mientras que la variable  $x_1$  es el nivel educativo y  $x_2$  es la habilidad innata del individuo. Como la habilidad innata no puede medirse, el investigador la excluye de la especificación del modelo. El modelo estimado, por tanto, es:

$$y_i = \beta_0 + \beta_1 \cdot x_{1i} + u_i$$

donde  $u_i = \beta_2 \cdot x_{2i} + \varepsilon_i$

- Si existe alguna correlación entre nivel educativo y habilidad innata (p.ej. como sugiere el modelo teórico de “*signalling*” elaborado por SPENCE (1970)), entonces la variable  $x_{1i}$  estará correlacionada con el error  $u_i$ , es decir, será endógena.

- **Casos en los que no se generan sesgos:** Los únicos 2 casos en los que omitir una variable no genera un sesgo ni problemas de inconsistencia son:

- Cuando la variable omitida no es relevante, es decir, cuando no afecta a  $y$ .
- Cuando la variable omitida es ortogonal a las variables incluidas (es decir, no está correlacionada con los demás regresores).

### ■ Soluciones:

- *Variable proxy* (p.ej. un proxy de la habilidad innata que podŕa servirnos es un índice de coeficiente intelectual).
- *Variables instrumentales* (p.ej. el barrio en el que nació el sujeto como VI de la educación<sup>7</sup>).

#### *b) Error de medición en las variables explicativas*

- **Caracterización del problema:** Cuando la variable explicativa que realmente influye sobre la variable endógena no puede ser medida adecuadamente por el econométra, entonces podemos demostrar que la estimación MCO será sesgada e inconsistente.

- Supongamos el siguiente modelo de regresión simple:

$$y_i = \beta_0 + \beta_1 \cdot x_{1i}^* + \varepsilon_i$$

No obstante, en la práctica, el investigador no puede observar con precisión el verdadero regresor  $x_{1i}^*$ , por lo que utiliza una variable con error  $x_{1i} = x_{1i}^* + u_i$ .

- En la mayoría de modelos de errores de medida, se supone que tanto  $\varepsilon_i$  como  $u_i$  son ruidos blancos (con media 0 y varianza constante, independientes entre sí). A pesar de ello, se puede demostrar que las estimaciones serán sesgadas e inconsistentes.
- Por tanto, el modelo realmente estimado sería:

$$y_i = \beta_0 + \beta_1 \cdot x_i + \varepsilon_i = \beta_0 + \beta_1 \cdot x_i^* + v_i$$

donde  $v_i = \varepsilon_i + \beta_1 \cdot u_i^*$ .

- La estimación de  $\beta_1$  cuando no existe error de medida sería:  $\hat{\beta}_1^* = \frac{Cov(x^*, y)}{Var(x^*)}$  que sabemos es insesgado y consistente.
- Sin embargo, cuando los datos disponibles contienen error de medida, tenemos:

$$\hat{\beta}_1 = \frac{Cov(x, y)}{Var(x)} = \frac{Cov(x^* + u, y)}{Var(x^* + u)} = \frac{Cov(x^*, y)}{Var(x^*) + Var(u)} = \frac{\hat{\beta}_1^*}{1 + \frac{Var(u)}{Var(x^*)}}$$

- Así pues, en la medida en que la ratio  $Var(u)/Var(x^*)$  sea elevado, el estimador MCO con errores de medida tendrá un sesgo hacia 0. Además, este sesgo no se anula para muestras grandes, por lo que el estimador también es inconsistente.

- **Casos donde no se generan sesgos:** Como señala JOHNSTON, es importante señalar que, en algunos casos, los agentes económicos reaccionan a los valores medidos de las variables económicas, y no a sus verdaderos valores.

- Por ejemplo, las empresas pueden basar sus decisiones de inversión en alguna extrapolación de las tendencias del PIB y, para hacerlo, utilizan las últimas estadísticas oficiales (aunque éstas contengan errores de medida). Por tanto, si las variables endógenas (en este caso, los niveles de inversión) responden a los datos medidos, entonces el error de medida es irrelevante a efectos de estimación y las técnicas de MCO seguirán siendo válidas.

#### *c) Endógena retardada y errores autocorrelacionados*

- **Caracterización del problema:**

- Supongamos el siguiente modelo en el que un retardo de la variable endógena actúa como variable explicativa:

$$y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot x_t + \varepsilon_t$$

donde el error del modelo  $\varepsilon_t$  está autocorrelacionado (sigue un proceso estacionario  $AR(1)$ ):

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t.$$

- Iterando hacia atrás, podemos comprobar que:  $y_{t-1} = \beta_0 + \beta_1 \cdot y_{t-2} + \beta_2 \cdot x_{t-1} + \varepsilon_{t-1}$
- Por tanto, en la primera ecuación, en la que  $y_{t-1}$  actúa como variable explicativa, vemos que esta variable está correlacionada con el error (ambos contienen  $\varepsilon_{t-1}$ ). En consecuencia,  $E[y_{t-1} \cdot \varepsilon_t] \neq 0$  lo que supone una violación de la condición de ortogonalidad.

<sup>7</sup> Es lógico pensar que el tipo de barrio está correlacionado positivamente con la educación (relevancia del instrumento: en los barrios ricos, el nivel de educación será superior) pero no con la habilidad innata del individuo (validez o exogeneidad del instrumento). Esta estrategia se ha utilizado en numerosos artículos de Economía de la educación.

d) Simultaneidad

- **Caracterización del problema:** Se da una situación de causalidad bidireccional entre la variable endógena y, al menos, una de las variables explicativas. En consecuencia, ya no puede estudiarse el comportamiento de la variable explicada y de las variables explicativas de una manera independiente [ver tema 5.B.4].

1.3.2. Estimación con endogeneidadMétodo de variables instrumentalesIdea

- 2 rasgos diferencian a la econometría de la estadística:
  - El interés por la identificación de relaciones causales, es decir, el interés de ir más allá del análisis descriptivo.
  - El conjunto de instrumentos desarrollados en su seno para la identificación y estimación de relaciones causales, siendo una de las más importantes las variables instrumentales.
- **Orígenes:**
  - Sus orígenes se remontan a la identificación de las ecuaciones simultáneas. En particular, la estimación simultánea de las curvas de oferta y demanda cuando la cantidad y el precio observados eran el resultado de la interacción entre ambas.
    - Fue WRIGHT (1928) quien resolvió este problema de simultaneidad al usar variables que sólo afectaban a un lado del mercado para comprender cómo el desplazamiento de una ecuación generaba cambios en el mercado interactuando con la otra.

Estimación general

- El método de variables instrumentales (VI) es el más utilizado para realizar estimaciones donde las variables explicativas presentan problemas de endogeneidad.
  - Volviendo al ejemplo de la estimación de la ecuación de MINCER, la educación está correlacionada con el término error (educación de los padres, renta familiar, coeficiente intelectual, etc.). No obstante, podemos utilizar como instrumento el barrio de nacimiento para eliminar la endogeneidad de la variable explicativa.
  - En particular, los instrumentos conformarán una matriz  $Z$  de dimensión  $N \times K$  (misma dimensión que  $X$ ) que debe cumplir las siguientes 3 propiedades:

$$\text{plim} \left( \frac{Z' \epsilon}{N} \right) = 0$$

$$\text{plim} \left( \frac{Z' X}{N} \right) = Q_{ZX}, \text{ donde } Q_{ZX} \text{ existe y es una matriz no singular}$$

$$\text{plim} \left( \frac{Z' Z}{N} \right) = Q_{ZZ}, \text{ donde } Q_{ZZ} \text{ existe y es una matriz no singular}$$

- Intuitivamente, las variables que componen la matriz  $Z$  cumplen:
  - Condición de exogeneidad: No están correlacionadas con el error del modelo.
  - Condición de relevancia: Están correlacionadas con  $X$ .
  - La tercera condición es trivial y garantiza que se puede obtener el estimador de VI obligando a que los instrumentos sean linealmente independientes entre sí (*para evitar multicolinealidad*).
- Transformamos el modelo original pre-multiplicando por la matriz  $Z$  traspuesta:
 
$$y = X \cdot \beta + \epsilon$$

$$Z' \cdot y = Z' \cdot X \cdot \beta + Z' \cdot \epsilon$$

$$y^* = X^* \cdot \beta + \epsilon^*$$
- Utilizamos el método MCO para estimar  $\beta$  en el nuevo modelo transformado:
 
$$\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} y^*$$

$$\hat{\beta} = (Z' X)^{-1} Z' y$$
  - Que es el **estimador VI**. Haciendo uso de las condiciones impuestas anteriormente, podemos demostrar que este estimador es consistente.



- En resumen, el modelo estimado será el siguiente:

$$y_i = \beta_0 + \rho \cdot z_i + \varepsilon_i$$

donde  $\rho = \text{cov}(y_i, z_i) / \text{cov}(x_i, z_i)$ .

#### Estimador de Wald

- El estimador más simple de VI usa una variable binaria (0-1) como instrumento para estimar un modelo con un único regresor endógeno y se conoce como **estimador de Wald**.

– Dado que  $\text{Cov}(y_i, z_i) = \{E[y_i | z_i = 1] - E[y_i | z_i = 0]\} \cdot p \cdot (1 - p)$ . Entonces  $\rho$  es:

$$\rho = \frac{\{E[y_i | z_i = 1] - E[y_i | z_i = 0]\}}{\{E[x_i | z_i = 1] - E[x_i | z_i = 0]\}}$$

- La fundamentación de este análisis para explotar la relación causal se basa en el supuesto de que la única razón por la que existe una relación entre la variable dependiente y el instrumento es el efecto que éste segundo tiene sobre la variable de interés.
  - Por ejemplo, ANGRIST (1990) utiliza el estimador de Wald para analizar el efecto que tuvo sobre el salario de los militares sus servicios durante la guerra de Vietnam. La endogeneidad se deriva de que los militares probablemente compartan más rasgos que influyan sobre su salario que simplemente haber participado en la guerra. Por ello, ANGRIST utiliza un proceso de lotería que hacía que aquellos que eran elegidos fueran más propensos a participar en la guerra<sup>8</sup>.

#### Identificación de las variables instrumentales

- En la literatura de la economía aplicada es común utilizar como variable instrumental cuestiones institucionales que dan lugar a *contextos cuasiexperimentales*.
  - Por ejemplo, en el caso de la educación, ANGRIST y KRUEGER (1991) explotan la fecha de nacimiento de los individuos para estudiar los efectos de una educación más larga. En particular, los niños nacidos en el último trimestre entran en el colegio antes de cumplir los 6 años, mientras que aquellos nacidos en el primer trimestre entran a la edad de los 6 y medio. Esto, junto con el derecho a abandonar la educación obligatoria a partir de los 16 años genera diferencias significativas en los años de educación entre ambos subconjuntos, lo cual se traduce en diferencias significativas estadísticamente en los niveles de salario.

#### Problemas y dificultades

- La principal dificultad del método VI consiste en seleccionar adecuadamente las variables instrumentales  $Z$ , buscando satisfacer las condiciones enunciadas previamente:
  - Condición de exogeneidad: El verdadero error del modelo no es observable, por lo que es difícil estar seguro de que los instrumentos realmente no están correlacionados en el límite con los errores.
  - Condición de relevancia: Los instrumentos de las variables endógenas<sup>9</sup> deben estar muy correlacionados con las variables explicativas  $x$  o, de lo contrario, el estimador tendría una varianza muy elevada.

### Sobreidentificación: mínimos cuadrados bietápicos (MC2E)

#### Idea

- Cuando existe un número mayor de instrumentos que de variables explicativas, entonces decimos que estamos en una situación de sobreidentificación. Existirán varias regresiones posibles con el método de variables instrumentales, todas consistentes, pero con distinta eficiencia (varianza).
  - El método de mínimos cuadrados bietápicos o en 2 etapas (MC2E) ayuda a realizar una estimación con variables instrumentales en situaciones de sobreidentificación. En particular, el estimador MC2E combina de manera eficiente la información de múltiples instrumentos para regresiones sobreidentificadas.

<sup>8</sup> <https://youtu.be/NLgB2WGGKUw> (ver principio hasta 5'35" y luego saltar al minuto 8'23").

<sup>9</sup> Nótese que los mejores instrumentos de las variables exógenas son ellos mismos, pues cumplen perfectamente las 2 condiciones.



Estimador▪ El **método** de MC2E:

- Se estructura en 2 etapas:
  - En una 1ª etapa, se determinan cuáles son las variables explicativas que son exógenas y cuáles son endógenas. A continuación, se construirán regresiones auxiliares de cada variable del modelo sobre las variables exógenas<sup>10</sup>. Obtenemos así el vector de variables explicadas.
  - En la 2ª etapa, se sustituirán en el modelo original las variables explicativas endógenas por sus estimaciones. Esta ecuación será estimada por MCO mediante el sistema de ecuaciones normales, ya conocido.
- Este método podrá emplearse para la estimación de cualquier ecuación que esté *exactamente identificada* o *sobreidentificada*, siendo el 2º caso el que hace que este método tenga mayor interés.
- La expresión del estimador MC2E es:

$$\widehat{\beta}_{MC2E} = [X'Z(Z'Z)^{-1}Z'X]^{-1}[X'Z(Z'Z)^{-1}Z'y]$$

- En el caso de que la ecuación esté exactamente identificada (matriz  $Z$  de tamaño igual a  $X$ ) vemos que esta expresión se reduce al caso más sencillo visto antes de VI.

Propiedades

- Puede demostrarse que el estimador MC2E es el estimador lineal de VI eficiente (es decir, el que tiene una mínima matriz de covarianzas).

**2. VERIFICACIÓN**

- En el momento en el que sospechamos estar en una situación de regresores estocásticos es importante **verificar** 2 cuestiones:
  - a) Exogeneidad de las variables del modelo: Como hemos visto, cuando se sospecha que existen variables endógenas, es necesario utilizar un procedimiento de VI para obtener estimaciones consistentes. Sin embargo, es aconsejable cuestionarse acerca de las propiedades de exogeneidad del resto de las variables explicativas, pues, de no satisfacerse, obtendríamos igualmente estimadores inconsistentes.
  - b) Validez y fiabilidad de los estimadores obtenidos con métodos consistentes: A través de los métodos de contraste de hipótesis y construcción de intervalos de confianza, que no se modifican sustancialmente respecto a la situación de regresores deterministas.

**2.1. Test de exogeneidad de HAUSMAN y WU**

- Cuando sospechamos que determinadas variables son endógenas, hemos visto que se debe utilizar un procedimiento de variables instrumentales para obtener estimaciones consistentes.
  - Sin embargo, es aconsejable verificar que las propiedades de exogeneidad del resto de variables se satisfacen, pues, de lo contrario, obtendríamos igualmente estimadores inconsistentes.
- El **problema** es que *la exogeneidad no puede ser testada directamente* porque:
  - La condición  $E[\varepsilon \cdot x] = 0$  no puede ser observada directamente (el parámetro poblacional  $\varepsilon$  es desconocido).
  - Utilizar los residuos de MCO también es inútil, puesto que, por construcción, se cumple siempre que  $E[\hat{\varepsilon} \cdot x] = 0$ , incluso en presencia de endogeneidad.
- **¿Cuál es entonces la solución?** Se puede comprobar la exogeneidad utilizando información adicional, en concreto, variables instrumentales. Ésta es la idea del test de HAUSMAN y WU.
  - Supongamos un modelo en el que se duda de la exogeneidad de  $r$  variables. El contraste consiste en:
    - Estimar el modelo por MCO y obtener la suma residual  $SR_0$ .
    - Estimar el modelo por VI y obtener la suma residual  $SR_1$ .

<sup>10</sup> Con estas regresiones auxiliares, se obtienen estimaciones de las variables explicativas endógenas sin recurrir a la información muestral disponible de dichas variables endógenas.

- El estadístico a estimar es:  $(\widehat{\beta}_{MCO} - \widehat{\beta}_{VI})' \cdot [var(\widehat{\beta}_{VI}) - var(\widehat{\beta}_{MCO})] \cdot (\widehat{\beta}_{MCO} - \widehat{\beta}_{VI}) = \frac{SR_0 - SR_1}{\sigma^2} \sim \chi_r^2$ 
  - Se distribuye como una chi-cuadrado con  $r$  grados de libertad bajo la hipótesis nula de que todas las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría tal supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.
- **Intuición:** La lógica del test de Hausman es la siguiente:
  - Bajo la hipótesis nula, tanto el estimador de MCO como el de VI son consistentes (aunque este último no eficiente). Por tanto, la diferencia  $d = \widehat{\beta}_{MCO} - \widehat{\beta}_{VI}$  converge en probabilidad a 0.
  - Bajo la hipótesis alternativa, el estimador de MCO será inconsistente, por lo que dicha diferencia será no nula.

## 2.2. Contrastación de hipótesis y construcción de intervalos de confianza

### Idea

- Hasta ahora, hemos obtenido estimaciones puntuales de los verdaderos parámetros poblacionales. Estos estimadores pueden considerarse “mejores predicciones” (para inferencia o inducción). A partir de aquí, muchas veces, el investigador está interesado en realizar:
  - Contrastación de hipótesis: Sirven para evaluar la significatividad (tanto individual como conjunta de los estimadores) y la estabilidad del modelo.
  - Construcción de intervalos de confianza: Rango de valores posibles que el parámetro poblacional podría tomar con cierta probabilidad.
- Los métodos generales de contrastes de hipótesis y de construcción de intervalos de confianza siguen siendo válidos en caso de regresores estocásticos, siempre que el método de regresión utilizado sea consistente, por lo que veremos la formulación general y posteriormente, mencionaremos algunas consideraciones específicas para el caso de estimadores MC2E.
  - Seguimos asumiendo que el término de error del modelo sigue una distribución normal (S6). Dicho supuesto permite obtener las distribuciones exactas de los contrastes estadísticos.

### Contrastes de hipótesis

#### Formulación general

- Nos vamos a centrar en la inferencia estadística sobre el vector de coeficientes  $\beta$ . Los pasos a seguir serían los siguientes:
  - i. Identificar de manera clara y consistente:
    - La hipótesis nula  $H_0$ , que nos da el modelo restringido.
    - La hipótesis alternativa  $H_1$ , que nos da el modelo no restringido.
  - ii. Elegir un test estadístico adecuado. Se trata de una función que depende sólo de la información muestral (no de ningún parámetro desconocido) y del que conocemos su distribución bajo  $H_0$ .
  - iii. Elegir una región crítica para el estadístico (teniendo en cuenta el nivel de significación deseado), dentro del cual rechazaremos  $H_0$  y fuera del cual no la rechazaremos.
- Existen distintas **estrategias** para llevar a cabo tales contrastes. Estas estrategias confieren determinadas características al contraste, que son de especial interés en inferencia. En concreto, para cualquier contraste, es importante evaluar su:
  - Nivel de significación ( $\alpha$ ): Probabilidad de cometer un error de tipo I, es decir, de rechazar la hipótesis nula a pesar de ser ésta cierta. Su complemento  $1 - \alpha$  es el nivel de confianza del contraste.

- **Probabilidad de error tipo II** ( $\beta$ ): Probabilidad de cometer un **error de tipo II**, es decir, de mantener la hipótesis nula a pesar de ser ésta falsa. Su complemento  $1 - \beta$  es la **potencia** del contraste<sup>11</sup>.
  - Lógicamente, el investigador quiere que un contraste tenga un gran nivel de confianza y una gran potencia al mismo tiempo.
  - Sin embargo, en general, sólo puede reducirse el error de tipo I de un contraste a costa de aumentar el error de tipo II, y viceversa, por lo que existirá un *trade-off*.

	no rechazar $H_0$	rechazar $H_0$
$H_0$ es cierta	Correcto Probabilidad $1 - \alpha$	Error tipo I Probabilidad $\alpha$
$H_0$ es falsa	Error tipo II Probabilidad $\beta$	Correcto Probabilidad $1 - \beta$

- En el MLG, se puede realizar un contraste acerca de un conjunto de restricciones lineales con la siguiente **formulación general**:

$$H_0: R\beta = r$$

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

donde:

- $R$  es una matriz  $q \times K$ , siendo:
  - $q$  el número de restricciones contrastadas.
  - $K$  el número de variables del modelo.
- $\beta$  es el vector de coeficientes del modelo.
- $r$  es un vector columna de dimensión  $q$  formado por constantes conocidas.
  - Se asume que:
    - $q \leq K$ .
    - $\text{rango}(R) = q$  (es decir, las restricciones son linealmente independientes).

#### Contraste de hipótesis sobre un conjunto de coeficientes del modelo (estadístico de Wald)

- **Contraste:** A partir de las hipótesis de distribución normal del estimador, puede **construirse el estadístico de Wald**, que refleja la distancia (ponderada por la varianza del estimador) entre  $R\hat{\beta}$  y  $r$  que se distribuirá como una  $\chi^2$  con  $q$  grados de libertad bajo la hipótesis nula.

- Por tanto, la expresión:  $W = (R\hat{\beta} - r)' [\sigma^2 R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r) \sim \chi_q^2$
- Como  $\sigma^2$  no es observable (no conocemos la varianza poblacional de los errores), se debe utilizar su estimación de MCO,  $\hat{\sigma}^2$ . A partir de ahí, podemos obtener la siguiente expresión:

$$\frac{1}{q} (R\hat{\beta} - r)' [\hat{\sigma}^2 R(X'X)^{-1} R']^{-1} (R\hat{\beta} - r) \sim F_{q, N-K}$$

que se **distribuye como una distribución  $F_{q, N-K}$**  bajo la hipótesis nula.

- **Intuición:** La **lógica** de este contraste es la siguiente:
  - Si  $H_0$  es cierta, entonces el valor del estadístico debería ser **próximo a 0** (la distancia entre  $R\hat{\beta}$  y  $r$  será reducida).
  - Por tanto, la hipótesis nula será rechazada si el valor obtenido para el estadístico  $F$  es superior al valor dado por los **valores tabulados** de la distribución  $F$  (con  $q$ ,  $n - k$  grados de libertad).
- **¿Por qué no contrastar los coeficientes de uno en uno?:** Esta estimación no resultaría fiable. Dado que esta pregunta implicaría 2 variables aleatorias ( $t_1$  y  $t_2$ ) para caracterizar el test, se debería asumir una distribución conjunta normal bivalente. De hacerlo una por una, no se contemplaría la posibilidad de haber cometido un error de tipo I en los anteriores contrastes.
- Aunque el test de Wald es el más habitual, existen **otros 2 métodos** de contraste de hipótesis:
  - **Ratio de verosimilitud:** Compara los valores de las funciones maximizadas de verosimilitud (o su logaritmo), bajo  $H_0$  y bajo  $H_1$ . Cuanto más grande sea la diferencia, más probable es que se rechace la hipótesis nula. Con errores normales, esto es equivalente a comparar las sumas residuales de modelo restringido y sin restringir.

<sup>11</sup> Un contraste es **insesgado** cuando su potencia es mayor que su nivel de significación ( $1 - \beta > \alpha$ ) para todos los posibles valores del parámetro desconocido.

- **Multiplicador Lagrangiano:** Cuantifica cómo de lejos se encontraría de 0 la derivada de la función de verosimilitud respecto al parámetro si se cumpliera la hipótesis nula (por definición, con el estimador de MV es 0 por la CPO).

#### **Caso particular: contraste de significación global del modelo**

Un test muy frecuente es el de significatividad conjunta del modelo, el cual contrasta la siguiente hipótesis nula:

$$H_0: \beta^* = 0_{K-1}$$

siendo  $\beta^*$  un vector columna de dimensión  $K-1$  (excluyendo el término independiente<sup>12</sup>).

Para este caso particular, el estadístico  $F$  adopta la siguiente representación alternativa:

$$F = \frac{R^2/(K-1)}{(1-R^2)/(N-K)}$$

distribuido como una  $F_{K-1, N-K}$ . Si el estadístico es suficientemente grande, se rechazará la hipótesis nula (es decir, el modelo sería globalmente válido). Nótese que esto dependerá, principalmente, del valor que tome el  $R^2$  (coeficiente de determinación).

#### Contraste de hipótesis sobre un coeficiente

- En el caso particular de que se quiera contrastar, únicamente, si un coeficiente del modelo toma o no un valor determinado, la matriz  $R$  pasará a ser un vector fila:

$$R = (0 \ 0 \ \dots \ 1 \ 0 \ \dots \ 0)$$

donde el número 1 ocupa la posición  $i$ -ésima correspondiente al coeficiente  $\beta_i$  que se quiere contrastar. Asimismo, tendremos que el término  $r$  es un escalar:

$$r = \beta_i^0$$

donde  $\beta_i^0$  es el valor que toma el coeficiente  $\beta_i$  para la hipótesis nula.

- Por tanto, el test puede presentarse de la siguiente manera:

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

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

- Siendo el estadístico:

$$t = \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}(\hat{\beta}_i)}$$

El cual seguirá una distribución  $t$  de Student con  $n-k$  grados de libertad<sup>13</sup>.

#### **Caso particular: contraste de significatividad individual**

El contraste para un solo coeficiente más habitual es el contraste de significatividad individual, en el que se contrasta si el parámetro de interés es estadísticamente no significativo ( $r = 0$ ). Tendríamos:

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0$$

Para rechazar la hipótesis nula de no significatividad del parámetro, el estadístico  $t$  del coeficiente estimado deberá ser mayor que el valor crítico fijado. Todo dependerá de la relación entre el valor que tome el coeficiente estimado y el de su desviación típica estimada:

$$t = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta}_i)}$$

#### Intervalos y regiones de confianza

- **Definición:** Otro elemento fundamental de la inferencia estadística es la construcción de intervalos de confianza, es decir, pares de valores entre los cuales se estima que se encontrará el valor desconocido de  $\beta$  con una cierta probabilidad. Un intervalo de confianza bilateral al  $x\%$  para el coeficiente  $\beta_j$  contiene el verdadero valor de  $\beta_j$  con una cierta probabilidad del  $x\%$ . Es decir, que contiene el valor de  $\beta_j$  en el  $x\%$  de todas las muestras escogidas al azar.
- **Estimación:** Partiendo del estadístico  $t$  anterior, puede construirse un intervalo con un nivel de confianza de  $(1 - \alpha)\%$ , siendo  $\alpha$  el nivel de significación:

$$-\lambda_{\alpha/2} \leq \frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}(\hat{\beta}_i)} \leq \lambda_{\alpha/2}$$

<sup>12</sup> La significación global del modelo no debe depender de la significatividad del término independiente, por lo que sólo se contrasta la significatividad conjunta de los  $k-1$  coeficientes restantes.

<sup>13</sup> Nótese que  $R\beta - r = \hat{\beta}_i - \beta_i^0$ .

$$\hat{\beta}_i - \lambda_{\alpha/2} \cdot \hat{\sigma}(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + \lambda_{\alpha/2} \cdot \hat{\sigma}(\hat{\beta}_i)$$

donde  $\lambda_{\alpha/2}$  son los percentiles de la distribución  $t$  de Student.

- Uno de los IC más empleados es el intervalo con nivel de confianza al  $(1 - \alpha) = 95\%$ :

$$IC_{95\%}(\beta_i) = \hat{\beta}_i \pm \lambda_{0,025} \cdot \hat{\sigma}(\hat{\beta}_i)$$

- En este caso, el valor crítico cuando la muestra es grande ( $N$  alto), es igual a  $\lambda_{0,025}$  es aproximadamente igual a 1,96.

#### ▪ **Contraste:**

- El intervalo de confianza permite también realizar un contraste en el que  $H_0: \beta_i = \beta_i^0$  al nivel de significación  $\alpha$ . Si  $\beta_i^0$  se encuentra dentro del intervalo obtenido, no se rechazará la hipótesis nula; en caso contrario, sí se rechazará.
- Razonando de manera análoga, pueden obtenerse regiones de confianza. En este caso el objetivo sería encontrar un rango de valores para varios coeficientes simultáneamente. Partiríamos del estadístico de Wald visto antes (distribuido como una  $F_{q,N-K}$ ).

#### Consideraciones específicas para estimadores MC2E

- Como decíamos, los métodos generales de contrastes de hipótesis y de construcción de intervalos de confianza siguen siendo válidos en caso de regresores estocásticos, siempre que el método de regresión utilizado sea consistente.
- En cualquier caso, cuando hayamos aplicado MC2E y queramos contrastar la significatividad de los coeficientes, deberemos prestar atención a 2 cuestiones a la hora de calcular los errores estándares adecuados para posteriormente realizar inferencia estadística:
  - Los errores no podrán estimarse utilizando los residuos de la regresión de la 2ª etapa, que son los que arrojan los programas informáticos habitualmente, pues éstos infraestiman el verdadero error (los residuos de la 2ª etapa son  $\hat{\varepsilon}^* = y - \hat{X}\hat{\beta}$ , pero, en realidad, los verdaderos residuos a utilizar para calcular la varianza del estimador MC2E deben incorporar el error que se comete también en la 1ª etapa al estimar la matriz de observaciones  $(X - \hat{X})$ ). Por tanto, el verdadero residuo de la regresión MC2E es  $\hat{\varepsilon} = y - X\hat{\beta}$ .
  - El término error del MC2E podrá ser o no heterocedástico. Por ello, es importante utilizar versiones de los errores estándar heterocedástico-robustos por las mismas razones que justifican su necesidad bajo MCO.

### 3. PREDICCIÓN

#### 3.1. Cálculo de predicciones

- En los años 30, WIENER y KOLMOGÓROV demostraron que, si el criterio es la *minimización del error cuadrático medio de predicción* (ECMP), entonces el *predictor óptimo* es la *esperanza condicionada a la información disponible*. Por tanto, con la información disponible hasta el instante  $T$  ( $\Omega_T$ ), la función a utilizar para predecir  $y_{T+1}$  es:

$$\begin{aligned}\hat{y}_{T+1} &= E[y_{T+1} | \Omega_T] \\ \hat{y}_{T+1} &= E[x_{T+1}' \cdot \beta | X_T] + E[\varepsilon_{T+1} | X_T] \\ \hat{y}_{T+1} &= E[x_{T+1}' | X_T] \cdot \hat{\beta}_T + E[\varepsilon_{T+1} | X_T]\end{aligned}$$

- Por tanto, la predicción de la variable endógena requiere estimar **3 parámetros desconocidos**:
  1. Predicción del vector de coeficientes: Utilizamos el vector  $\hat{\beta}_T$ , es decir, el vector de estimaciones MCO obtenido con el tamaño muestral hasta  $T$ . Aquí es clave el *supuesto de estabilidad*: la mejor relación estimada entre la variable endógena y las variables explicativas durante el periodo muestral sigue siendo válida en el periodo  $T + 1$  y posteriores.
  2. Predicción de las variables explicativas  $x_{T+1}$ . Podemos distinguir 2 casos:
    - Los valores futuros de  $x$  se conocen de antemano (no hay necesidad de predecirlos). 2 ejemplos:
      - Si el modelo se ha especificado para explicar la cifra de ventas de una determinada empresa como función del precio de su producto, la empresa fija la variable

explicativa (precio) para el siguiente período y utiliza ese precio conocido para predecir el volumen de ventas.

- Los salarios reales, si sus valores para los 2 años próximos han sido fijados en virtud de un convenio colectivo, serían otro ejemplo de una variable explicativa (en una ecuación de empleo, por ejemplo) cuyos valores futuros son conocidos de antemano.
  - *Los valores futuros de las variables explicativas se desconocen*, por lo que habría que estimarlos previamente. Estas estimaciones pueden estar basadas simplemente en nuestra intuición o ser algo más rigurosas (p.ej. modelos ARIMA [ver tema 5.B.6]).
- 3. Predicción de  $\varepsilon_{T+1}$ : Como hemos supuesto que el término de error es una sucesión de variables aleatorias independientes entre sí,  $\varepsilon_{T+1}$  es independiente de la información muestral disponible en  $T$ . En consecuencia, predecimos  $\varepsilon_{T+1}$  por medio de su esperanza matemática, es decir:  $E[\varepsilon_{T+1}|\mathbf{X}_T] = E[\varepsilon_{T+1}] = 0$ .

- En resumen, bajo nuestros supuestos, la estimación MCO es:

$$\hat{y}_{T+1} = \mathbf{x}_{T+1}' \cdot \hat{\boldsymbol{\beta}}_T$$

- Podemos deducir una serie de **condiciones necesarias para que la predicción sea fiable**:
  - Supuesto de estabilidad: La relación lineal estimada entre variable endógena y variables explicativas se mantiene en el futuro.
  - Coefficientes suficientemente estables como para que sus estimaciones obtenidas con la muestra actual sean una buena aproximación a los valores que se obtendrían al incorporar observaciones futuras a la muestra.
  - Conocimiento de los valores futuros de las variables  $x$  o, alternativamente, que los modelos de predicción que utilicemos para dichas variables sean suficientemente fiables.
  - Especificación correcta del modelo lineal que hemos estimado.
  - Horizonte de predicción no muy lejano.

### 3.2. Error de predicción

- **Concepto**: El error de predicción un período hacia el futuro se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida:

$$\begin{aligned} e_T(1) &= y_{T+1} - E_T[y_{T+1}] \\ &= \mathbf{x}_{T+1}' \cdot \boldsymbol{\beta} + \varepsilon_{T+1} - \mathbf{x}_{T+1}' \cdot \hat{\boldsymbol{\beta}}_T \\ &= \mathbf{x}_{T+1}' \cdot (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) + \varepsilon_{T+1} \end{aligned}$$

- Nótese que, en el instante  $T$ , la variable  $e_T(1)$  es una variable aleatoria con valor desconocido, puesto que su realización ocurrirá en el período  $T + 1$ .
- Puede verse en esta expresión cuáles son las fuentes de error de predicción:
  - Error en la estimación del vector  $\boldsymbol{\beta}$ .
  - Error en la predicción del vector  $\mathbf{x}_{T+1}$ .
  - Error estocástico inherente al modelo  $\varepsilon_{T+1}$ .
- **Propiedades**: Como el estimador MCO es insesgado, entonces se tiene que, bajo nuestros supuestos, el error de predicción tiene esperanza nula:  $E[e_T(1)] = 0$ .
  - En este sentido, se dice que cuando las variables  $\mathbf{x}_{T+1}$  son conocidas de antemano, entonces la predicción obtenida a partir del estimador MCO del modo que aquí hemos descrito es insesgada.

### 3.3. Varianza del error de predicción e intervalos de confianza

- Ningún problema de estimación es completo si la estimación que se obtiene no viene acompañada de un intervalo de confianza, y la predicción no es diferente en este respecto. Para obtener dichos intervalos, es necesario calcular la varianza del error de estimación.

$$\text{Var}(e_T(1)) = E[e_T(1) \cdot e_T(1)']$$

## CONCLUSIÓN





### 5.B.3 : AUTOCORRELACIÓN, MULTICOLINEALIDAD Y HETEROCEDASTICIDAD: CONSECUENCIAS, ORIGEN, DETECCIÓN Y SOLUCIONES.

#### INTRODUCCIÓN

##### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

##### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

▪ **Contextualización:**

- Desde un punto de vista histórico, desde las primeras aportaciones de GAUSS a principios del s. XIX, el modelo más utilizado y estudiado por los econométricos ha sido el modelo lineal general (o modelo de regresión clásico).
  - Este modelo ha sido aplicado con gran éxito en el estudio de un elevado nº de temas económicos. Además, sirve de base para el desarrollo de modelos más complejos que levantan los supuestos clásicos, que pueden ser restrictivos en determinadas aplicaciones (series temporales, relaciones no lineales...).
  - El MLG puede estimarse a través de varios métodos, entre los que destacaremos los 2 más importantes:
    - El método de mínimos cuadrados ordinarios (MCO); y
    - El método de máxima verosimilitud (MV).
- En esta exposición, nos vamos a centrar en una **desviación de los supuestos clásicos: autocorrelación, multicolinealidad y heterocedasticidad**. Y es que, con frecuencia, los datos disponibles para estimar un modelo de regresión no se adaptan exactamente a los supuestos que subyacen al MLG.

▪ **Problemática (Preguntas clave):**

–

▪ **Estructura:**

<b>0. CONSIDERACIONES PREVIAS</b>
0.1. Supuestos clásicos del Modelo Lineal General (MLG)
Supuesto 1: Linealidad del modelo
Supuesto 2: Rango completo (no multicolinealidad perfecta)
Supuesto 3: Esperanza condicionada nula de las perturbaciones
Supuesto 4: Perturbaciones esféricas
Supuesto 5: Regresores deterministas
Supuesto 6: Distribución normal de los errores
0.2. Método de Mínimos Cuadrados Ordinarios (MCO)
0.3. Violación de los supuestos
<b>1. MATRIZ X NO ES DE RANGO COMPLETO: MULTICOLINEALIDAD</b>
1.1. Definición
1.2. Origen
1.3. Consecuencias
1.4. Detección
1.5. Soluciones
<b>2. PERTURBACIONES NO ESFÉRICAS</b>
2.1. Definición y estimador MCG
2.1.1. Definición
2.1.2. Consecuencias
2.1.3. Estimador de Mínimos Cuadrados Generalizados (MCG)
2.2. Heterocedasticidad
2.2.1. Definición
2.2.2. Origen
2.2.3. Consecuencias
2.2.4. Detección
2.2.5. Soluciones
2.3. Autocorrelación
2.3.1. Definición
2.3.2. Origen
2.3.3. Consecuencias
2.3.4. Detección
2.3.5. Solución

## 0. CONSIDERACIONES PREVIAS

Esto muy rápido en este tema.

### 0.1. Supuestos clásicos del Modelo Lineal General (MLG)

- El **Modelo Lineal General (MLG)** se basa en un conjunto de supuestos sobre cómo los datos muestrales serán generados en un proceso subyacente de generación. Como veremos, bajo estos supuestos del MLG, los estimadores cumplen una serie de “buenas propiedades” si se cumplen los supuestos clásicos enunciados originalmente por GAUSS a principios del s. XIX.
  - Se utilizará como referencia de “*Econometric Analysis*” realizada por GREENE para la explicación.

#### Supuesto 1: Linealidad del modelo

- El modelo que relaciona la variable endógena con las variables explicativas es **lineal en los coeficientes  $\beta$** .
  - Por tanto, el modelo se expresa del siguiente modo:
 
$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ki}, \varepsilon_i) = \beta_0 + \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + \dots + \beta_k \cdot x_{ki} + \varepsilon_i \quad \forall i = 1, \dots, N$$
 donde,
    - $\beta_0$  es el término independiente o *intercepto*. El término independiente puede interpretarse como la esperanza de la variable dependiente cuando el resto de las variables es nulo o puede interpretarse aquel que acompaña a una primera variable explicativa  $x_{0i}$  cuyo valor es siempre igual a 1.
    - $\beta_1, \dots, \beta_k$  son los coeficientes o *pendientes* de la regresión. Recogen la magnitud del impacto de cada una de las variables explicativas sobre la variable endógena.

#### Supuesto 2: Rango completo (no multicolinealidad perfecta)

- $X$  es una matriz de dimensión  $N \times K$  con rango  $K$ , por lo que es una matriz con rango completo. Esto requiere que se cumplan 2 condiciones:
  - Las columnas de  $X$  son linealmente independientes, lo que implica que las variables explicativas son linealmente independientes, es decir, ninguna variable explicativa es combinación lineal de otra.
  - Condición de identificación: Disponemos de al menos  $K$  observaciones ( $N \geq K$ )<sup>3</sup>.

#### Supuesto 3: Esperanza condicionada nula de las perturbaciones

- Para todo el conjunto de observaciones, se asume que los términos de error tienen esperanza condicionada a  $X$  nula:
  - Ello implica que la esperanza de cada error, condicionado a la información contenida en el vector  $X$  es nula, lo que significa que ninguna observación de  $x$  provee información sobre el valor esperado del error (son ortogonales).
  - Por la ley de expectativas iteradas se obtiene que  $E[\varepsilon_i | X] = E_X[E[\varepsilon_i | X]] = E_X[0] = 0 \quad \forall i$  esto implica<sup>4</sup> que la media no condicionada del término error también es nula. Por otro lado, el supuesto también implica lo siguiente:  $Cov[X_i, \varepsilon_i] = 0 \quad \forall i = 1, \dots, N$
  - Este último supuesto “débil”, conocido como exogeneidad de las variables explicativas, garantiza la consistencia del método MCO. Sin embargo, su incumplimiento en casos como, por ej., la presencia de variables omitidas, errores de medida o presencia de retardos de la variable endógena como variable explicativa, puede generar problemas de endogeneidad y requiere un tratamiento especial de los datos.

<sup>3</sup> Si el nº de observaciones fuera menor que  $K$ ,  $X$  no podría tener rango completo. En principio, basta con que  $N = K$  para que  $X$  pueda llegar a tener rango  $K$ . Sin embargo, cuanto mayor sea  $N$ , más precisas tenderán a ser nuestras estimaciones. La diferencia  $(N - K)$  se conoce como “grados de libertad” de la regresión.

<sup>4</sup> La relación no es cierta en el sentido contrario.

#### Supuesto 4: Perturbaciones esféricas

- El supuesto de perturbaciones esféricas queda resumido en la siguiente forma de la matriz de varianzas-covarianzas de los términos de error (matriz escalar):

$$E[\varepsilon \cdot \varepsilon' | \mathbf{X}] = \sigma^2 \cdot \mathbf{I} = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{bmatrix}$$

- Este supuesto contiene, por tanto, 2 sub-supuestos:
  - Homocedasticidad: Todos los términos de error tienen la misma varianza:

$$\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma^2 \quad \forall i = 1, \dots, N$$

- No autocorrelación: No existe correlación serial entre las perturbaciones:

$$\text{Cov}[\varepsilon_i, \varepsilon_j | \mathbf{X}] = 0 \quad \forall i \neq j$$

- Todo ello se resume en el siguiente supuesto:

$$E[\varepsilon \cdot \varepsilon' | \mathbf{X}] = \sigma^2 \cdot \mathbf{I} \rightarrow \text{Var}[\varepsilon] = E[\text{Var}[\varepsilon | \mathbf{X}]] + \text{Var}[E[\varepsilon | \mathbf{X}]] = \sigma^2 \cdot \mathbf{I}$$

#### Supuesto 5: Regresores deterministas

- Un supuesto habitual del MLG es que las variables explicativas  $x$  son deterministas o no estocásticas.
  - Es decir, se supone que las observaciones de las variables explicativas son fijas en muestras repetidas, como sucedería, por ejemplo, en una situación experimental, en la que el científico elige los valores de las variables explicativas y luego observa  $y_i$ .
    - Nótese, por tanto, que el supuesto no incluye a la variable endógena  $y$  (que es aleatoria por ser función del término de error).
  - Sintetizando, el supuesto establece que  $\mathbf{X}$  es una matriz  $N \times K$  de constantes conocidas.
  - Este supuesto garantiza, en coherencia con el S3, que la esperanza condicionada a  $x$  de  $\varepsilon$  es nula, lo cual, como veremos después, es crucial para que los estimadores habituales sean insesgados. Por tanto, podríamos relajar el supuesto de regresores deterministas a un bajo coste puesto que, en definitiva, el supuesto crucial es el S3.

#### Supuesto 6: Distribución normal de los errores

- Es conveniente añadir el supuesto de que los términos de error se distribuyen como una normal, con media cero y varianza constante, de tal manera que:  $\varepsilon | \mathbf{X} \sim N(0, \sigma^2 \cdot \mathbf{I})$ .
  - Este supuesto es razonable si tenemos en cuenta que las condiciones para la aplicación del Teorema Central del Límite son generalmente aplicables en este caso. Es decir, aunque los errores tuvieran una distribución distinta, este teorema asegura que, en muestras grandes, su distribución tenderá a la Normal. Ello implica que los errores son estadísticamente independientes e incorrelados.
  - Aunque no es un supuesto imprescindible, es útil para el posterior contraste de hipótesis.

#### 0.2. Método de Mínimos Cuadrados Ordinarios (MCO)

- Método:**

- El estimador MCO es aquel que minimiza la suma de los residuos al cuadrado:

$$\min_{\{\hat{\beta}\}} SR(\hat{\beta}) = \hat{\varepsilon}' \cdot \hat{\varepsilon}$$

- Resolviendo el anterior problema, se encuentra la expresión clave para el estimador de MCO (vector columna de dimensión  $K$ ):

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

▪ **Propiedades:** Dados, los supuestos del MLG, el estimador MCO será:

- Insesgado: Dado que  $y$  es una variable estocástica afectada directamente por un componente error inobservable  $\varepsilon$ , la estimación se verá afectada. No obstante, tomando esperanzas sobre el beta estimado ( $\hat{\beta}$ ), constatamos que éste coincide con el beta poblacional ( $\beta$ ):

$$\left. \begin{array}{l} \hat{\beta} = (X'X)^{-1}X'y \\ y = X\beta + \varepsilon \end{array} \right\} E[\hat{\beta}] = E[(X'X)^{-1}X'(X\beta + \varepsilon)] = \underbrace{E[(X'X)^{-1}X'X\beta]}_{=E[\hat{\beta}]=\beta} + \underbrace{E[(X'X)^{-1}X'\varepsilon]}_{=((X'X)^{-1}X') \cdot E[\varepsilon]=0} = \beta$$

- Eficiente: Bajo los supuestos del MLG, el estimador de MCO es el estimador lineal insesgado óptimo (*Best Linear Unbiased Estimator, BLUE*) por el Teorema de Gauss-Márkov. Es decir, tiene una varianza menor o igual a cualquier otro estimador lineal o insesgado. En otras palabras, el estimador MCO es eficiente.
- Distribuido normalmente: Asumiendo que  $\varepsilon$  se distribuye como una normal multivariante (S6), cada elemento de  $\hat{\beta}|X$  se distribuirá normalmente (por el teorema del mapeo continuo de Mann-Wald, pues es una combinación lineal de elementos distribuidos normalmente):

$$\hat{\beta}|X \sim N[\beta, \sigma^2(X'X)^{-1}]$$

- Asintóticamente consistente: A medida que el tamaño muestral aumenta hasta el infinito, el beta estimado converge con el beta poblacional.

### 0.3. Violación de los supuestos

- Con frecuencia, los datos disponibles para estimar un modelo de regresión no se adaptan exactamente a los supuestos que subyacen al MLG:
  - La multicolinealidad perfecta supone una violación del S2. La multicolinealidad imperfecta no supone la violación de ningún supuesto, pero generará problemas de precisión del estimador.
  - La heterocedasticidad y autocorrelación suponen la violación del S4.

## 1. MATRIZ X NO ES DE RANGO COMPLETO: MULTICOLINEALIDAD

### 1.1. Definición

- La **multicolinealidad**<sup>5</sup> se da cuando existe, en alguna medida, correlación entre distintas variables explicativas del modelo.
  - Multicolinealidad perfecta: Una de las variables explicativas es combinación lineal determinista de, al menos, otra de las restantes. Es decir, se viola el supuesto de rango completo de la matriz  $X$  (S2), ya que la matriz  $X$  es  $N \times K$  y su rango  $L < K$  (al ser las columnas linealmente dependientes).

$$x_j = \sum \lambda_n \cdot x_n$$

- Este problema es poco habitual y suele manifestarse ante especificaciones incorrectas del modelo. Un ejemplo es la *trampa de los dummies* (p.ej. incluir, además del intercepto, una *dummy* para personas que sí utilizan gafas y otra para personas que no las utilizan).
- Multicolinealidad aproximada: No es combinación lineal determinista, pero existe alguna relación. Este caso es mucho más habitual y provoca un problema de carácter estadístico: la pérdida de precisión en las estimaciones.

$$x_j = \sum \lambda_n \cdot x_n + u_i$$

- Por ejemplo, la alta correlación entre la educación de un individuo y la de sus padres.
- Evidentemente, la correlación entre las variables económicas suele ser habitual. La cuestión reside en saber discernir cuándo este nivel de correlación puede llegar a originar problemas.

<sup>5</sup> El término multicolinealidad se atribuye a RAGNAR FRISCH.

## 1.2. Origen

- Podemos destacar las siguientes **causas** de multicolinealidad:
  - i) Restricciones en el modelo o en la población objeto de muestreo: Por ejemplo, en la regresión del consumo de electricidad sobre el ingreso ( $X_2$ ) y el tamaño de las viviendas ( $X_3$ ) hay una restricción física en la población, pues las familias con ingresos más altos suelen habitar viviendas más grandes que las familias con ingresos más bajos.
  - ii) Especificación del modelo: Por ejemplo, la adición de términos polinomiales a un modelo de regresión, en especial cuando el rango de la variable  $X$  es reducido.
  - iii) Un modelo sobredimensionado: Esto sucede cuando el modelo tiene más variables explicativas que número de observaciones.
  - iv) Tendencia común: Especialmente común en las series temporales. Por ejemplo, ingresos y riqueza suelen evolucionar de forma paralela.

## 1.3. Consecuencias

- Las consecuencias de la multicolinealidad perfecta son claras:
  - La matriz de variables explicativas,  $X$ , no es de rango completo (se viola el S2 de rango completo). Como consecuencia, no existe la inversa de  $(X'X)$  y el sistema de ecuaciones normales no tiene una solución única.
  - Los coeficientes de regresión de las variables  $X$  son indeterminados y sus errores estándar infinitos.
- En cuanto a las implicaciones de la multicolinealidad aproximada:
  - La inversa de  $(X'X)$  sí existe y el sistema de ecuaciones normales tiene una *solución única*, que seguirá siendo el estimador lineal insesgado óptimo (ELIO, por el teorema de Gauss-Márkov).
  - Sin embargo, incorporar o reducir un reducido número de observaciones muestrales, aunque suponga pequeños cambios en la matriz  $X$  y el vector  $y$ , probablemente ocasionará grandes cambios en la solución del sistema ( $\hat{\beta}$ ).
    - Cuanto más próxima esté la matriz  $(X'X)$  a ser singular<sup>6</sup>, más pequeña será ésta (quedará reflejado en su norma) y, por tanto, mayor será la matriz de varianzas-covarianzas (recuérdese que la matriz  $(X'X)$  entra como inversa).
  - En definitiva, la multicolinealidad aproximada no implica la violación de ningún supuesto pero *se perderá precisión en las estimaciones*. Ello implicará que:
    - i) Aunque todos los estimadores de MCO son ELIO, presentan varianzas y covarianzas grandes que dificultan la estimación precisa. Matemáticamente, ello se puede ver en la fórmula de la varianza del estimador en una regresión con 2 variables explicativas:
 
$$var[\hat{\beta}_2] = \frac{\sigma^2}{\sum x_{2i}^2 \cdot (1 - r_{23}^2)}$$
 donde  $r_{23}^2$  es la correlación entre  $X_2$  y  $X_3$ .
    - ii) Debido a esto, los intervalos de confianza tienden a ser mucho más amplios, lo cual propicia una aceptación más fácil de la hipótesis nula ( $H_0$ ).
    - iii) A su vez, la ratio  $t$  de uno o más coeficientes tiende a ser estadísticamente no significativa.
    - iv) A pesar de que los coeficientes no sean significativos, estos modelos tienden a presentar un  $R^2$  elevado.
    - v) Los estimadores de MCO y sus errores estándar serán sensibles a cambios pequeños en los datos.

## 1.4. Detección

- La multicolinealidad perfecta es *fácilmente detectable*, ya que, en presencia de ésta, no se podrá invertir la matriz  $(X'X)$  por ser singular, por lo que el estimador MCO no podrá calcularse.

<sup>6</sup> Una matriz singular es una matriz cuadrada cuyo determinante es igual a cero. Una matriz singular no tiene matriz inversa. En el caso de multicolinealidad perfecta, la matriz  $(X'X)$  sería singular.



- La multicolinealidad aproximada es más difícil de detectar, pues  $(X'X)$  sí es invertible y el estimador mínimo cuadrático puede obtenerse. No obstante, las mismas consecuencias que tiene la multicolinealidad aproximada nos ayudarán a detectar la existencia de este problema. Así, se podrán observar ciertos indicios de multicolinealidad aproximada cuando:
  - Pequeños cambios en los datos produzcan grandes cambios en las estimaciones por MCO.
  - Los coeficientes estimados tienen desviaciones típicas elevadas, por lo que tienden a considerarse no significativos en contrastes de significación individual, a pesar de que son significativos conjuntamente (p.ej.  $R^2$  elevado).

### 1.5. Soluciones

- La solución que se tome ante la presencia de multicolinealidad aproximada dependerá de la **finalidad** que se busque del modelo econométrico:
  - Si la finalidad es *predictiva*, entonces la inclusión de variables redundantes no impide conseguir un buen ajuste global y, con ello, obtener buenas predicciones de la variable endógena. Por tanto, la multicolinealidad no debería de preocuparnos.
  - Si se pretende hacer un *análisis estructural* que contribuya al conocimiento descriptivo de la economía, entonces la multicolinealidad puede suponer un problema grave, debido a que las estimaciones de MCO serán poco precisas.
- Las **soluciones** al problema de la multicolinealidad son variadas, pero generalmente *poco satisfactorias*.
  - Eliminación de una de las variables*, de tal forma que se elimina el origen de la multicolinealidad.
    - No obstante, ello puede derivar en problemas asociados al sesgo de especificación del modelo.
  - Combinación de información de corte transversal de series temporales*:
    - Por ejemplo, si buscamos regresar el número de ventas de automóviles a partir del precio promedio de los mismos y los ingresos de la empresa, estas 2 variables explicativas pueden encontrarse altamente correlacionadas. Por ello, TOBIN sugiere que, en caso de tener datos de panel, utilizar el retardo de una de las 2 variables, eliminando la alta correlación que suele manifestarse en un mismo periodo.
  - Regresión cresta*: El estimador de la regresión cresta trata de aumentar los elementos de la diagonal de  $(X'X)$  en una constante para evitar el problema de su aproximada singularidad.
    - En ese caso el beta estimado pasará a ser  $\hat{\beta}_c = (X'X + c \cdot I)^{-1}X'y$ , el cual si bien estará sesgado podría tener una matriz de varianzas-covarianzas menor.
    - El problema es que la elección de la  $c$  será arbitraria.

## 2. PERTURBACIONES NO ESFÉRICAS

### 2.1. Definición y estimador MCG

#### 2.1.1. Definición

- Pasemos ahora a analizar el incumplimiento del supuesto de perturbaciones esféricas (S4).
  - Es decir, supondremos la matriz de varianzas-covarianzas deja de ser una matriz escalar y tiene, potencialmente  $N \cdot (N + 1)/2$  elementos distintos, lo cual ocurriría si se dan a la vez *heterocedasticidad* y *autocorrelación*. Denotamos dicha matriz  $\Omega$ .

#### 2.1.2. Consecuencias

- Las **consecuencias** del incumplimiento del supuesto de perturbaciones esféricas son:
  - Insesgadez: El estimador MCO sigue siendo insesgado, puesto que esta propiedad no depende de la estructura de la matriz de varianzas-covarianzas de los errores, sino de su esperanza condicionada nula.

- Matriz de varianzas-covarianzas del estimador MCO: Como vimos antes, cuando se cumple el S4, puede deducirse que ésta es igual a  $\sigma^2 \cdot (X'X)^{-1}$ . En ausencia de dicho supuesto, la anterior expresión se encontraría sesgada, por lo que debemos utilizar una formulación más general:

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

- Eficiencia: El estimador de MCO, aun siendo insesgado, ya no es el de mínima varianza, ya que se incumplen las condiciones del *Teorema de Gauss-Márkov*. En este caso, el nuevo ELIO pasa a ser el estimador de mínimos cuadrados generalizados (MCG).

- En resumen, si se cumplen el resto de los supuestos, el estimador MCO sigue siendo insesgado, pero ya no es de mínima varianza, pues se incumplen las condiciones del teorema de Gauss-Márkov. En este caso, el nuevo ELIO es el **estimador de mínimos cuadrados generalizados (MCG)**.

### 2.1.3. Estimador de Mínimos Cuadrados Generalizados (MCG)

- La idea que subyace al estimador de MCG es transformar el modelo original para que vuelva a cumplir con las condiciones del Teorema de Gauss-Márkov.
  - En presencia, por ejemplo, de heterocedasticidad, el estimador usual de MCO, a pesar de ser insesgado deja de ser ELIO. En el caso, por ejemplo, de que quisiéramos analizar los efectos de la renta sobre el consumo, sabemos que la varianza aumenta a medida que aumenta la renta, por lo que los errores no se distribuirán de manera homogénea.
  - Idealmente, querríamos diseñar un esquema de estimación de manera que las observaciones que surgen de poblaciones con mayor variabilidad reciban menos peso que las que provienen de poblaciones con menor variabilidad. Por desgracia, el método de MCO no sigue esta estrategia y, por consiguiente, no aprovecha la información sobre la variabilidad desigual de la variable dependiente. No obstante, los mínimos cuadrados generalizados sí tomarán en cuenta esta información explícitamente.
- La **estimación del estimador MCG** se realiza de la siguiente manera:
  - Se pre-multiplica el modelo original por una matriz  $V^{-1}$ :

$$V^{-1} \cdot y = V^{-1} \cdot X \cdot \beta + V^{-1} \cdot \varepsilon$$

donde la matriz  $V$  descompone la matriz de varianzas-covarianzas del término de error en el modelo original  $\Omega = VV'$ , lo que se conoce como **descomposición de Cholesky**.

$$V^{-1} = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & 1/\sigma_i & \vdots \\ 0 & \cdots & 1/\sigma_N \end{pmatrix}$$

$$V^{-1} \cdot y = \begin{pmatrix} y_1/\sigma_1 \\ y_2/\sigma_2 \\ \vdots \\ y_N/\sigma_N \end{pmatrix} \quad V^{-1} \cdot X = \begin{pmatrix} x_{11}/\sigma_1 & \cdots & x_{K1}/\sigma_1 \\ \vdots & \ddots & \vdots \\ x_{1N}/\sigma_N & \cdots & x_{KN}/\sigma_N \end{pmatrix}$$

- La expresión del estimador MCG sería:  $\widehat{\beta}_{MCG} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y$

- **¿Cuál es el propósito de transformar el modelo original?** A través de la transformación, la varianza del término error es ahora homocedástica. Como aún se conservan el resto de supuestos del MLG, si aplicamos MCO al modelo transformado, los estimadores volverán a ser ELIO.

- Diferencias con entre MCG y MCO:

- Bajo MCG, se minimiza la suma ponderada de los residuos al cuadrado, donde cada observación será ponderada por  $w_i = 1/\sigma_i^2$ , pero en MCO se reduce la suma del cuadrado de los residuos sin ponderar.
- En consecuencia, bajo MCG las observaciones que provienen de un subconjunto con una  $\sigma_i^2$  más grandes tendrán una ponderación relativamente menor, y las de una población con una menor  $\sigma_i^2$  tendrán una ponderación proporcionalmente mayor.

- El **problema del método MCG ideal** es que, en la práctica, no se conocen los elementos de la matriz  $\Omega$ . Por eso, el investigador deberá realizar distintos supuestos sobre la estructura de la *heterocedasticidad* o *autocorrelación* que presenta el modelo, con el objetivo de reducir el nº de parámetros a estimar.

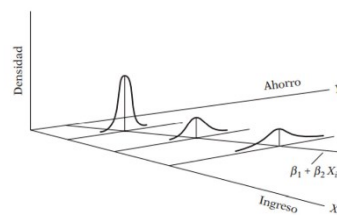
## 2.2. Heterocedasticidad

### 2.2.1. Definición

- La **heterocedasticidad** implica la violación del supuesto de perturbaciones esféricas ya que la varianza del error deja de ser la misma para todo  $i$ :  $Var(\varepsilon_i|X) \neq \sigma^2$ .
  - En este caso, la matriz de covarianzas del término de error,  $\Omega$ , sería una matriz definida positiva y diagonal. Es decir,

$$E[\varepsilon\varepsilon'|X] = \Omega = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \sigma_i^2 & \vdots \\ 0 & \cdots & \sigma_N^2 \end{pmatrix}$$

- Los elementos en la diagonal no son idénticos para todas las perturbaciones ( $\sigma_i^2$  diferentes).
- Pero presenta todo ceros fuera de la diagonal porque suponemos que no hay autocorrelación



### 2.2.2. Origen

- La heterocedasticidad puede surgir por las siguientes **causas**:
  - Escala de las variables: Con datos de sección cruzada donde la escala de las variables dependiente o independientes y el poder explicativo del modelo varían a lo largo de las observaciones.
    - Por ejemplo, en el caso de que quisiésemos estudiar los efectos de la renta sobre el consumo, sabemos que la varianza del consumo y del ahorro aumenta a medida que aumenta la renta, por lo que los errores no se distribuirán de manera homogénea.
  - Aprendizaje de los errores: A medida que la gente aprende, sus errores de comportamiento van disminuyendo con el tiempo. Por ejemplo, los errores cometidos en el teclado a medida que se practica la mecanografía.
  - Técnicas de recolección de datos: A medida que mejoran las técnicas de recolección de datos, es probable que  $\sigma_i^2$  se reduzca.
  - Datos atípicos: La existencia de outliers dentro de la base de datos muestral generará heterocedasticidad en los tramos de la muestra donde se encuentren.
  - Mala especificación del modelo: La heterocedasticidad puede ser el resultado de la omisión de alguna variable explicativa importante.
    - Por ejemplo, si en la estimación de la demanda de un bien se omiten los precios de los bienes complementarios (*sesgo por variable omitida*), los residuos dejarán de tener varianza constante. Al incluir estas variables el problema desaparece.

### 2.2.3. Consecuencias

- Como dijimos, las estimaciones por MCO seguirán siendo insesgadas y consistentes, pero dejarán de ser eficientes, incluso para muestras muy grandes (asintóticamente). Esto se debe a que ya no se cumplen las condiciones para el Teorema Gauss-Márkov.
  - Ello se deduce de que la varianza del estimador pasará a ser  $var(\hat{\beta}) = \sum x_i^2 \sigma_i^2 / (\sum x_i^2)^2$ , lo cual difiere de la varianza usual obtenida bajo homocedasticidad:  $var(\hat{\beta}) = \sigma^2 / \sum x_i^2$ .
- En relación a su **estimación**:
  - Si se realiza la estimación mediante MCO teniendo en cuenta la heterocedasticidad, estimando la varianza del estimador bajo la fórmula señalada, la estimación es insesgada pero la inferencia estadística se va a ver afectada. Las pruebas  $t$  y  $F$  se verán afectadas, ya que la varianza del estimador óptimo (MCG) es inferior a la del MCO. Ello implica que los intervalos de confianza serán innecesariamente grandes. Como resultado, es posible que las pruebas  $t$  y  $F$  den resultados imprecisos.
  - Si se realiza la estimación mediante MCO sin tener en cuenta la heterocedasticidad, la situación se torna más grave, pues la fórmula habitual de varianza para el beta estimado será generará un sesgo.

- En definitiva, la característica más sobresaliente de estos resultados es que los MCO, con o sin corrección por heterocedasticidad, sobreestiman consistentemente el verdadero error estándar obtenido mediante el estimador ELIO de MCG.

#### 2.2.4. Detección

- Como con la multicolinealidad, la heterocedasticidad resulta un problema serio en el análisis econométrico. No obstante, en su detección no existirá una única regla precisa y rápida, sino un conjunto de métodos que nos permitirán intuir la presencia de este tipo de problemas. Por ello, para detectar la heterocedasticidad se pueden emplear diferentes contrastes:

##### a) Método gráfico informal:

- Si el econométra sospecha de la existencia de heterocedasticidad (p.ej. a través de la comprensión del problema analizado en ocasiones se puede intuir), una primera aproximación sería llevar a cabo un análisis de regresión bajo el supuesto de que no existe heterocedasticidad y analizar *a posteriori* los residuos al cuadrado para observar si existe algún patrón.

##### b) Prueba de Park:

- Formaliza el método gráfico, bajo el supuesto de que  $\sigma_i^2$  es algún tipo de función de la variable explicativa  $X_i$  (p.ej. la forma funcional puede ser  $\sigma_i^2 = \sigma^2 + X_i^\beta + e^{v_i}$ , donde si aplicamos una transformación logarítmica se obtiene  $\ln(\sigma_i^2) = \ln(\sigma^2) + \beta \cdot \ln(X_i) + v_i$  y contrastamos  $H_0: \beta = 0$ ).
- En resumen, se procede en 2 etapas:
  - i) Se efectúa la regresión MCO ignorando la posible heterocedasticidad. Se obtienen los errores estimados de la regresión.
  - ii) Se estima la regresión  $\ln(\sigma_i^2) = \ln(\sigma^2) + \beta \cdot \ln(X_i) + v_i$  y contrastamos  $H_0: \beta = 0$ .

##### c) Contraste general de White:

- Plantea la ventaja de no tener que especificar la forma funcional de la heterocedasticidad.
- En este caso distinguimos las siguientes etapas:
  - i) Estimación del modelo original por MCO, ignorando la heterocedasticidad.
  - ii) Estimación de una nueva regresión tomando como variable dependiente los cuadrados de los residuos de la primera etapa, y con  $p$  variables dependientes: la constante, los regresores originales, sus cuadrados y sus productos.
  - iii) Obtener el coeficiente de determinación  $R^2$  de la segunda regresión.
- WHITE demuestra que  $N \cdot R^2$  se distribuye asintóticamente como una  $\chi^2$  con  $p - 1$  grados de libertad, lo que nos permite verificar la existencia de heterocedasticidad mediante un contraste de hipótesis.
- Si se cumple la hipótesis nula de homocedasticidad, a medida que crece el número de observaciones, el coeficiente de determinación  $R^2$  debería tender a 0 (los regresores  $x$  no tienen poder explicativo sobre los términos de error).

##### d) Contraste de Goldfeld-Quandt:

- Este contraste parte del supuesto de que la magnitud de  $\sigma_i^2$  depende de una variable  $Z$ . Generalmente, esta variable es una de las variables explicativas, aunque no siempre.
- Distinguimos las siguientes etapas:
  - i) Se ordenan las observaciones por valores de la variable  $Z$ , de menor a mayor y se divide la muestra original en 2 submuestras de igual tamaño, omitiendo cierto número de observaciones en la mitad de la muestra.
  - ii) Se realizan 2 regresiones, una para cada uno de los 2 subconjuntos de observaciones.
  - iii) Se obtiene la suma residual de cada una de las regresiones y se construye un estadístico  $SR_2/SR_1 \sim F_{N_1-K, N_2-K}$  que nos permitirá contrastar si existen diferencias significativas entre las sumas de los errores al cuadrado en ambos subconjuntos de la población.

### 2.2.5. Soluciones

- Una vez contrastada la existencia de heterocedasticidad, el estimador de  $\beta$  debe ser modificado convenientemente.
- En función de si la matriz  $\Omega$  es o no conocida, pueden distinguirse **2 enfoques** fundamentales para la estimación del modelo.

#### 1) Estimación del modelo cuando $\Omega$ es conocida (MCP):

- En este caso, el estimador MCG ideal adopta una forma particular conocida como mínimos cuadrados ponderados (MCP).
- Concretamente, la aplicación del método anteriormente descrito se traduce en que las observaciones son ponderadas por la inversa de la desviación típica de  $\varepsilon_i$ .
- Valoración:
  - Esta ponderación logra que las perturbaciones pasen a ser homocedásticas en el nuevo modelo transformado, con lo que MCO vuelve a ser el ELIO.
  - Pero, es poco probable que se conozca precisamente  $\Omega$ .

#### 2) Estimación cuando $\Omega$ es desconocida:

- En ese caso el enfoque más habitual es la estimación robusta de la matriz de varianzas-covarianzas asintótica, propuesta por WHITE.
- WHITE demuestra cómo no es necesario conocer  $\Omega$ , bastando con estimar la expresión  $S = X'\Omega X$  en base a los propios cuadrados de los residuos resultantes de la estimación por MCO.
- Valoración:
  - No requiere conocer  $\Omega$ .
  - Pero no es aconsejable con muestras pequeñas, dado que las propiedades de este estimador son solamente asintóticas.

## 2.3. Autocorrelación

### 2.3.1. Definición

- Pasando al problema de la autocorrelación, implica también la violación del supuesto de perturbaciones esféricas al ser la covarianza entre errores para diferentes observaciones no nula:

$$\text{Cov}[\varepsilon_i, \varepsilon_j] \neq 0 \quad i \neq j$$

- En este caso, la matriz de covarianzas del término de error,  $\Omega$ :

$$E[\varepsilon\varepsilon'|X] = \Omega = \begin{pmatrix} \sigma^2 & \cdots & \text{Cov}[\varepsilon_N, \varepsilon_1] \\ \vdots & \sigma^2 & \vdots \\ \text{Cov}[\varepsilon_1, \varepsilon_N] & \cdots & \sigma^2 \end{pmatrix}$$

- Ya no sería una matriz diagonal.
- Aun suponiendo la homocedasticidad en principio, ningún componente será nulo.

### 2.3.2. Origen

- La autocorrelación es un problema que puede tener distintas **causas**:
  - Es un problema típicamente común de los datos de series temporales.
    - En efecto, es frecuente que las series temporales económicas presenten una “memoria”, debido a que la variación alrededor de la función de regresión no es independiente entre un periodo y el siguiente<sup>7</sup>.
    - Concretamente, la autocorrelación en muchas de las variables macroeconómicas suele estar causada por la existencia de ciclos y tendencias, que no son bien explicados por las variables independientes del modelo [ver tema 5.B.8].
  - También puede ser resultado de variables relevantes omitidas, o del proceso de agregación de datos.

<sup>7</sup> En la práctica, a medida que nos alejamos de la diagonal principal, el valor de las covarianzas tiende a cero. Esto se explica por el decrecimiento progresivo de la memoria, característico de los procesos autorregresivos de raíz no unitaria.

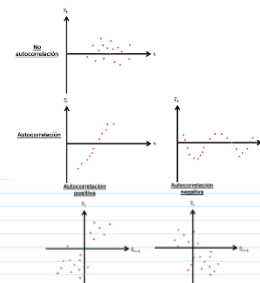


### 2.3.3. Consecuencias

- De nuevo, las estimaciones por MCO seguirán siendo insesgadas y consistentes, pero dejarán de ser eficientes, incluso para muestras muy grandes (asintóticamente). Esto se debe a que ya no se cumplen las condiciones para el Teorema Gauss-Márkov.
- En relación a su **estimación**:
  - Si se realiza la estimación mediante MCO teniendo en cuenta la autocorrelación, corrigiendo la varianza del estimador para tener en cuenta la autocorrelación, la estimación es insesgada pero la inferencia estadística se va a ver afectada. Las pruebas  $t$  y  $F$  se verán afectadas, ya que la varianza del estimador óptimo (MCG) es inferior a la del MCO. Ello implica que los intervalos de confianza serán innecesariamente grandes. Como resultado, es posible que las pruebas  $t$  y  $F$  den resultados imprecisos.
  - Si se realiza la estimación mediante MCO sin tener en cuenta la autocorrelación, la situación se torna más grave, pues la fórmula habitual de varianza para el beta estimado generará un sesgo.
- En definitiva, la característica más sobresaliente de estos resultados es que los MCO, con o sin corrección por autocorrelación, sobreestiman consistentemente el verdadero error estándar obtenido mediante el estimador ELIO de MCG.

### 2.3.4. Detección

- Una primera forma de detectar la autocorrelación sería realizar **contrastes gráficos** bien de la serie de residuos o de residuos consecutivos enfrentados entre sí. Por ejemplo, en el segundo tipo, una mayoría de puntos en el primer y tercer cuadrantes sería indicio de autocorrelación positiva.
- Una segunda opción es realizar un **contraste de Durbin-Watson** (ver Luis) el test más habitual si se sospecha que el error del modelo sigue un esquema autorregresivo de primer orden  $AR(1)$ .
- También puede realizarse un **contraste de Breusch-Godfrey**, que permite que las variables explicativas sean estocásticas, o esquemas autorregresivos distintos de primer orden, entre otros.
- Otros**: El *contraste h de Durbin* o el *Contraste Ljung-Box*.



#### Detección

**Contrastes gráficos**: el examen gráfico de los residuos generados en la estimación puede ser una primera aproximación muy útil para detectar rápidamente la existencia de autocorrelación. Ejemplos:

- Gráfico de la serie de residuos**: se deberán de buscar rachas de residuos de igual signo, claro indicio de autocorrelación positiva
- Gráfico de  $\varepsilon_t$  contra  $\varepsilon_{t-1}$** : revelando posibles autocorrelaciones positivas o negativas.

**Contraste de Durbin-Watson**: la prueba más conocida para detectar la correlación serial es la de los estadísticos *Durbin & Watson*. Se le conoce como estadístico  $D$  de *Durbin & Watson*, que se define como  $d = \frac{\sum (\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$ , que es simplemente la razón de las sumas de las diferencias al cuadrado de los residuos sucesivos sobre la SCR.

- Para muestras lo suficientemente grandes, el estadístico  $d$  puede expresarse como  $d \approx 2(1 - \rho)$ , siendo  $\rho = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$ , es decir, el grado de correlación entre los residuos consecutivos.

La idea del contraste es la siguiente:

- El coeficiente de correlación puede tomar valores comprendidos entre -1 y 1, por lo que el estadístico  $d$  tomará valores entre 0 y 4.
- El estadístico tomará un valor 2 cuando la correlación sea nula. A medida que  $d$  se aleje de 2, se tenderá a rechazar más fácilmente la hipótesis nula de  $d=2$ .

A pesar de su popularidad, este contraste presenta un conjunto de limitaciones prácticas:

- Para determinados valores de  $d$ , el investigador no podrá tomar una decisión basada en el contraste, ya que la distribución de probabilidad del estadístico es difícil de derivar.
- Las variables explicativas se suponen deterministas
- No permite detectar esquemas de autocorrelación generados por procesos autorregresivos de orden mayor que 1.

**Contraste de Breusch-Godfrey**: las principales ventajas de este test con respecto al de *Durbin-Watson* son que: *i)* las variables explicativas pueden ser estocásticas, *ii)* el contraste permite detectar esquemas de autocorrelación correspondientes a esquemas autorregresivos de orden superior a 1, *iii)* el test puede emplearse para modelos en los que uno o varios de los regresores son retardos de la variable endógena y *iv)* puede ser utilizado para identificar estructuras de autocorrelación correspondientes a procesos *MA*.

- Para elaborar el contraste puede suponerse que el término error sigue un proceso  $AR(p)$  como  $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \dots + \rho_p \varepsilon_{t-p} + u_t$ . Siendo  $u$  una variable aleatoria que sigue un proceso de ruido blanco y los parámetros  $\rho$  los coeficientes de autocorrelación parcial.
- El test contrastará la hipótesis nula de ausencia de autocorrelación del siguiente modo:  $H_0 = \rho_1 = \rho_2 = \dots = \rho_p = 0$ .
- Los pasos a seguir serán:

- Estimar el modelo por MCO y obtener así los residuos del mismo.
- Estimar una regresión de los residuos obtenidos sobre los  $p$  retardos y sobre las variables explicativas del modelo, obteniendo  $R^2$  el de la regresión auxiliar.
- Si el tamaño de la regresión es lo suficientemente grande, puede demostrarse que:  $(N - p)R^2 \sim \chi_p^2$ . Para aceptar la hipótesis nula de ausencia de relación serial el coeficiente de determinación de la regresión auxiliar debe ser muy pequeño. La idea que subyace es que, cuando  $N$  tiende a infinito, para unos residuos incorrelados con sus retardos y ortogonales a  $X$ ,  $R^2$  tiende a 0.

**Otros**: el *Contraste h de Durbin* o el *Contraste Ljung-Box*.



### 2.3.5. Solución

- En caso de que se haya detectado que existe autocorrelación y que ésta no es fruto de problemas de especificación<sup>8</sup>, se podrá:

– Estimar por MCG haciendo uso del procedimiento de Cochrane-Orcutt.

- Transforma el modelo original convenientemente para que se vuelva a cumplir el supuesto de perturbaciones esféricas (ver largo para desarrollo).
- Pero requiere que el coeficiente de autocorrelación  $\rho$  sea conocido.

Por sencillez expositiva, supondremos un modelo de regresión simple en el que el término error sigue un proceso  $AR(1)$  estacionario:  $y_t = \beta_0 + \beta_1 X_t + e_t$ ;  $e_t = \rho e_{t-1} + u_t$ .

Corrección MCG cuando  $\rho$  es conocido (procedimiento Cochrane-Orcutt)

El procedimiento comienza con unas transformaciones de las variables del modelo original conocidas como *cuasi-diferencias*. Se multiplicará cada término del modelo en el periodo  $t-1$  por el coeficiente  $\rho$  y se resta al modelo original en  $t$ , obteniendo así el siguiente modelo transformado  $y_t - \rho y_{t-1} = \beta_0(1 - \rho) + \beta_1(x_t - \rho x_{t-1}) + e_t - \rho e_{t-1}$ .

Dado que este nuevo modelo tendrá un término de error que sigue un proceso de ruido blanco, se cumple de nuevo el supuesto de perturbaciones esféricas. Podemos estimar los parámetros de este nuevo modelo por MCO de manera consistente y recuperar a continuación los parámetros de interés de la ecuación estructural original.

- Este método puede generalizarse en los modelos en los que el error sigue un proceso  $AR(p)$  y los parámetros son conocidos

- Emplear el método Newey-West para la corrección de errores estándar de los estimadores MCO, tal que estos sean robustos a autocorrelación:

- La idea es que para muestras grandes, el método de MCO sigue siendo adecuado, ya que la autocorrelación no afecta la consistencia.
- No obstante, para la inferencia sí se necesita una estimación adecuada de  $\Omega$ , que se puede obtener aplicando la corrección de Newey-West (ver largo)

Corrección MCG cuando  $\rho$  es desconocido (Newey-West)

**Concepto:** En la práctica,  $\rho$  será desconocido en la mayoría de los casos, por lo que el estimador robusto de autocorrelación de Newey-West es uno de los grandes avances del análisis econométrico y supone una generalización del estimador consistente de Eicker-White visto antes.

- La idea es que, para muestras grandes, el estimar MCO es adecuado, puesto que sigue siendo consistente en presencia de autocorrelación. Sin embargo, para realizar tareas de inferencia, es necesario contar con una estimación adecuada de la matriz de varianzas-covarianzas del estimador MCO:  $VCov(\hat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$

- Newey y West demuestran que la siguiente expresión es una estimación consistente y robusta a autocorrelación de  $S = X'\Omega X$ . Tenemos que:  $s^* = \hat{S} + \frac{1}{N} \sum_{j=1}^L \sum_{i=j+1}^N w(j) \hat{e}_i \hat{e}_{i-j} (x_i x_{i-j}' + x_{i-j} x_i')$  donde,  $\hat{S} = \sum_{i=1}^N \hat{e}_i^2 x_i x_i'$  (estimador consistente de White visto antes),

-  $w(j) = 1 - \frac{j}{L+1}$

El parámetro  $L$  será determinado previamente, de manera que sea lo suficientemente grande como para que las autocorrelaciones en retardos mayores que  $L$  sean lo suficientemente pequeñas para ignorarlas. De hecho, cuanto más grande sea  $L$ , más términos serán incluidos para corregir la autocorrelación.

- Nótese que Newey-West será igual al de White cuando  $L = 0$  (no consideración de la autocorrelación).

#### ○ Valoración

##### • Ventajas:

- Es robusto a formas de autocorrelación más allá de los procesos  $AR(1)$ .
- No exige que las variables explicativas sean estrictamente exógenas para ser consistente.

##### • Desventajas:

- No obstante, si la autocorrelación es considerable y la muestra pequeña las estimaciones por MCO serán muy ineficientes (y, en consecuencia, los coeficientes podrían ser mucho menos significativos).
- El criterio de elección de  $L$  no es obvio. La respuesta dependerá de si el modelo subyacente es un  $AR$ ,  $MA$  o  $ARMA$ , o de si los datos tienen una frecuencia anual, trimestral o mensual. En definitiva, el investigador deberá tomar una decisión discrecional.

## CONCLUSIÓN

▪

<sup>8</sup> En cuyo caso la solución pasaría por incorporar la variable omitida relevante o por hallar la forma funcional correcta del modelo.



## 5.B.4 : MODELOS DE ECUACIONES SIMULTÁNEAS: IDENTIFICACIÓN, ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

**▪ Contextualización:**

- Desde un punto de vista histórico, desde las primeras aportaciones de GAUSS a principios del s. XIX, el modelo más utilizado y estudiado por los econométricos ha sido el modelo lineal general (o modelo de regresión clásico).
  - Este modelo ha sido aplicado con gran éxito en el estudio de un elevado nº de temas económicos. Además, sirve de base para el desarrollo de modelos más complejos que levantan los supuestos clásicos, que pueden ser restrictivos en determinadas aplicaciones (series temporales, relaciones no lineales...).
- En efecto, la teoría económica señala que a menudo es necesario construir sistemas de relaciones, por ejemplo, si lo que se desea es estudiar el vaciado de mercado, o modelos macroeconómicos de consumo y de inversión.
  - Estos modelos generarán nuevas dificultades en la estimación, por lo que la econometría ha tenido que diseñar marcos teóricos para su adecuada estimación.
- En esta exposición, nos vamos a centrar en una **los modelos de ecuaciones simultáneas**.

**▪ Problemática (Preguntas clave):**

–

**▪ Estructura:****1. IDENTIFICACIÓN****1.1. Introducción al Modelos de Ecuaciones Simultáneas (MES)**

- 1.1.1. Idea
- 1.1.2. Ejemplo: Modelo de consumo keynesiano
- 1.1.3. Conclusiones

**1.2. Modelo estructural y modelo reducido****1.3. Problema de la identificación**

- 1.3.1. Idea
- 1.3.2. Restricciones para lograr la identificabilidad
- 1.3.3. Condiciones de orden y de rango con restricciones de exclusión

**2. MÉTODOS DE ESTIMACIÓN****2.1. Introducción: métodos de información limitada y métodos de información completa****2.2. Métodos uniecuacionales (información limitada)**

- 2.2.1. Mínimos Cuadrados Ordinarios (MCO)
- 2.2.2. Mínimos Cuadrados Indirectos (MCI)
- 2.2.3. Mínimos Cuadrados en 2 Etapas (MC2E)
- 2.2.4. Estimación por variables instrumentales
- 2.2.5. Máxima Verosimilitud Limitada (MVL)

**2.3. Métodos de sistemas (información completa)**

- 2.3.1. Mínimos Cuadrados en Tres Etapas (MC3E)
- 2.3.2. Máxima verosimilitud con información completa (MVIC)

**3. VERIFICACIÓN: EL TEST DE HAUSMAN Y WU****4. PREDICCIÓN**

## 1. IDENTIFICACIÓN

### 1.1. Introducción al Modelos de Ecuaciones Simultáneas (MES)

#### 1.1.1. Idea

- El **Modelo Lineal General (MLG)** presupone un modelo uniecuacional, es decir, existe *una única variable dependiente* y una o más variables explicativas. En tales modelos, la relación causa-efecto iba de los regresores a la variable dependiente.
- No obstante, en muchas situaciones, nos encontramos que la **relación causa-efecto es bidireccional**. Es decir, *la variable dependiente está determinada por las variables explicativas y, a su vez, algunas de las explicativas están determinadas por la variable dependiente*.
  - Así, en tales modelos hay más de una ecuación, concretamente una para cada una de las variables conjuntamente endógenas.
  - Como consecuencia, a diferencia de los modelos uniecuacionales, no va a ser posible estimar los parámetros de una ecuación aisladamente, sin tener en cuenta la información proporcionada por las demás ecuaciones del sistema.
    - En efecto, si los parámetros de cada ecuación se estimasen por MCO sin considerar las demás ecuaciones los estimadores serán no sólo sesgados, sino también inconsistentes.

#### 1.1.2. Ejemplo: Modelo de consumo keynesiano

- Para ilustrar esta situación, consideramos el **modelo básico de consumo keynesiano**:
  - En primer lugar, analicemos las ecuaciones del modelo:
    - Función de consumo:  $C_t = \beta_0 + \beta_1 \cdot Y_t + \varepsilon_t$ 
      - El consumo es una función lineal de la renta.
      - $\beta_1$  sería la propensión marginal a consumir, situada entre 0 y 1.
    - Identidad del ingreso:  $Y_t = C_t + I_t$ 
      - El ingreso total es igual al gasto de consumo total más el gasto de inversión (suponiendo una economía cerrada y sin gasto público).
  - Vemos cómo en este modelo existen las siguientes variables:
    - El consumo agregado y el nivel de renta están determinadas de forma *endógena*.
    - La inversión viene determinada *exógenamente* (i.e. sin influencia del consumo o del nivel de producto).
  - De lo anterior, se desprende que consumo y renta son *interdependientes*. Un alto valor de la perturbación de la ecuación de consumo en un periodo determinado generará 2 efectos:
    - *Efecto directo*: Un flujo de consumo superior a su propio valor medio.
    - *Efecto indirecto a través de la identidad contable*: El mayor consumo aumenta la renta, lo que a su vez impulsara nuevamente el consumo (*multiplicador keynesiano*).
    - Por tanto, el ingreso está correlacionado con el término de perturbación ( $cov(Y_t, u_t)$ )<sup>3</sup>.
- En lo que respecta a la **estimación econométrica del modelo**:
  - Así, si se estima la función de consumo anterior por MCO el estimador  $\widehat{\beta}_1^{MCO}$  atribuirá a la renta tanto el incremento directo como el indirecto.
  - En consecuencia, se produce un sesgo a la alza en la estimación de la propensión marginal al consumo:
 
$$\widehat{\beta}_1^{MCO} = \frac{\sum C_t \cdot Y_t}{\sum Y_t^2} = \frac{\sum (\beta_0 + \beta_1 \cdot Y_t + \varepsilon_t) \cdot Y_t}{\sum Y_t^2}$$

$$E[\widehat{\beta}_1^{MCO}] = \beta_1 + E\left[\frac{\sum Y_t \cdot \varepsilon_t}{\sum Y_t^2}\right]$$
  - Pero, además, se puede demostrar cómo llevará también a estimadores inconsistentes<sup>4</sup>.

<sup>3</sup> Analíticamente:  $cov[Y_t, \varepsilon_t] = E(Y_t - E[Y_t])(\varepsilon_t - E[\varepsilon_t])$ .

Se puede demostrar (GUJARATI, p. 680) como  $Y_t - E[Y_t] = \frac{\varepsilon_t}{1-\beta_1}$ , y, dado que  $E[\varepsilon_t] = 0$ , tenemos que  $cov[Y_t, \varepsilon_t] = \frac{E[\varepsilon_t^2]}{1-\beta_1} = \frac{\sigma^2}{1-\beta_1} \neq 0$ .

<sup>4</sup>  $plim(\widehat{\beta}_1^{MCO}) = plim(\beta_1) + plim\left(\frac{\sum Y_t \varepsilon_t}{\sum Y_t^2}\right) = \beta_1 + \frac{plim(\sum Y_t \varepsilon_t/n)}{plim(\sum Y_t^2/n)}$  y por el pie de página 3,  $plim(\sum Y_t \varepsilon_t/n) = cov[Y_t, \varepsilon_t] \neq 0$ .

### 1.1.3. Conclusiones

- **En suma**, en los Modelos de Ecuaciones Simultáneas (MES):
  - Se recogen *interacciones entre variables más complejas* que las que aparecen en el modelo uniecuacional.
  - La *estimación aislada de una de las ecuaciones por procedimientos habituales* (p.ej. MCO) producirá *sesgos* e *inconsistencia*.
    - En concreto, el MES implica que, al menos una de las variables explicativas de cada ecuación es endógena, de manera que estará correlacionada con el término error  $cov[Y_t, \varepsilon_t] \neq 0$ .
    - Esto se debe al problema de *causalidad simultánea* anteriormente descrito, que impide el estudio del comportamiento de la variable explicada y las explicativas de forma independiente.

## 1.2. Modelo estructural y modelo reducido

- El modelo de ecuaciones simultáneas (MES) tiene **variables de 2 tipos**:
  - *Endógenas*: Aquellas determinadas dentro del modelo.
  - *Exógenas*: Aquellas cuyos valores no están determinados por el modelo en el periodo de tiempo actual.
- La siguiente cuestión relevante es la distinción entre **modelo estructural** y en **forma reducida**.
  - *Modelo estructural*:

- El modelo básico parte de un sistema formado por  $g$  ecuaciones (por tanto,  $g$  variables endógenas) y  $k$  variables exógenas en cada periodo.
- Por tanto, en un periodo concreto  $t$ , tendríamos un sistema de  $g$  ecuaciones con la forma:

$$y_h = \gamma_{h1} \cdot y_1 + (\dots) + \gamma_{hg} \cdot y_g + \beta_{h1} \cdot x_1 + (\dots) + \beta_{hk} \cdot x_k + \varepsilon_h$$

donde  $h = 1, \dots, g$

- En forma matricial, para el periodo  $t$  podríamos expresar el modelo tal que:

$$\mathbf{y}_t = \mathbf{\Gamma} \cdot \mathbf{y}_t + \mathbf{B} \cdot \mathbf{x}_t + \varepsilon_t$$

donde,

- $\mathbf{y}_t$  y  $\varepsilon_t$  son vectores columna de dimensión  $g$ .
- $\mathbf{x}_t$  es un vector columna de dimensión  $k$ .
- $\mathbf{\Gamma}$  (gamma) es una matriz de dimensión  $g \times g$  que contiene en la primera fila los coeficientes de las  $g$  variables endógenas en la primera ecuación, y así sucesivamente
- $\mathbf{B}$  (beta) es una matriz  $g \times k$  que contiene los  $k$  coeficientes de las variables exógenas en cada una de las ecuaciones.

- Si disponemos de  $T$  observaciones, tenemos  $\mathbf{Y} = \mathbf{Y}\mathbf{\Gamma}' + \mathbf{X}\mathbf{B}' + \mathbf{E}$ .
- Esto es lo que se conoce como modelo estructural, **pues muestra la estructura del modelo económico**.

- *Modelo reducido*:

- A partir de las ecuaciones estructurales se pueden resolver para las  $g$  variables endógenas, derivando las *ecuaciones en forma reducida*.
- Estas son aquellas que expresan únicamente una variable endógena en términos de las variables predeterminadas y las perturbaciones estocásticas.
- Con el modelo anterior:

$$\mathbf{y}_t \cdot (1 - \mathbf{\Gamma}) = \mathbf{x}_t \cdot \mathbf{B}' + \varepsilon_t \rightarrow \mathbf{y}_t = \mathbf{x}_t \cdot \mathbf{\Pi}' + \mathbf{v}_t$$

donde  $\mathbf{\Pi} = \mathbf{B}' \cdot (1 - \mathbf{\Gamma})^{-1}$ , con dimensión  $k \times g$  y  $\mathbf{v}_t = \varepsilon_t \cdot (1 - \mathbf{\Gamma})^{-1}$  con dimensión  $T \times g$ .

- Para facilitar la comprensión, podemos tomar de nuevo el modelo básico de consumo:
  - *Modelo estructural*: Las ecuaciones de consumo y renta son las estructurales.
    - $C_t = \beta_0 + \beta_1 \cdot Y_t + \varepsilon_t$
    - $Y_t = C_t + I_t$



– Reducida:

- Operando para expresar las variables endógenas solo en términos de las exógenas y el error:

$$Y_t = \pi_0 + \pi_1 \cdot I_t + v_t$$

$$C_t = \pi_2 + \pi_3 \cdot I_t + v_t$$

donde obtenemos que:

$$\pi_0 = \frac{\beta_0}{1 - \beta_1} \quad \pi_1 = \frac{1}{1 - \beta_1} \quad \pi_2 = \frac{\beta_0}{1 - \beta_1} \quad \pi_3 = \frac{\beta_1}{1 - \beta_1} \quad v_t = \frac{\varepsilon_t}{1 - \beta_1}$$

## – Por tanto, ¿cómo podemos estimar el modelo?:

- Dado que por definición las variables predeterminadas no están correlacionadas con los términos de error, el método MCO puede aplicarse para estimar los coeficientes de las ecuaciones reducidas  $\pi$ .
  - A partir de estos coeficientes se podrían potencialmente obtener los coeficientes estructurales  $\beta$ .
  - A este procedimiento se le conoce como **Mínimos Cuadrados Indirectos**.
- Para poder estimar el modelo, es necesario realizar ciertos **supuestos sobre el término del error**:
    - i) Primero, la esperanza del error es nula en todas las ecuaciones del sistema ( $E[\varepsilon_{ht}] = 0 \forall h = 1, \dots, g$ ).
    - ii) Segundo, no puede haber correlación serial, es decir, las perturbaciones en 2 instantes de tiempo no están nunca correlacionadas, pertenezcan o no a la misma ecuación del sistema ( $\text{Cov}[\varepsilon_{ht}, \varepsilon_{ht'}] = 0 \forall t \neq t'$ ).
    - iii) Pero, sí se admite la existencia de correlaciones contemporáneas entre los términos de error de 2 ecuaciones distintas, que suponemos constante ( $\text{Cov}[\varepsilon_{ht}, \varepsilon_{ht'}] = \sigma_{hh'}$ ).
      - De hecho, es precisamente la existencia de estas correlaciones contemporáneas lo que constituye la esencia del modelo MES, y lleva a tratarlo como un modelo independiente en lugar de un conjunto de modelos uniecuacionales aislados.
      - Un cambio en la perturbación de una ecuación afecta a todas las variables endógenas del modelo, puesto que todas se determinan simultáneamente.
      - Esta simultaneidad proviene de 2 características fundamentales del modelo:
        - i) Todas las variables endógenas del modelo pueden aparecer en todas las ecuaciones; y
        - ii) Los términos de error de todas las ecuaciones están correlacionados contemporáneamente.

### 1.3. Problema de la identificación

#### 1.3.1. Idea

- El **problema de la identificación** pretende establecer si las estimaciones numéricas de los parámetros de una ecuación estructural pueden obtenerse de los coeficientes en forma reducida estimados.
  - En caso negativo, se dice que la ecuación bajo consideración está no identificada o subidentificada.
  - En caso afirmativo, se dice que la ecuación está identificada. Concretamente, podemos hablar de:
    - Exactamente identificada: Pueden obtenerse valores numéricos únicos de los parámetros estructurales.
    - Sobreidentificada: Puede obtenerse más de un valor numérico para algunos de los parámetros de las ecuaciones estructurales. Esto se deriva de la existencia de demasiada información (i.e. más ecuaciones que incógnitas).

### 1.3.2. Restricciones para lograr la identificabilidad

- A priori, esta tarea de recuperación podría resultar imposible, ya que la forma estructural cuenta con un mayor número de parámetros que la forma reducida. Sin embargo, existe un rango sorprendentemente amplio de **recursos para lograr la identificabilidad de una ecuación**.
  - Así, en primer lugar, la teoría económica puede proporcionarnos información adicional para resolver el problema.
    - Por ejemplo, podemos saber que *ciertas variables no deben aparecer en ciertas ecuaciones* (i.e. sus coeficientes deben ser 0 por pura lógica) o que *la combinación lineal de algunos de ellos debe ser 0* (p.ej. para que se cumpla una condición de equilibrio o una identidad contable). De este modo, imponiendo estas restricciones, *reduciríamos el número de incógnitas*.
    - Precisamente el recurso más habitual para lograr la identificabilidad de una ecuación suele ser imponer restricciones de exclusión sobre los coeficientes, que fuerzan determinados de entre ellos a cero. Así, la omisión de una ecuación coloca ceros en las matrices  $\Gamma$  y  $B$  que veíamos antes, reduciendo el número de incógnitas.
      - Por ejemplo, vamos a considerar un modelo sencillo de equilibrio de mercado:
        - $Q_t^d = \alpha_0 + \alpha_1 \cdot P_t + u_{t,1}$
        - $Q_t^s = \beta_0 + \beta_1 \cdot P_t + u_{t,2}$
 donde  $Q_t^d$  y  $Q_t^s$  son las cantidades demandadas y ofertadas en el periodo  $t$ ,  $P_t$  es el precio y  $u_{t,i}$  son los términos de error.
        - Si obtenemos forma reducida, veríamos que no hay forma de recuperar los coeficientes estructurales, puesto que se dispone tan solo de 2 ecuaciones en forma reducida para 4 coeficientes estructurales ( $\alpha_0, \beta_0, \alpha_1, \beta_1$ ).
      - No obstante, supongamos que se dispone de información adicional, acerca de determinadas variables que sólo afectan a la demanda o a la oferta.
        - Se incorpora el ingreso a la función de demanda, lo que proporcionará información adicional acerca del comportamiento del consumidor.
 
$$Q_t^d = \alpha_0 + \alpha_1 \cdot P_t + \alpha_2 \cdot I_t + u_{t,1}$$
        - Se incorpora el precio rezagado un periodo a la función de la oferta:
 
$$Q_t^s = \beta_0 + \beta_1 \cdot P_t + \beta_2 \cdot P_{t-1} + u_{t,2}$$
  - Si operamos para obtener las ecuaciones reducidas, veríamos cómo en este caso se disponen de 6 ecuaciones en forma reducida, para 6 coeficientes estructurales del modelo ( $\alpha_i, \beta_i$ , para  $i = \{1, 2, 3\}$ ).
  - Por tanto, los parámetros de cada una de las demandas y el sistema en su totalidad pueden ser identificados.
- Existen otras herramientas para lograr la identificabilidad. Por ejemplo:
  - Normalizaciones: Se podría realizar una normalización, de modo que en cada ecuación se tendrá una variable que tenga un 1 como coeficiente, reduciendo de esta manera el número de valores indeterminados.
  - Restricciones lineales (identidades): En algunos modelos, las definiciones de las variables o las condiciones de equilibrio implican que todos los coeficientes de una ecuación concreta son conocidos. No se tendrán problemas de identificación para ecuaciones que sean identidades contables o condiciones de equilibrio (recuérdese el ejemplo del modelo de consumo keynesiano).
  - Restricciones no lineales: Las restricciones en los parámetros estructurales pueden emplearse, también, para rechazar estructuras falsas.
  - Restricciones en la matriz de covarianzas de las perturbaciones: Son semejantes a las restricciones en los parámetros de las pendientes.

- *No linealidades*: En bastantes modelos, las variables y/o los coeficientes aparecen de modo no lineal. Esto puede complicar mucho el análisis. No obstante, las no linealidades pueden ayudar a la identificación.

### 1.3.3. Condiciones de orden y de rango con restricciones de exclusión

- Tal y como hemos visto, en principio es posible recurrir a las ecuaciones de forma reducida para determinar la identificación en un MES. Para que esto sea posible es necesario satisfacer las denominadas **condiciones de orden y de rango de identificación**.
  - Condición de orden de identificación: Una condición necesaria pero no suficiente para la identificación es la conocida como *condición de orden*, que podrá expresarse de 2 formas diferentes pero equivalentes:
    - En un modelo de  $g$  ecuaciones simultáneas (y  $g$  variables endógenas), para que una ecuación esté identificada debe excluir al menos  $g - 1$  variables (endógenas y predeterminadas) que aparecen en el modelo.
      - Si excluye exactamente  $g - 1$  variables, la ecuación está *exactamente identificada*. En caso de excluir más, la ecuación se encontrará *sobreidentificada*.
    - En un modelo de  $g$  ecuaciones simultáneas (y  $g$  variables endógenas), para que una ecuación esté identificada, el número de variables predeterminadas excluidas de esa ecuación no debe ser menor que el número de variables endógenas incluidas en esa ecuación menos 1.
      - Analíticamente, si  $k_{total} - k_{incluidas} = g_{incluidas} - 1$  la ecuación se encuentra *exactamente identificada*, y si  $k_{total} - k_{incluidas} > g_{incluidas} - 1$  la ecuación está *sobreidentificada*.
  - Condición de rango: No obstante, la condición de orden es una condición necesaria pero no suficiente para la identificación. Es decir, aunque una ecuación cumpla la condición de orden puede no estar identificada porque las variables excluidas de esa ecuación pero presentes en el modelo, quizá no sean independientes, de modo que no tal vez no existe una correspondencia uno a uno entre los coeficientes estructurales y los coeficientes en forma reducida. Es decir, probablemente, no sea posible estimar los parámetros estructurales a partir de los coeficientes en forma reducida.
    - Por ello, se recurre a la *condición de rango*, que establece que en un modelo con  $g$  ecuaciones y  $g$  variables endógenas, una ecuación estará identificada si y sólo si se puede construir por lo menos un determinante diferente de cero de orden  $(g - 1)(g - 1)$ , a partir de los coeficientes de las variables (endógenas y exógenas) excluidas de esa ecuación particular pero que sí son incluidas en las otras ecuaciones del modelo [ver GUJARATI pág. 701 para ejemplo concreto].
    - Esto es una condición suficiente.

## 2. MÉTODOS DE ESTIMACIÓN

### 2.1. Introducción: métodos de información limitada y métodos de información completa

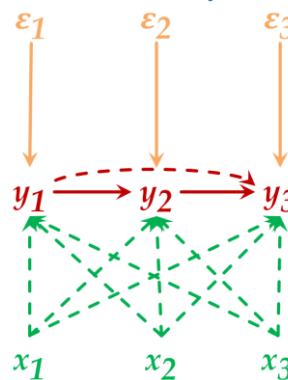
- Si se considera el modelo general de  $g$  ecuaciones con  $g$  variables endógenas, pueden adoptarse **2 enfoques para estimar las ecuaciones estructurales**:
  - Métodos uniecuacionales o de información limitada: Cada ecuación del sistema se estima individualmente, considerando sólo las restricciones impuestas sobre ella.
  - Métodos de sistemas o de información completa: Se estiman todas las ecuaciones del modelo de manera simultánea, teniendo en cuenta las restricciones ocasionadas por la omisión o ausencia de algunas variables sobre dichas ecuaciones.

- Para preservar el espíritu de los MES, idealmente debería utilizarse el método de sistemas (p.ej. el método de máxima verosimilitud con información completa que veremos posteriormente). No obstante, en la práctica no son de uso frecuente por múltiples razones:
  - i) Primero, por su enorme carga computacional<sup>5</sup>.
  - ii) Segundo, por conducir a soluciones altamente no lineales en los parámetros (por ende, difíciles de determinar).
  - iii) Tercero, porque si hay un error de especificación en alguna de las ecuaciones, dicho error es transmitido al resto del sistema.

## 2.2. Métodos uniecuacionales (información limitada)

### 2.2.1. Mínimos Cuadrados Ordinarios (MCO)

- La estimación de forma aislada de las ecuaciones por **métodos como MCO**:
  - Como vimos anteriormente, puede conducir a estimadores sesgados e inconsistentes debido a la interdependencia entre variables explicativas endógenas y la perturbación.
  - Sin embargo, hay una situación en la que el método de MCO puede ser aplicado apropiadamente. Este es el caso de los modelos triangulares, recursivos o causales:
    - Supongamos que tenemos 3 ecuaciones con 3 incógnitas y 3 términos de error como las reflejadas:
 
$$\begin{aligned} Y_{1t} &= \beta_{01} && + \gamma_{11} \cdot X_{1t} + (\dots) + \gamma_{k1} \cdot X_{kt} + \varepsilon_{1t} \\ Y_{2t} &= \beta_{02} + \beta_{12} \cdot Y_{1t} && + \gamma_{12} \cdot X_{1t} + (\dots) + \gamma_{k2} \cdot X_{kt} + \varepsilon_{2t} \\ Y_{3t} &= \beta_{03} + \beta_{13} \cdot Y_{1t} + \beta_{23} \cdot Y_{2t} && + \gamma_{13} \cdot X_{1t} + (\dots) + \gamma_{k3} \cdot X_{kt} + \varepsilon_{3t} \end{aligned}$$
    - En estos modelos, suponemos además que las perturbaciones de diferentes ecuaciones en un mismo periodo no están correlacionadas, es decir, no existe correlación contemporánea.
      - Por ello, no hay interdependencia entre las variables endógenas, es decir, si suponemos un MES de 3 variables endógenas,  $Y_1$  afecta a  $Y_2$  pero no al revés, e  $Y_1$  e  $Y_2$  influyen en  $Y_3$ , sin que esta última las influya.



- Así se puede aplicar MCO de forma separada a cada ecuación:
  - Considerando la primera ecuación, ésta únicamente contiene variables exógenas en el lado derecho y, como hemos supuesto que no están correlacionadas con el término error, esta ecuación satisface los supuestos críticos para aplicar MCO.
  - Para las restantes ecuaciones también será posible aplicar MCO ya que:
 
$$\text{corr}[Y_i, \varepsilon_j] = 0 \quad \forall i \neq j$$
- En la práctica, la mayoría de MES no presentan tal relación unilateral y MCO por lo general resultaría apropiado para estimar una sola ecuación en el contexto de un MES.

<sup>5</sup> Por ejemplo, el modelo de Klein-Goldberger de la economía estadounidense para 1955 de 20 ecuaciones (comparativamente pequeño) tenía 151 coeficientes distintos de cero, de los cuales sólo se estimaron 51 debido a la enorme carga computacional. Aunque modelos tan elaborados pueden proporcionar detalles complejos sobre diversos sectores de la economía, los cálculos representan un enorme esfuerzo.

### 2.2.2. Mínimos Cuadrados Indirectos (MCI)

- Como vimos anteriormente, para una ecuación estructural exactamente identificada, el método para obtener las estimaciones de los coeficientes estructurales a partir del modelo residual se conoce como **Mínimos Cuadrados Indirectos (MCI)**, y las estimaciones así obtenidas se realizan en 3 pasos:

- i) Primer paso: Se obtienen las ecuaciones en forma reducida por el procedimiento ya descrito.
- ii) Segundo paso: Se aplica MCO individualmente a las ecuaciones de forma reducida, lo cual es posible porque las variables explicativas no están correlacionadas con las perturbaciones. Las estimaciones de esta etapa serán consistentes.
- iii) Tercer paso: Se obtienen estimaciones de los coeficientes estructurales originales a partir de los coeficientes en forma reducida estimados. Si la ecuación está exactamente identificada, habrá una correspondencia uno a uno entre los coeficientes estructurales y los coeficientes en la forma reducida, por lo que podrán derivarse estimaciones únicas.
  - Los estimadores MCI heredan la propiedad de consistencia de los estimadores en forma reducida.

### 2.2.3. Mínimos Cuadrados en 2 Etapas (MC2E)

- Para las ecuaciones sobreidentificadas se puede proceder por el método de MC2E, que se estructura en 2 etapas:

- Primera etapa:
  - Se construyen las *regresiones auxiliares*, una para cada variable explicativa del modelo.
  - En cada una de estas regresiones auxiliares, la *variable dependiente será una de las variables endógenas*, y las *variables explicativas serán todas las variables exógenas* del modelo.
  - La idea es obtener estimaciones de las variables explicativas endógenas, sin tener que recurrir a la información muestral disponible para ellas.
- Segunda etapa:
  - Se sustituirán en el modelo original las variables explicativas endógenas por sus estimaciones de la etapa anterior.
  - Esta ecuación será estimada por MCO mediante el sistema de ecuaciones normales.

- **Características de la estimación mediante MC2E:**

- A diferencia de MCI (que proporciona múltiples estimaciones de los parámetros en las ecuaciones sobreidentificadas), MC2E proporciona sólo una estimación por parámetro. Por lo tanto, este método podrá emplearse para la estimación de cualquier ecuación que esté exactamente identificada o sobreidentificada:
  - Puede aplicarse a ecuaciones exactamente identificadas, coincidiendo MCI y MC2E.
  - El caso de ecuaciones sobreidentificadas es el que hace que este método tenga mayor interés.
- Puede aplicarse a una ecuación individual en el sistema sin tener en cuenta directamente ninguna otra ecuación o ecuaciones en el mismo. Por tanto, para resolver modelos econométricos que contienen un gran número de ecuaciones, MC2E ofrece un método económico.
- Es fácil de aplicar porque solamente se necesita saber el número de variable exógenas o predeterminadas en el sistema sin conocer ninguna otra variable en el mismo.

### 2.2.4. Estimación por variables instrumentales

- Es lógico que, para un modelo en el que el principal inconveniente, en términos de estimación, resulta ser la correlación existente entre las perturbaciones y al menos una variable explicativa, una de las opciones válidas sea el método de **variables instrumentales (VI)** [ver tema 5.B.2].
  - Los candidatos, posiblemente idóneos, a ser utilizados como instrumentos serán las variables exógenas del modelo que formen parte de una ecuación que no se desea estimar.
  - En tal caso se deberá tener en cuenta la condición de orden para tener la certeza de que se disponen de instrumentos suficientes.



### 2.2.5. Máxima Verosimilitud Limitada (MVL)

- El **método de Máxima Verosimilitud Limitada (MVL)** se propone maximizar la función de verosimilitud de la ecuación de interés, bajo normalidad del término de error y sujeto a las restricciones que suponen los coeficientes.
  - SCHMIDT (1976) demuestra que, para el caso de una ecuación exactamente identificada, el estimador MVL coincide con el de MC2E y, en consecuencia, con el MCI.
  - Asimismo, comparte la propiedad de eficiencia asintótica con el MC2E.

## 2.3. Métodos de sistemas (información completa)

### 2.3.1. Mínimos Cuadrados en Tres Etapas (MC3E)

- Similar a MC2E pero considerando las correlaciones mencionadas.
- Será válido para cualquier conjunto de ecuaciones *exactamente identificadas* o *sobreidentificadas*.
- Puede demostrarse que satisface los requisitos del estimador VI y que, por tanto, es un estimador consistente.

### 2.3.2. Máxima verosimilitud con información completa (MVIC)

- Bajo perturbaciones normalmente distribuidas, MVIC es eficiente.
- No obstante, no permite una representación analítica, precisando ser obtenido por métodos numéricos.

## 3. VERIFICACIÓN: EL TEST DE HAUSMAN Y WU

Coordinar en temas 5.B.2 y 5.B.4

- Hasta ahora hemos visto cómo el problema de simultaneidad puede llevar a que los estimadores MCO sean sesgados e inconsistentes, debiendo proceder por métodos alternativos como los propuestos.
  - Cuando sospechamos que determinadas variables son endógenas, hemos visto que se debe utilizar un procedimiento de variables instrumentales para obtener estimaciones consistentes. Sin embargo, es aconsejable verificar que las propiedades de exogeneidad del resto de variables se satisfacen, pues, de lo contrario, obtendríamos igualmente estimadores inconsistentes.
- El **problema** es que *la exogeneidad no puede ser testada directamente* porque:
  - La condición  $E[\varepsilon \cdot x] = 0$  no puede ser observada directamente (el parámetro poblacional  $\varepsilon$  es desconocido).
  - Utilizar los residuos de MCO también es inútil, puesto que, por construcción, se cumple siempre que  $E[\hat{\varepsilon} \cdot x] = 0$ , incluso en presencia de endogeneidad.
- **Solución:** Se puede comprobar la exogeneidad utilizando información adicional, en concreto, variables instrumentales. Ésta es la idea del *test de HAUSMAN y WU*.
  - Supongamos un modelo en el que se duda de la exogeneidad de  $r$  variables. El contraste consiste en:
    - Estimar el modelo por MCO y obtener la suma residual  $SR_0$ .
    - Estimar el modelo por VI y obtener la suma residual  $SR_1$ .
  - El estadístico a estimar es:  $(\widehat{\beta}_{MCO} - \widehat{\beta}_{VI})' \cdot [var(\widehat{\beta}_{VI}) - var(\widehat{\beta}_{MCO})] \cdot (\widehat{\beta}_{MCO} - \widehat{\beta}_{VI}) = \frac{SR_0 - SR_1}{\sigma^2} \sim \chi_r^2$ 
    - Se distribuye como una chi-cuadrado con  $r$  grados de libertad bajo la hipótesis nula de que todas las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría tal supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.
  - **Intuición:** La lógica del test de Hausman es la siguiente:
    - Bajo la hipótesis nula, tanto el estimador de MCO como el de VI son consistentes (aunque este último no eficiente). Por tanto, la diferencia  $d = \widehat{\beta}_{MCO} - \widehat{\beta}_{VI}$  converge en probabilidad a 0.
    - Bajo la hipótesis alternativa, el estimador de MCO será inconsistente, por lo que dicha diferencia será no nula.



#### 4. PREDICCIÓN

- Por último, pasando a la **predicción**, si el objetivo del investigador es únicamente la predicción de valores futuros de las variables endógenas (*y no conocer los verdaderos coeficientes del modelo estructural*), bastaría con estimar el modelo reducido.
  - Por tanto, se puede hacer uso de MCO, que será insesgado y consistente.
  - La predicción óptima que minimiza el Error Cuadrático Medio de Pronóstico (ECMP) vendría dada por la esperanza de las variables condicionadas a los valores de las variables explicativas.
- Desde un punto de vista histórico:
  - Hasta los años 70, los MES fueron ampliamente utilizados en las tareas de predicción por parte de las autoridades económicas, en consonancia con las recomendaciones de la Comisión Cowles.
    - En concreto, se fueron desarrollando modelos macroeconómicos cada vez más complejos, que reflejaban la interdependencia de distintas variables económicas.
  - Sin embargo, desde mediados de los años 70, la utilidad de estos modelos es puesta en duda por varios autores, entre los que destaca la *crítica de LUCAS (1976)* y la *crítica de SIMS (1980)*.
    - Este último autor considera que la capacidad predictiva de los modelos MES era dudosa por varias deficiencias:
      - i) Primero, cuestionaba la *validez de las restricciones de exclusión* que según él no siempre venían respaldadas por la teoría económica.
      - ii) Segundo, critica también la *clasificación arbitraria de las variables en exógenas y endógenas* en muchos casos.
    - Estos aspectos movieron a SIMS a proponer una alternativa en la modelización de la relación entre varias variables, los métodos de vectores autorregresivos (VAR).
      - Estos son una generalización para el caso multivariante de los modelos dinámicos uniecuacionales.
      - Este método permite evitar la imposición de restricciones subjetivas sobre los coeficientes del modelo.
      - También permite considerar relaciones dinámicas generales entre las variables.
- **Valoración de los VAR:**
  - Los defensores de los VAR destacan las siguientes virtudes del método:
    - No es preciso preocuparse por determinar que variables son endógenas y cuáles son exógenas, ya que el VAR las toma todas como endógenas.
    - Estimación sencilla, ya que aplica MCO a cada ecuación por separado.
    - Predicciones obtenidas en muchos casos mejores que las obtenidas con MES más complejos.
  - No obstante, estos modelos también tendrán una serie de críticas:
    - A diferencia de los MES es ateoórico, al utilizar menos información previa.
    - Debido a su acento en el pronóstico, los modelos VAR son menos apropiados para el análisis de políticas.
    - Se requiere que todas las variables del modelo sean estacionarias.
    - La selección de la longitud apropiada de los retardos plantea un desafío práctico.
    - Puede ser difícil interpretar los coeficientes individuales estimados.
      - No obstante, se pueden estimar las funciones impulso-respuesta, que analizan la respuesta de una variable dependiente del sistema VAR ante shocks en los términos de error.

#### CONCLUSIÓN

- En conclusión, el MLG presupone un modelo uniecuacional, pero, la teoría económica señala que a menudo es necesario construir sistemas de relaciones.
- Hemos visto cómo estos modelos generarán nuevas dificultades en la estimación, así como diferentes marcos teóricos para su adecuada estimación.



## 5.B.5 : MODELOS UNIECUACIONALES DE REGRESIÓN NO LINEAL: ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

---

▪ **Contextualización:**

- Desde un punto de vista histórico, desde las primeras aportaciones de GAUSS a principios del s. XIX, el modelo más utilizado y estudiado por los econométricos ha sido el modelo lineal general (o modelo de regresión clásico).
  - Este modelo ha sido aplicado con gran éxito en el estudio de un elevado nº de temas económicos. Además, sirve de base para el desarrollo de modelos más complejos que levantan los supuestos clásicos, que pueden ser restrictivos en determinadas aplicaciones (series temporales, relaciones no lineales...).
- En este sentido, *cuando se incumplen algunos de los supuestos básicos en los que se basa el MLG*, los métodos de estimación habituales (p.ej. MCO o MV), pueden dejar de cumplir algunas de sus propiedades de insesgadez, consistencia asintótica o eficiencia. Por ello, *será siempre necesario evaluar con espíritu crítico, caso por caso, la validez de los supuestos del MLG antes de realizar cualquier ejercicio econométrico*. En caso de detectar algún problema, deberemos *evaluar sus consecuencias* y, de ser graves, aplicar algún método econométrico para solucionarlo.
  - En concreto, uno de los supuestos del MLG más controvertidos es el de la linealidad, pues la teoría económica señala que existen numerosas relaciones entre variables económicas que no son lineales. Sin embargo, este supuesto no es tan restrictivo como cabría pensar en un primer momento:
    - Un gran nº de relaciones que, en principio, no son lineales, pueden llegar a serlo si se aplican simples transformaciones.
    - En otros casos, el modelo podría ser considerado una aproximación lineal en serie de Taylor del verdadero modelo no lineal.
  - Sin embargo, como veremos a lo largo de la exposición, existen casos en los que la no linealidad puede ocasionar graves sesgos para la estimación y la predicción, por lo que es necesario emplear nuevos métodos de estimación que tengan en cuenta estas no linealidades.

▪ **Problemática (Preguntas clave):**

—

▪ **Estructura:****1. ESTIMACIÓN***1.0. Idea**1.1. Distintas formas funcionales dan pie a distintos tipos de no linealidad*

- 1.1.1. No linealidad en las variables explicativas
- 1.1.2. No linealidad en los coeficientes
- 1.1.3. No linealidad en las variables explicativas y en los coeficientes
- 1.1.4. No linealidad en la variable endógena

*1.2. Estimación no lineal*

- 1.2.1. Regresión linealizada (aproximación lineal al modelo no lineal)
- 1.2.2. Mínimos Cuadrados No Lineales (MCNL)
- 1.2.3. Máxima Verosimilitud No Lineal (MVNL)
- 1.2.4. Transformación Box-Cox
- 1.2.5. Otros métodos

**2. VERIFICACIÓN***2.1. Contraste de linealidad del modelo*

- 2.1.1. Análisis de los residuos: Contraste gráfico
- 2.1.2. Regresiones por tramos: Contraste de estabilidad de Chow
- 2.1.3. Uso de polinomios de las variables
  - Contraste RESET de mala especificación (RAMSEY, 1969)
  - Prueba del multiplicador de Lagrange

*2.2. Contrastes de hipótesis*

- 2.2.1. Contraste F
- 2.2.2. Contraste de Wald
- 2.2.3. Contraste de la razón de verosimilitud
- 2.2.4. Contraste de multiplicadores de Lagrange

**3. PREDICCIÓN****4. MODELOS DE REGRESIÓN NO LINEAL: RESPUESTA CUALITATIVA***4.1. Idea: Naturaleza de los modelos de respuesta cualitativa**4.2. Posibilidades de estimación*

- 4.2.1. Modelo lineal de probabilidad (MLP)
- 4.2.2. Modelo Logit
  - Definición
  - Propiedades
- 4.2.3. Modelo Probit
  - Definición
  - Efectos marginales

*4.3. Valoración*

## 1. ESTIMACIÓN

### 1.0. Idea

- En un gran número de aplicaciones económicas, **el modelo que relaciona las variables endógenas con las explicativas es no lineal**. Por ejemplo, habitualmente, en los modelos microeconómicos, las funciones de producción dependen de forma no lineal de los inputs utilizados.
  - Por tanto, es importante analizar las posibles especificaciones del modelo de regresión, para comprobar cuáles son las estimaciones en términos de métodos de estimación.
  - En este contexto, la forma del modelo de regresión se puede generalizar de la siguiente manera:

$$y_i = f(x_i, \beta) \quad \forall i = 1, 2, \dots, N$$

donde  $f(x_i, \beta)$  es una función, generalmente, no lineal de los componentes de los vectores  $x_i$  y  $\beta$ . Así, el MLG sería un caso particular en el que  $f(x_i, \beta) = x_i' \beta$ .

### 1.1. Distintas formas funcionales dan pie a distintos tipos de no linealidad

- Las **implicaciones** serán *muy distintas en función de cuál sea la forma funcional del modelo*:
  - En algunos casos, es posible realizar una *transformación previa* del modelo que nos permitirá seguir aplicando los métodos habituales de estimación del MLG. Decimos, por tanto, que estos modelos son *intrínsecamente lineales*.
  - En otros casos, esto no será posible, y será recomendable recurrir a *métodos no lineales*.
- A continuación, veremos las principales formas funcionales en las que se pueden presentar no linealidades y comprobaremos si es posible realizar una linealización del modelo para estimar por métodos lineales.

#### 1.1.1. No linealidad en las variables explicativas

- En algunos modelos no lineales, la dependencia no lineal sólo se da entre la variable endógena y las explicativas, siendo la relación entre la variable endógena y los coeficientes lineal. En estos casos, el problema se resuelve fácilmente mediante la *transformación de datos* oportuna:
  - $y_i = \beta_0 + \beta_1 \cdot e^{x_{1i}} + \varepsilon_i \rightarrow y_i = \beta_0 + \beta_1 \cdot z_{1i} + \varepsilon_i$ , donde  $z_{1i} = e^{x_{1i}}$
  - $y_i = \beta_0 + \beta_1 \cdot x_{1i} \cdot x_{2i} + \varepsilon_i \rightarrow y_i = \beta_0 + \beta_1 \cdot z_{2i} + \varepsilon_i$ , donde  $z_{2i} = x_{1i} \cdot x_{2i}$
- En definitiva, basta definir unas nuevas variables explicativas para obtener un modelo lineal.

#### 1.1.2. No linealidad en los coeficientes

- En otras ocasiones, la no linealidad afecta a los coeficientes del modelo, pero no a las variables:
  - $y_i = \beta_0 + \ln(\beta_1) \cdot x_{1i} + \varepsilon_i$ 
    - En este caso, tampoco existen grandes dificultades para encontrar estimadores de  $\beta_0$  y  $\beta_1$ :
      - En primer lugar, se estima la constante del modelo y el coeficiente  $\beta_1^* = \ln(\beta_1)$  que acompaña a  $x_{1i}$ .
      - En segundo lugar, se procede a la recuperación de  $\beta_1$  mediante  $\hat{\beta}_1 = e^{\hat{\beta}_1^*}$ . No obstante, debe tenerse en cuenta que el estimador de  $\beta_1$  así obtenido no heredará directamente las propiedades estadísticas que pudiera tener el estimador de  $\ln(\beta_1)$ .
  - $y_i = \beta_0 + \beta_1 \cdot e^{\beta_2} \cdot x_{1i} + \varepsilon_i$ 
    - En este caso, pueden estimarse el término independiente y el coeficiente  $\beta_1^* = \beta_1 \cdot e^{\beta_2}$  que acompaña a  $x_{1i}$ , pero las estimaciones de los coeficientes  $\beta_1$  y  $\beta_2$  no podrían recuperarse (a no ser que se tenga información adicional acerca de sus valores numéricos y su suma o producto).



### 1.1.3. No linealidad en las variables explicativas y en los coeficientes

- También se pueden producir los 2 fenómenos mencionados simultáneamente, como en los siguientes ejemplos:

$$i) y_i = \beta_0 \cdot x_{1i}^{\beta_1} \cdot x_{2i}^{\beta_2} \cdot e^{\varepsilon_i}$$

- En este caso, tomando logaritmos a ambos lados de la igualdad, el modelo puede ser transformado en uno lineal. Hablamos entonces de *modelos log-lineales*.
- Por ejemplo, la función de producción Cobb-Douglas<sup>3</sup> que relaciona producción y factores productivos, puede ser linealizada tomando logaritmos en ambos lados de la ecuación:

$$y = A \cdot K^\alpha \cdot L^\beta \cdot e^\varepsilon \rightarrow \ln(y) = \ln(A) + \alpha \cdot \ln(K) + \beta \cdot \ln(L) + \varepsilon$$

$$ii) y_i = \beta_0 + \beta_1 \cdot x_{1i}^{\beta_2} + \varepsilon_i$$

- En este caso, *no se puede log-linealizar el modelo*. Por tanto, deberemos recurrir a métodos específicos de estimación.
- Por ejemplo, una especificación de la función de producción muy usada en la literatura económica reciente por su versatilidad es la función de elasticidad constante (CES):

$$y = A \cdot [\alpha \cdot K^\sigma + (1 - \alpha) \cdot L^\sigma]^{1/\sigma}$$

Pese a su uso habitual, esta función no es log-linealizable, por lo que deberán utilizarse otras técnicas para su estimación, como la aproximación de Taylor.

### 1.1.4. No linealidad en la variable endógena

- Finalmente, puede darse el caso en el que la no linealidad afecte a la variable dependiente de tal modo que sea imposible expresarla de manera explícita como función de los vectores  $x_i$  y  $\beta$ :

$$i) y_i \cdot x_{1i} + \beta_1 \cdot \ln(y_i) = \beta_2 \cdot x_{1i} + \varepsilon_i$$

- La forma funcional en este tipo de modelos será una función implícita:

$$g(y_i, x_i, \beta) = \varepsilon_i$$

- ii) Por ejemplo, en modelos probabilísticos, como los modelos *logit*, la variable endógena puede seguir una distribución logística, de tal forma que esta quede expresada de la siguiente forma:

$$\frac{1 - y_i}{y_i} = e^{\beta_0 + \beta_1 \cdot x_{1i} + \varepsilon_i}$$

- A través de las transformaciones pertinentes este modelo será estimable por MCO.

<sup>3</sup> La función de producción Cobb-Douglas es el ejemplo más común de función de producción utilizada en macroeconomía.

$$Y_t = A_t \cdot K_t^\alpha \cdot L_t^\beta$$

Concretamente consideraremos una función de producción Cobb-Douglas con rendimientos constantes a escala ( $\alpha + \beta = 1$ ) y productividades marginales decrecientes ( $\alpha \in (0,1)$  y  $\beta \in (0,1)$ ):

$$Y_t = K_t^\alpha \cdot (A_t \cdot L_t)^{1-\alpha}$$

PAUL DOUGLAS fue un senador por Illinois entre 1949 y 1966. Cuando todavía era profesor de economía, DOUGLAS descubrió un hecho sorprendente: la división de la renta nacional entre trabajadores y capitalistas permanecía más o menos constante en el tiempo. En particular, descubrió que los trabajadores en Estados Unidos se quedan con, más o menos, el 70 % de la renta total mientras que los capitalistas se quedan con el 30 %. Esto le llevó a indagar las condiciones bajo las cuales las rentas de los factores mantenían proporciones constantes.

Como no sabía solucionar el problema, DOUGLAS le preguntó a un matemático amigo suyo llamado CHARLES COBB si había una función de producción tal que, si los factores de producción cobraban sus productos marginales, la proporción de la renta agregada que se quedaba cada uno de ellos fuera constante. La función de producción, pues, debería tener las dos propiedades siguientes:

$$\text{Renta del capital} = \text{Producto marginal del capital} \cdot K = \alpha \cdot Y$$

$$\text{Renta del trabajo} = \text{Producto marginal del trabajo} \cdot L = (1 - \alpha) \cdot Y$$

Donde  $\alpha \in (0,1)$  es una constante que mide la fracción de la renta que se queda el capital (*partición del capital*). CHARLES COBB demostró que tal función de producción existía y tomaba la forma  $Y_t = A_t \cdot K_t^\alpha \cdot L_t^{1-\alpha}$ . Esta función de producción pasó a llamarse Cobb-Douglas. La función de producción Cobb-Douglas posee las siguientes características:

- 1) *Monótona*;
- 2) *Estrictamente convexa*;
- 3) *Curvas isocuantas asintóticas a los ejes* (i.e. si un factor de producción no se usa, el nivel de producción es cero);
- 4) *Rendimientos constantes a escala en  $K_t$ ,  $L_t$*  (por lo tanto, la función de producción es homogénea de grado 1  $\Rightarrow$  homotética  $\Rightarrow$  forma polar de Gorman);
- 5) *Productividades marginales positivas pero decrecientes*;
- 6) *Cumple las condiciones de Inada*;
- 7) *Los factores productivos son cooperativos*;
- 8) *La proporción de la renta agregada atribuible al trabajo  $(1 - \alpha)$  y al capital  $(\alpha)$  es constante*;
- 9) *La elasticidad de sustitución es constante e igual a uno*.
- 10) *Se convierte en una función lineal en logaritmos*, lo que hace que se haya convertido en una función popular en el campo de la econometría.

- Finalmente, puede comprobarse que, en los modelos de regresión no lineal vistos, el número de variables explicativas y el número de coeficientes no están asociados necesariamente (a diferencia de lo que ocurriría en el MLG, en el que existían  $k$  variables independientes y  $k+1$  parámetros por la existencia del término independiente).

## 1.2. Estimación no lineal

- Un modelo de regresión no lineal es aquel cuyas CPO para la estimación de sus parámetros por mínimos cuadrados son funciones no lineales de sus parámetros.

### 1.2.1. Regresión linealizada (aproximación lineal al modelo no lineal)

- Este método funciona del siguiente modo:
  - En primer lugar, se obtiene una aproximación lineal del modelo original, mediante un desarrollo en serie de Taylor de la función  $f(x_i, \beta)$  alrededor de un estimador inicial  $\tilde{\beta}$ .
  - En segundo lugar, se estima el modelo lineal resultante por los métodos ya conocidos.

- Supongamos el siguiente modelo de regresión no lineal:

$$y_i = f(x_i, \beta) + \varepsilon_i$$

- Si realizamos un desarrollo en serie de Taylor de primer orden alrededor de un vector  $\tilde{\beta}$  de la función  $f$ , obtenemos:

$$y_i \simeq f(x_i, \tilde{\beta}) + \left( \frac{\partial f(x_i, \tilde{\beta})}{\partial \beta} \right)' \cdot (\beta - \tilde{\beta}) + \varepsilon_i$$

- Reordenando, de forma que pongamos todas las constantes en el lado izquierdo:

$$y_i - f(x_i, \tilde{\beta}) + \left( \frac{\partial f(x_i, \tilde{\beta})}{\partial \beta} \right)' \cdot \tilde{\beta} \simeq \left( \frac{\partial f(x_i, \tilde{\beta})}{\partial \beta} \right)' \cdot \beta + \varepsilon_i$$

- Renombrando:

$$y_i^* \simeq x_i^{*'} \cdot \beta + \varepsilon_i \quad \forall i = 1, 2, \dots, N$$

donde  $x_i^*$  e  $y_i^*$  son variables transformadas, que podemos calcular a partir de la información muestral y los gradientes de la función no lineal.

- En definitiva, hemos obtenido una expresión del modelo que es lineal en los coeficientes de  $\beta$ . Por tanto, podemos ahora aplicar el método MCO habitual para encontrar estimaciones de estos coeficientes:

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

- Evidentemente, al basarse en un modelo aproximado, la estimación de  $\beta$  será sesgada, pero puede ser una buena aproximación.
- Cuanto menos marginales sean las variaciones que buscan ser estimadas, mayor sesgo.

### 1.2.2. Mínimos Cuadrados No Lineales (MCNL)

- La **idea** que subyace al método de mínimos cuadrados no depende en modo alguno de la linealidad del modelo, sino que es aplicable en condiciones más generales.
  - Por tanto, la minimización de los cuadrados de los residuos del modelo vuelve a ser una opción al problema de estimación. No obstante, la diferencia es que, ahora, la resolución analítica del problema será de mayor complejidad que en el modelo lineal.
- El **estimador mínimo cuadrático** es aquel que minimiza la norma euclídea del vector  $\varepsilon$ , es decir, de la suma de residuos al cuadrado (en adelante, SR):

$$\min_{\beta} SR(\beta) = \sum_{i=1}^N \hat{\varepsilon}_i^2 = \sum_{i=1}^N [y_i - f(x_i, \beta)]^2 = \hat{\varepsilon}' \cdot \hat{\varepsilon} = y'y - 2\tilde{\beta}X'y + \tilde{\beta}X'X\tilde{\beta}$$

- Tomando derivadas con respecto a cada uno de los componentes del vector  $\tilde{\beta}$ , se tienen las CPO para el problema de minimización:

$$\frac{\partial SR(\tilde{\beta})}{\partial \beta_1} = -2 \cdot \sum_{i=1}^N [y_i - f(x_i, \tilde{\beta})] \cdot \frac{\partial f(x_i, \tilde{\beta})}{\partial \beta_1} = 0$$

$$\frac{\partial SR(\hat{\beta})}{\partial \beta_k} = -2 \cdot \sum_{i=1}^N [y_i - f(x_i, \hat{\beta})] \cdot \frac{\partial f(x_i, \hat{\beta})}{\partial \beta_k} = 0$$

- De forma abreviada, utilizando la notación vectorial  $\beta$ :

$$\left[ \frac{\partial f(x_i, \hat{\beta})}{\partial \beta} \right]' \cdot \hat{\varepsilon} = 0_k$$

- Obsérvese que, cuando la forma funcional es lineal, esta última condición se reduce a que los residuos son ortogonales a los regresores:  $X'\hat{\varepsilon} = 0_k$ .

- En términos generales, el estimador de mínimos cuadrados es aquel vector de coeficientes  $\hat{\beta}$  que genera un vector de residuos ortogonal a cada uno de los componentes del vector gradiente de la función  $f(x_i, \hat{\beta})$  evaluado en  $\hat{\beta}$ .

- Podemos realizar una serie de **observaciones** sobre el método de MCNL:

- Se tendrá un sistema de ecuaciones normales no lineales. Por tanto, el estimador MCNL dependerá de  $y$  (y en consecuencia de  $\varepsilon$ ) de forma no lineal. Dado que la prueba de insesgadez depende de esta característica, las estimaciones serán, en general, sesgadas.
- La solución a dicho sistema puede no ser única o incluso no existir (un sistema de ecuaciones no lineales no siempre tiene solución).
- En general, no se puede obtener una expresión analítica del estimador MCNL, por lo que el problema deberá ser resuelto por métodos numéricos (algoritmos). Con frecuencia, se emplea el *método Gauss-Newton*.

- En relación a las **propiedades** del estimador MCNL:

- Puede demostrarse que, asintóticamente (cuando el tamaño muestral crece)  $\hat{\beta} \sim N[\beta, Var(\hat{\beta})]$ , donde  $Var(\hat{\beta}) = \sigma^2 \cdot \left[ \sum_{i=1}^N \left( \frac{\partial f(x_i, \hat{\beta})}{\partial \beta} \right) \left( \frac{\partial f(x_i, \hat{\beta})}{\partial \beta} \right)' \right]^{-1}$ .
- Además, el estimador<sup>4</sup> de la varianza del error es  $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}$  que es un estimador consistente de  $\sigma^2$ . Bajo normalidad,  $\hat{\sigma}^2$  coincide con el de máxima verosimilitud.
  - En resumen, las propiedades fundamentales atribuibles al estimador MCNL son insesgadez, consistencia y normalidad asintótica. Sin embargo, a excepción del caso en que los residuos estén normalmente distribuidos, no se puede asegurar que tal estimador sea el más eficiente.

### 1.2.3. Máxima Verosimilitud No Lineal (MVNL)

- De nuevo, la **idea** subyacente al método de regresión por máxima verosimilitud es totalmente independiente de la linealidad del modelo, por lo que las implicaciones de introducir no linealidad son principalmente de carácter técnico.
  - Sin embargo, a diferencia del método de MCO, el método de MV descansa directamente en un supuesto sobre la distribución del vector de errores, que en la mayoría de los casos se asume normal (S6). Si el supuesto es correcto, se ganará en eficiencia utilizando tal información adicional.
- La idea de este método es elegir los estimadores de  $\beta$  y  $\sigma^2$  que maximizan la probabilidad de que el verdadero proceso generador de datos haya generado la muestra efectivamente observada<sup>5</sup>.
  - La función de verosimilitud ("*likelihood function*") es la función de probabilidad conjunta de la muestra:

$$L = f(y_1, \dots, y_N | X, \beta, \sigma^2)$$

<sup>4</sup> Nótese que, a pesar de que frecuentemente puede encontrarse una corrección,  $1/(N-k)$ , ésta es irrelevante en este contexto, puesto que se están exponiendo resultados asintóticos. Por supuesto, en muestras pequeñas, las diferencias sí pueden ser considerables.

<sup>5</sup> Intuitivamente, se acepta que, de varios sucesos posibles, ocurrirá el más probable. Por ejemplo, si en una bolsa hay más bolas blancas que negras y extraemos una al azar, lo más probable es que ésta sea blanca. Con el método de máxima verosimilitud, si no sabemos qué tipo de bolas son las más numerosas y extraemos una bola blanca (muestra), entonces inferiremos que el número de bolas blancas es superior al de negras en la bolsa (población).

- Cuando la muestra es aleatoria simple (i.i.d.), las distintas observaciones muestrales son independientes entre sí, por lo que la función de verosimilitud es el producto de los valores de la función de densidad para cada una de las observaciones:

$$L = f(y_1 | x_1, \beta, \sigma^2) \cdot \dots \cdot f(y_N | x_N, \beta, \sigma^2)$$

$$L = \prod_{i=1}^N f(y_i | x_i, \beta, \sigma^2)$$

- Teniendo en cuenta el supuesto de distribución normal de  $\varepsilon$  (S6), podemos sustituir la forma concreta de la función de densidad normal:

$$L = (2 \cdot \pi \cdot \sigma^2)^{-N/2} \cdot e^{\left( -\frac{(y-f(x,\beta))' \cdot (y-f(x,\beta))}{2 \cdot \sigma^2} \right)}$$

- Si el parámetro  $\sigma^2$  no depende de ninguno de los parámetros  $\beta$ , entonces escoger el vector de parámetros  $\hat{\beta}_{MV}$  que maximice la función de verosimilitud (o su logaritmo) es equivalente a escoger el vector  $\hat{\beta}$  que minimice la suma residual.

- Por tanto, los estimadores MV y MCO coinciden si la distribución del error es normal.

#### 1.2.4. Transformación Box-Cox

- La familia de transformaciones Box-Cox puede ser utilizada para especificar una relación del tipo:

$$y_i = \alpha + \beta \cdot g(x_i) + \varepsilon_i$$

donde se tiene que:

$$g(x_i) = \frac{x_i^\lambda - 1}{\lambda}$$

- En la mayoría de casos, el valor de  $\lambda$  se encontrará entre -1 y 1, lo que dará lugar a distintas formas funcionales del modelo. Algunas de las más habituales son:

- $\lambda = 0 \Rightarrow g(x_i) = \ln(x_i)$ <sup>6</sup>.
  - $\lambda = 1 \Rightarrow g(x_i) = x_i - 1$  (i.e. *modelo lineal*).
  - $\lambda = -1 \Rightarrow g(x_i) = 1/x_i$ .

- Nótese que:

- Si el parámetro  $\lambda$  es conocido, la relación expuesta es un modelo de regresión intrínsecamente lineal.
  - Si  $\lambda$  es un coeficiente a estimar, el modelo es intrínsecamente no lineal (se tendría una regresión no lineal en los parámetros).

- Un posible tratamiento para este último caso sería:

- Fijar un valor para  $\lambda$  y realizar un cambio de variables inmediato que nos conduciría a un modelo lineal. Estimar el modelo por MCO.
- Realizar lo mismo para varios valores de  $\lambda$ . Por ejemplo, se podrían realizar estimaciones con  $\lambda$  en un rango de -2 a 2, con incrementos de 0,1 en cada estimación.
- Finalmente, se seleccionaría el modelo buscando aquel valor de  $\lambda$  con el que se minimice la suma residual. Dicho valor de  $\lambda$ , junto con las estimaciones de  $\alpha$  y  $\beta$  asociadas, constituirán el estimador mínimo cuadrático del modelo (que también coincidirá con el estimador MV bajo el supuesto de normalidad de los residuos).
  - Sin embargo, hay que tener los errores estándar de mínimos cuadrados infraestiman los verdaderos errores estándar, ya que los primeros no incorporan el error en la estimación de  $\lambda$ , que ha sido forzado a un determinado valor antes de la regresión.

<sup>6</sup> En este caso, la transformación se efectúa utilizando la regla de L'Hopital:  $\lim_{\lambda \rightarrow 0} (x_i^\lambda - 1)/\lambda = \ln(x_i)$ .

- Otra alternativa sería estimar el modelo con la transformación Box-Cox a partir del método de máxima verosimilitud. Ahora, las CPO para maximizar el logaritmo de la función de verosimilitud serían:

$$\frac{\partial \ln L}{\partial \beta} = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = 0$$

$$\frac{\partial \ln L}{\partial \lambda} = 0$$

- Desafortunadamente, este método no tiene tampoco solución analítica y hay que recurrir a métodos numéricos.

### 1.2.5. Otros métodos

- **Otros métodos de regresión** que se proponen para el caso no lineal, los cuales no se desarrollan aquí, son:
  - Variables instrumentales no lineales.
  - Mínimos cuadrados no lineales en 2 etapas (véase WILLIAM H. GREENE “Análisis econométrico”).

## 2. VERIFICACIÓN

- Se plantean 2 cuestiones:
  - 1) Contraste de linealidad del modelo.
  - 2) Contraste de hipótesis.

### 2.1. Contraste de linealidad del modelo

- Cuando la forma funcional del modelo no es conocida a priori, entonces existen **3 enfoques** principales para ayudar a identificar no linealidades a partir de la información muestral:
  - i) Análisis de los residuos.
  - ii) Regresiones por tramos.
  - iii) Uso de polinomios de las variables.

#### 2.1.1. Análisis de los residuos: Contraste gráfico

- Graficar la regresión no lineal estimada para confirmar que se ajusta razonablemente bien a la información muestral resulta muy recomendable.
  - También es útil comprobar que la relación obtenida entre la variable explicada y los regresores es consistente con la teoría económica y con la evidencia empírica.

#### 2.1.2. Regresiones por tramos: Contraste de estabilidad de Chow

- El **test de Chow** contrasta la hipótesis nula de ausencia de cambio estructural. Esto es, de que 2 submuestras han sido generadas por una misma estructura económica.
  - Normalmente, este test se aplicará cuando el investigador sospeche que se ha dado una variación estructural en algún momento del periodo muestral observado, que haya podido alterar los coeficientes del modelo.
- El método seguido en el contraste es el siguiente:
  - Se parte de los modelos restringido y sin restringir:
    - Modelo Restringido (MR):
      - $y_t = x_t' \cdot \beta + \varepsilon_t, \quad t = 1, 2, \dots, T_1, \dots, T_2$
    - Modelo Sin Restringir (MSR):
      - $y_t = x_t' \cdot \beta_1 + \varepsilon_t, \quad t = 1, 2, \dots, T_1$
      - $y_t = x_t' \cdot \beta_2 + \varepsilon_t, \quad t = T_1 + 1, \dots, T_2$
  - Nótese que hemos adoptado el subíndice  $t$ , dado que este contraste suele realizarse en un contexto de modelos de *series temporales*. Además, el MSR consta de 2 regresiones, una para cada subperiodo muestral.
  - La hipótesis nula sería:

$$H_0: \beta_1 = \beta_2$$

- Es decir, cada uno de los coeficientes  $\beta_i$  en las 2 submuestras del MSR son iguales, lo que constituye un conjunto de  $k$  hipótesis lineales.

- Teniendo todo lo anterior en consideración el estadístico a utilizar en este contraste es:

$$F = \frac{[SSR - (SSR_1 + SSR_2)]/K}{(SSR_1 + SSR_2)/(T - 2 \cdot K)}$$

que se distribuye como una  $F_{K, T-2K}$  con  $T - 2K$  grados de libertad del MSR (pues el número total de observaciones empleadas será  $K$  considerando ambas submuestras y el número total de parámetros es  $2 \cdot K$  puesto que se estiman los  $K$  parámetros en las 2 regresiones).

### 2.1.3. Uso de polinomios de las variables

#### Contraste RESET de mala especificación (RAMSEY, 1969)

- Cuando tenemos sospechas de no linealidad, debido a la teoría económica o a la información muestral, una forma sencilla de comprobar nuestras sospechas es estimar el modelo lineal y, a continuación, realizar un contraste de mala especificación de la forma funcional.
  - En concreto, destaca el test general de especificación **RESET** (*Regression Specification Error Test*), propuesto por RAMSEY en 1969.
- La idea detrás de este test es sencilla:
  - Si el verdadero modelo es  $y = \beta \cdot X + \varepsilon$ , entonces, ninguna función no lineal de las variables explicativas  $X$  debería ser significativa cuando se añada a la ecuación anterior.
  - El test RESET añade polinomios de estimaciones MCO de las variables endógenas para detectar formas genéricas de mala especificación funcional, como la no linealidad. Por ejemplo:
 
$$y = \beta_0 + \beta_1 \cdot x_1 + \dots + \beta_k \cdot x_k + \delta_1 \cdot \hat{y}^2 + \delta_2 \cdot \hat{y}^3 + \varepsilon$$
  - La idea es que las funciones polinómicas de las estimaciones de  $y$  contienen varias no linealidades de las variables  $x$ , cuya significatividad se quiere comprobar.
  - El test es:
 
$$H_0: \delta_1 = \delta_2 = 0$$

$$H_1: \delta_1 \neq 0 \text{ y/ó } \delta_2 \neq 0 \text{ (problema en la especificación: indicios de no linealidad)}$$
  - En este caso, se utiliza un contraste  $F$  con grados de libertad 2 y  $N - k - 3$ .
- Valoración de la prueba RESET:
  - En la práctica, la prueba RESET puede no ser particularmente buena para detectar alguna alternativa específica para un modelo propuesto, y su utilidad radica en que sirve como indicador general de que algo está mal<sup>7</sup>.
  - Además, una cuestión complicada es decidir cuántas funciones polinómicas de las estimaciones de  $y$  deben incluirse. No hay una respuesta sencilla y dependerá de cada caso.

#### Prueba del multiplicador de Lagrange

- La **prueba del multiplicador de Lagrange** es una alternativa para la prueba RESET de RAMSEY. Para explicar esta prueba partiremos del siguiente ejemplo; buscamos estimar una función de costes, donde dudamos entre una especificación lineal y una cúbica. La primera es una versión restringida de la última, que supone que los coeficientes de los términos de producción elevados al cuadrado y al cubo son iguales a cero. Para comprobar esto, se realiza el siguiente análisis:
  - 1) Se estima la regresión restringida mediante MCO y se obtienen los residuos  $\hat{\varepsilon}_i$ .
  - 2) Si la regresión no restringida resulta ser la verdadera regresión, los residuos obtenidos deberían estar correlacionados con el polinomio de los términos de producción. Por tanto, estimaremos la siguiente regresión:  $\hat{\varepsilon}_i = \alpha_0 + \alpha_1 \cdot X^2 + \alpha_2 \cdot X^3 + v_i$ .
  - 3) Para un tamaño de muestra grande, ENGLE demostró que el tamaño de la muestra multiplicado por  $R^2$  estimado en la regresión auxiliar sigue una distribución chi cuadrado con

<sup>7</sup> Por esta razón, una prueba como RESET se describe en ocasiones como una prueba de especificación en lugar de una función de especificación. Esta distinción es muy sutil, pero la idea básica es que una prueba de especificación examina algún aspecto particular de una ecuación dada, teniendo en mente claras hipótesis nula y alternativa. Una prueba de especificación incorrecta, por su parte, puede detectar varias opciones e indica que algo está al según la hipótesis nula, sin ofrecer necesariamente guía clara en cuanto a la hipótesis alternativa apropiada.



grados de libertad iguales al número de restricciones impuestas en la versión restringida:

$$n \cdot R^2 \sim \chi^2_{n^{\circ} \text{ de restricciones}}$$

- 4) Si el valor de la chi cuadrado obtenida excede el valor crítico de la chi cuadrado, rechazamos la restricción restringida.

## 2.2. Contrastes de hipótesis

De este apartado se puede recortar.

- Los distintos contrastes de hipótesis pueden clasificarse básicamente en:
  - Contrastes con restricciones lineales.
  - Contrastes con restricciones no lineales.
- Independientemente de que se esté contrastando un conjunto de  $q$  restricciones lineales o no lineales, acerca de los coeficientes del modelo de regresión, la hipótesis nula a contrastar será:

$$H_0: R(\beta) = r$$

siendo necesario que las restricciones (lineales o no lineales) sean sobreidentificadas.

- Considerando que no hay diferencias conceptuales considerables entre ambos contrastes citados debe asumirse que  $(\partial R(\beta)/\partial \beta)$  es una matriz  $q \times K$ , con rango completo de fila ( $q$ ), para la que se debe cumplir estrictamente que  $q < K$ .
  - Por tanto, se tienen menos restricciones que parámetros y las restricciones no son redundantes. Puede hacerse una analogía entre esta matriz y la de coeficientes de las restricciones en el modelo lineal.

### 2.2.1. Contraste F

- No obstante, cabe advertir que los contrastos  $t$  y  $F$  quedan invalidados para el caso de modelos no lineales, puesto que tales distribuciones no podrán justificarse de la misma manera.
- Sin embargo, para el caso de no linealidad, puede proponerse un contraste  $F$  que será válido asintóticamente:

$$F = \frac{[SR(\widehat{\beta}_R) - SR(\widehat{\beta})]/q}{SR(\widehat{\beta})/(T - K)}$$

donde  $\widehat{\beta}$  es el estimador de MCNL sin restringir y  $\widehat{\beta}_R$  el estimador MCNL considerando las restricciones mencionadas.

Debe entenderse que ni numerador ni denominador seguirán de manera exacta una distribución  $\chi^2$ , por lo que la distribución  $F$  del estadístico será únicamente aproximada. Es decir, se tiene una  $F_{q, T-K}$  asintótica.

### 2.2.2. Contraste de Wald

- Este test se basa en la distancia existente entre  $R(\widehat{\beta})$  y  $r$ . Si las estimaciones no restringidas no satisfacen las restricciones, surgirá la duda acerca de la validez de las restricciones. El estadístico será:

$$W = [R(\widehat{\beta}) - r] \cdot [Var R(\widehat{\beta})]^{-1} \cdot [R(\widehat{\beta}) - r]$$

que sigue una distribución chi cuadrado con tantos grados de libertad como restricciones a contrastar ( $q$ ).

Bajo la hipótesis nula de que las restricciones son correctas, este estadístico es asintóticamente equivalente a  $q$  veces el estadístico  $F$  que se expuso previamente.

El problema que se daba para este contraste en el modelo lineal se da aquí, de manera análoga. Puesto que este test no emplea la hipótesis alternativa, el estadístico de Wald será distinto según el modo en que se formule la hipótesis. En los casos en los que haya más de una forma de especificar  $R(\widehat{\beta}) = r$ , según cuál de éstas se escoja,  $W$  puede dar resultados bien diferentes.

### 2.2.3. Contraste de la razón de verosimilitud

Suponiendo normalidad en lo que al término de error se refiere se tiene que el logaritmo de la función de verosimilitud es:

$$\ln L = \text{jacobiano} - \frac{N}{2} \cdot \ln(2\pi\sigma^2) - \frac{\varepsilon'\varepsilon}{2\sigma^2}$$

Recordando que  $\hat{\sigma}_{MV}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}$  se tiene la función de verosimilitud concentrada:

$$\ln L = -\frac{N}{2} \cdot \left[ 1 + \ln(2\pi) + \ln\left(\frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}\right) \right]$$

Tras obtener el logaritmo de la función de verosimilitud evaluado en el estimador restringido  $\widehat{\beta}_R$  puede obtenerse el estadístico:

$$\lambda = -2 \cdot (\ln L_R - \ln L)$$

que se distribuye como una  $\chi^2$  con  $q$  grados de libertad.

### 2.2.4. Contraste de multiplicadores de Lagrange

No resulta difícil entender que la suma residual del modelo restringido será siempre mayor a la del modelo sin restringir. Este contraste se basa, precisamente, en la reducción que se tendría en la suma de los residuos al cuadrado si se suprimiesen las restricciones del modelo restringido.

Para la realización del test, se deberá maximizar la función de verosimilitud  $L(\theta)$  sujeta a las restricciones  $R(\beta) = r$ , donde el vector  $\theta$  se compone de los coeficientes del modelo ( $\beta$ ) y de los parámetros de la matriz de covarianzas. El lagrangiano será:

$$\ln L_R(\theta) = \ln L(\theta) + \lambda' \cdot [R(\beta) - r]$$

Teniéndose que  $\lambda$  es un vector columna de dimensión  $q$  compuesto por los multiplicadores de Lagrange, mientras que  $L(\theta)$  y  $L_R(\theta)$  serán las funciones de verosimilitud sin restringir y restringida, respectivamente. Por otro lado, las condiciones necesarias para el problema de maximización deben ser:

$$\begin{aligned} \frac{\partial \ln L(\theta_R)}{\partial \theta_R} &= \left( \frac{\partial \ln L(\theta)}{\partial \theta} \right) + R' \cdot \lambda = 0_K \\ \frac{\partial \ln L(\theta_R)}{\partial \lambda} &= 0_q \end{aligned}$$

Si las restricciones fueran válidas, su imposición no debería reducir de manera significativa el valor máximo de la función de verosimilitud y, por tanto, las derivadas  $\partial \ln L_R(\theta_R)/\partial \theta_R$  y  $\partial \ln L(\theta)/\partial \theta$  deberían ser muy similares.

En consecuencia, el vector  $R' \cdot \lambda$  y, en particular, el valor óptimo de los multiplicadores de Lagrange, debería ser muy reducido (es decir, se está contrastando que  $H_0: \lambda = 0$ ):

$$\frac{\partial \ln L(\theta_R)}{\partial \theta_R} \approx \left( \frac{\partial \ln L(\theta)}{\partial \theta} \right) = -R' \cdot \lambda \approx 0_K$$

El estadístico a emplear en este contraste será:

$$\begin{aligned} LM &= [\partial \ln L(\widehat{\theta}_R)/\partial \widehat{\theta}_R]' \cdot [I(\widehat{\theta}_R)]^{-1} \\ LM &= [\partial \ln L(\widehat{\theta}_R)/\partial \widehat{\theta}_R] \end{aligned}$$

Bajo la hipótesis nula, LM seguirá una distribución  $\chi^2$  con  $q$  grados de libertad. Todos los términos que figuren en la expresión se evalúan con el estimador restringido.

Una vez expuesto el contraste de una manera formal, hay que señalar que, en el caso de regresiones no lineales, LM adopta una forma particularmente atractiva y muy útil en el trabajo aplicado:

$$\begin{aligned} LM &= \frac{\tilde{\varepsilon}' \tilde{X}' (\tilde{X}' \tilde{X})^{-1} \tilde{X} \tilde{\varepsilon}}{\frac{\tilde{\varepsilon}' \tilde{\varepsilon}}{N}} \\ LM &= N \cdot \tilde{R}^2 \end{aligned}$$

donde:

- $\tilde{\varepsilon} = [y_i - f(x_i, \hat{\beta}_R)]$ : Vector de residuos calculado utilizando las estimaciones restringidas.
  - $\tilde{X}$ : Matriz  $N \times K$  de observaciones del gradiente de la función  $f$ , evaluado en los estimadores restringidos.
  - $\tilde{R}^2$ : Coeficiente de determinación, no centrado, de la regresión del vector de residuos sobre la matriz de variables explicativas  $\tilde{X}$ .
- Una gran ventaja de este test es que sólo se precisa del estimador restringido, con lo que el contraste de multiplicadores de Lagrange cobrará especial interés en los casos en que la imposición de las restricciones conduzca a un modelo más sencillo que el que se tuviese originalmente.

### 3. PREDICCIÓN

- La utilización de una aproximación lineal cuando el verdadero modelo es no lineal puede conducir a errores de predicción muy elevados.
- Sin embargo, como reconoce el *International Journal of Forecasters* en una review de 2006 sobre predicción, la literatura relativa a predicción con modelos no lineales está *todavía en una fase muy incipiente*.
  - Los artículos más importantes se han concentrado en comparar la capacidad predictiva en términos del Error Cuadrático Medio de Predicción (ECMP) de métodos lineales frente a métodos como *Bootstrapping*, simulaciones de Monte Carlo o de redes neuronales.
    - Si bien autores como DE GOOIJER defienden que los modelos no lineales son más eficientes, otros consideran que el aumento de la complejidad en el cálculo no necesariamente arroja mejores resultados.

### 4. MODELOS DE REGRESIÓN NO LINEAL: RESPUESTA CUALITATIVA

#### 4.1. Idea: Naturaleza de los modelos de respuesta cualitativa

- Los modelos de respuesta cualitativa son aquellos donde, en su versión más simple, nos enfrentamos a una respuesta dicotómica, que será representada con los valores 0 y 1.
  - Por ejemplo, estas serán las respuestas obtenidas cuando a través de encuestas como la EPA preguntemos a la muestra si en el momento actual se encuentran desempleados (0) o trabajando (1).

#### 4.2. Posibilidades de estimación

##### 4.2.1. Modelo lineal de probabilidad (MLP)

- El modelo ordinal de probabilidad buscará darnos una esperanza de la probabilidad condicional de la variable dependiente dado el vector  $X$ :  $Pr(Y_i = 1|X_i)$ . Dado que  $Y$  sólo toma valores 0 y 1, diremos que sigue una distribución de Bernoulli. Éstos modelos tendrán las siguientes **características**:
  - No normalidad de las perturbaciones: Aunque MCO no requiere que las perturbaciones se encuentren normalmente distribuidas, su existencia es deseable para fines predictivos. En esta ocasión, las perturbaciones seguirán, de nuevo, una distribución de Bernoulli.
  - Varianzas heterocedásticas de las perturbaciones: En una distribución de Bernoulli, la media y la varianza teóricas son  $p$  y  $p \cdot (1 - p)$ , respectivamente, donde  $p$  es la probabilidad de éxito (*de que algo suceda*), lo cual revela que la varianza es una función de la media. Por tanto, la varianza del error es heterocedástica.
    - Ello nos obligará a estimar la regresión a través de mínimos cuadrados ponderados.
  - Valor cuestionable del  $R^2$  como medida de bondad de ajuste: El valor pronosticado se suele encontrar entre 0 y 1, pero nunca será ninguno de estos dos. Por tanto, el éxito del modelo no debe de ser valorado a través de la bondad de ajuste, sino a través del éxito de la predicción.
  - No cumplimiento de  $0 \leq E[Y_i|X_i] \leq 1$ : Como  $E[Y_i|X_i]$  en los modelos lineales de probabilidad mide la probabilidad condicional de que ocurra el suceso  $Y$  dado  $X$ , ésta debería encontrarse necesariamente entre 0 y 1. Sin embargo, con el modelo lineal de probabilidad, no hay garantía

de que los valores predichos de  $Y$  se encuentren en este rango. Éste es el verdadero problema de estimación del *MLP* por *MCO*.

- El modelo es inherentemente no lineal, por ello su estimación ha descansado sobre aproximaciones como los modelos *logit* o *probit*, que pasamos a ver.

#### 4.2.2. Modelo Logit

##### Definición

- Supondremos que la variable dependiente sigue una función logística como la siguiente:

$$P_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot X_{1i} + e_i)}} \Rightarrow P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}}$$

- Si por otra parte consideramos que:  $1 - P_i = \frac{1}{1 + e^{Z_i}}$ , entonces podemos reescribir la ratio de probabilidades como:  $\frac{P_i}{1 - P_i} = e^{Z_i}$ .
- Tomando logaritmos, podemos estimar la ecuación por *MCO*:  $L_i = \ln \frac{P_i}{1 - P_i} = \beta_0 + \beta_1 \cdot X_{1i}$ .

##### Propiedades

- A medida que  $Z$  aumenta de  $-\infty$  a  $\infty$ ,  $P$  aumenta de 0 a 1. Es decir, los valores predichos se encuentran acotados entre  $[0, 1]$ .
- En consecuencia, a diferencia del *MLP*, las probabilidades no serán lineales en  $X$ .
- La interpretación de los coeficientes será la siguiente: cada estimador mide el cambio en  $L$ , es decir, el logaritmo de probabilidades, como consecuencia de un cambio unitario en  $X$ .

#### 4.2.3. Modelo Probit

##### Definición

- Para explicar el comportamiento de una variable dependiente dicotómica es preciso utilizar una función de distribución acumulativa (*FDA*) seleccionada apropiadamente. El modelo *logit* toma una función logística, mientras que el **modelo probit** tomará una distribución normal.
  - Para estimar el modelo probit supondremos que la decisión dicotómica por parte del individuo  $i$  dependerá de un **índice de conveniencia** no observable  $I_i$ , también conocido como variable latente, determinado por una o varias variables explicativas. Expresaremos tal índice como:  $I_i = \beta_0 + \beta_1 \cdot X_{1i} + e_i$ .
  - Además, supondremos que existe un **nivel crítico** o **umbral del índice** ( $I^*$ ) tal que, si  $I > I^*$ , entonces  $Y = 1$ ; si es inferior, entonces  $Y = 0$ . Si bien el umbral  $I^*$  y el índice  $I$  no son observables, si suponemos que se distribuye de manera normal con las mismas media y varianza, será posible estimar el índice de conveniencia, tal que:  $P_i = P(Y = 1|X) = P(I_i^* \leq I_i) = P(Z_i \leq \beta_0 + \beta_1 \cdot X_{1i}) = F(\beta_0 + \beta_1 \cdot X_{1i})$ .
  - Dado que:  $F(I_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-\frac{z^2}{2}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_0 + \beta_1 \cdot X} e^{-\frac{z^2}{2}} dz \Rightarrow I_i = F^{-1}(I_i) = F^{-1}(P_i) = \beta_0 + \beta_1 \cdot X$ .

##### Efectos marginales

- El resultado de la regresión puede ser difícil de interpretar.
  - Supongamos que deseamos averiguar el efecto de una unidad de  $X$  sobre la probabilidad de que  $Y = 1$ . Para ello, deberemos de calcular lo siguiente:  $\partial P_i / \partial X_i = f(\beta_0 + \beta_1 \cdot X_i) \cdot \beta_1$ .

#### 4.3. Valoración

- Entre un modelo *logit* y uno *probit*, ¿Cuál deberíamos de elegir?
  - Para la mayoría de aplicaciones, los 2 modelos son muy similares. La principal diferencia es que la distribución logística tendrá unas colas un poco más anchas. Ello implicará que la probabilidad condicional de  $P_i$  se aproximará de 0 a 1 con una tasa menor en el modelo *logit* en comparación con el *probit*.

- Ahora bien, aunque los 2 modelos sean similares, se debe tener cuidado al interpretar los coeficientes estimados de cada modelo, pues no serán análogos.

## CONCLUSIÓN

- En la teoría económica, se conoce un amplio abanico de relaciones de dependencia no lineales entre variables. Sin embargo, el supuesto clásico de relación de dependencia lineal (S1) ha tenido tradicionalmente un peso importante en la mayoría de trabajos econométricos porque facilita mucho las cosas al investigador y porque:
  - Se ha demostrado que, habitualmente, se podrá hacer una aproximación lineal a un modelo de regresión no lineal.
  - En los casos de no linealidad que no afecten a coeficientes, podemos retomar el modelo lineal con unas simples transformaciones.
- Sin embargo, en los casos de no linealidad en los coeficientes del modelo, será necesario introducir técnicas de estimación no lineales. En este sentido, tradicionalmente, trabajar con una relación de dependencia intrínsecamente no lineal podía suponer problemas técnicos considerables en la especificación y estimación de los modelos. Afortunadamente, en la actualidad, existen diversos paquetes informáticos que cuentan con la posibilidad de aplicar el estimador de MCNL, lo cual facilita enormemente esta labor y supone un gran avance para el estudio empírico.
- En definitiva, la clave para decidir si utilizar o no un modelo no lineal está en la especificación del modelo. En muchos trabajos empíricos, la relación entre las variables explicada y las explicativas no será lineal en los coeficientes del modelo y, en consecuencia, se deberá especificar la forma funcional que mejor se adecue a tal relación. El investigador deberá hacer uso de la teoría y de sus conocimientos e información disponibles para discernir cuál es el modelo de regresión correcto a estimar. A partir de ahí, las dificultades serán menores y sólo quedará verificar que los resultados son los que se esperaban.
- Por último, STOCK y WATSON subrayan la importancia de considerar la posible interacción no lineal entre distintas variables independientes de un mismo modelo. En concreto, el efecto que se produce debido a un cambio en determinada variable independiente depende del valor de otra variable independiente distinta.





## 5.B.6 : MODELOS ARIMA: IDENTIFICACIÓN, ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del *método deductivo contrastable*<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

### ▪ **Contextualización:**

- En Econometría aplicada, se trabaja con 3 tipos de datos muestrales:
  - Datos de sección cruzada: Conjunto de datos recogidos observando diversas unidades económicas (como familias, empresas, ciudades...) en un mismo instante de tiempo o haciendo abstracción de la dimensión temporal.
  - Datos de series temporales: Conjunto ordenado de datos sobre una unidad económica a lo largo del tiempo.
  - Datos de panel (combinación de los 2 anteriores): Datos de sección cruzada observados a lo largo de varios periodos de tiempo.
- Esta exposición se centra en el análisis de **series temporales**. Podemos distinguir 2 grandes **tipos de modelos dinámicos uniecuacionales**:
  - Modelos univariantes: La evolución de una variable se explica utilizando sólo su propio pasado, a través de modelos ARIMA, donde la evolución de una variable se explica utilizando su propio pasado.
  - Modelos multivariantes: Modelos con variables explicativas.
    - En esta exposición, nos vamos a centrar en los modelos univariantes.
- Desde un punto de vista histórico,
  - Tradicionalmente, las series temporales se veían como un proceso determinista.
  - Esto cambia a partir de las aportaciones de YULE (1927), que interpreta las series temporales como la realización de un proceso estocástico, lo cual tiene importantes implicaciones para su modelización a partir de modelos probabilísticos, entre los que destacan los *modelos ARIMA*.
  - En 1970, BOX y JENKINS propusieron una metodología que se ha convertido en una herramienta habitual en el análisis de series económicas y que permite ajustar los modelos ARIMA a series temporales reales. La metodología consiste en un proceso iterativo en **3 etapas**:
    - 1) Identificación: Identificar el posible modelo ARIMA que sigue una determinada serie temporal requiere:
      - Decidir qué transformaciones aplicar para convertir la serie observada en una serie estacionaria.
      - Determinar un modelo ARMA para dicha serie estacionaria, que implica identificar:
        - Los órdenes  $p$  y  $q$  de las estructuras autorregresiva y de media móvil, respectivamente.
        - Si el proceso es estacional, los órdenes  $P$  y  $Q$  de la estructura ARMA estacional.
    - 2) Estimación: Una vez seleccionado provisionalmente el modelo ARMA, los parámetros de las partes AR y MA se estiman por máxima verosimilitud (u otros métodos), y se obtienen sus errores estándar y los residuos del modelo.
    - 3) Verificación: Finalmente, se verifica que los residuos carecen de una estructura de dependencia y siguen un proceso de ruido blanco:
      - Si los residuos no contienen información, aceptamos el modelo como correcto y lo utilizaremos para la previsión o la toma de decisiones.
      - Si los residuos presentan estructura, modificaremos el modelo para incorporarla a éste y repetiremos, nuevamente, las 3 etapas hasta obtener el modelo adecuado.

### ▪ **Problemática (Preguntas clave):**

- En esta exposición, procederemos en 3 apartados:
  - En primer lugar, introduciremos qué es un modelo ARIMA.

- En segundo lugar, analizaremos en más detalle cada una de las distintas etapas propuestas por BOX y JENKINS.
- Finalmente, una vez hemos asegurado que el modelo ARIMA ha sido adecuadamente ajustado, veremos también cómo puede usarse para la predicción.

▪ **Estructura:**

## 1. ASPECTOS GENERALES SOBRE MODELOS ARIMA

### 1.1. Procesos estocásticos

- 1.1.1. Definición de proceso estocástico
- 1.1.2. Estacionariedad (estricta y débil)

### 1.2. Ejemplos de procesos estocásticos (Ruido blanco, Paseo aleatorio, AR y MA)

### 1.3. Procesos ARMA como aproximación de la representación de Wold

- 1.3.1. Teorema de la descomposición de Wold
- 1.3.2. Procesos ARMA

### 1.4. Procesos integrados

### 1.5. Conclusión: Procesos ARIMA

## 2. METODOLOGÍA PARA EL ANÁLISIS DE PROCESOS ARIMA DE BOX Y JENKINS (1970)

### 2.1. Identificación

- 2.1.1. Transformaciones para lograr la estacionariedad
  - 1) Transformación para estabilizar la varianza
  - 2) Diferenciación para estabilizar la media
- 2.1.2. Especificación de la estructura ARMA

### 2.2. Estimación

- 2.2.1. Estimación por MCO (para procesos AR puros)
- 2.2.2. Estimación por MV (para procesos MA puros y ARMA)
  - Procedimiento de MV
  - Complejidad
  - Propiedades de los estimadores del método de máxima verosimilitud
- 2.2.3. Método Generalizado de los Momentos (MGM)
- 2.2.4. Enfoque bayesiano

### 2.3. Verificación

- 2.3.1. Contraste de media cero (i.e.  $E\epsilon_t = 0 \quad \forall t = 1, 2, \dots$ )
- 2.3.2. Contraste de homocedasticidad (i.e.  $\text{Var}\epsilon_t = \sigma^2 \quad \forall t = 1, 2, \dots$ )
- 2.3.3. Contrastes de ausencia de autocorrelación residual (i.e.  $\text{Cov}\epsilon_t, \epsilon_{t-k} = 0 \quad \forall k = 1, 2, \dots$ )
  - 1) Análisis de la Función de Autocorrelación y del correlograma
  - 2) Significatividad estadística de los coeficientes de autocorrelación
- 2.3.4. Contraste de normalidad (i.e.  $\epsilon_t \sim N(0, \sigma^2)$ )
- 2.3.5. Otros contrastes
  - Prueba de raíz unitaria
  - Criterios de información (AIC, BIC)

## 3. PREDICCIÓN

### 3.1. Predictor óptimo

- 3.1.1. Concepto de predictor
- 3.1.2. Predictor óptimo (criterio de minimización del ECMP)

### 3.2. Cálculo de las predicciones

### 3.3. Varianza de la predicción e intervalos de predicción

## 1. ASPECTOS GENERALES SOBRE MODELOS ARIMA

- El acrónimo **ARIMA** proviene de “*Autoregressive Integrated Moving Average*”. Por tanto, los modelos ARIMA son representaciones teóricas de:
  - **Procesos estocásticos**,
  - **Autorregresivos**,
  - **De media móvil**,
  - **Integrados**.
    - Para entender en qué consisten, podemos analizar uno a uno los elementos que componen esta definición.

### 1.1. Procesos estocásticos

#### 1.1.1. Definición de proceso estocástico

- Llamamos **proceso estocástico** a un conjunto de variables aleatorias  $\{y_t\}$ . En nuestro caso, este conjunto es ordenado y corresponde a los instantes temporales (días, meses, años, etc.). Para cada valor  $t$ , está definida una variable aleatoria (i.e.  $y_1, y_2, \dots$ ). Los valores observados de las variables aleatorias en estos distintos instantes conforman una *serie temporal*.
  - En el análisis estadístico, se suele interpretar la serie temporal como una realización del proceso estocástico, es decir, una muestra. En este punto, es importante distinguir 2 casos:
    - En *situaciones experimentales*, es posible obtener varias series temporales (varias muestras) del mismo proceso estocástico.
      - Por ejemplo, podemos estudiar cómo evoluciona el tamaño de una planta en los meses 1, 2, ..., 12 desde que se planta la semilla ( $t = 0$ ) y volver a repetir el experimento en las mismas condiciones.
    - Sin embargo, en la mayoría de *aplicaciones económicas*, sólo podemos observar una realización del proceso.
      - Por ejemplo, la serie de crecimiento anual del PIB es una realización concreta de un proceso estocástico, pero no podemos volver atrás en el tiempo para observar otra realización.
      - En este último caso, el objetivo será estimar las características “transversales” del proceso (medias, varianzas, etc.) a partir de su evolución “longitudinal”.
        - Para que ello sea posible, es necesario suponer que las propiedades transversales (distribución de cada variable en cada instante) son estables a lo largo del tiempo. Aquí entra el concepto de **estacionariedad**.

#### 1.1.2. Estacionariedad (estricta y débil)

- Podemos distinguir 2 tipos de **estacionariedad**:
  - Estacionariedad en sentido estricto: Un proceso estocástico  $\{y_t\}$  es estacionario en sentido estricto si para toda  $m$ -tupla  $(t_1, t_2, \dots, t_m)$  y todo entero  $k$ , el vector de variables  $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$  tiene la misma distribución de probabilidad conjunta que el vector  $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$ .
    - En otras palabras, la función de distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo  $k$  periodos. Esto implica que todos los momentos que caracterizan a los 2 conjuntos son iguales (esperanza, varianza/covarianza, asimetría, curtosis...).
    - La estacionariedad estricta es una condición muy fuerte y excesiva para nuestras necesidades prácticas. Por ello, generalmente, nos conformamos con un concepto menos exigente.

- Estacionariedad en sentido débil (o de 2º orden): Se produce cuando todos los momentos de 1ª y 2ª orden del proceso estocástico son invariantes en el tiempo. Estos momentos incluyen la esperanza matemática y la varianza de las variables  $y_t$ , pero también las covarianzas y correlaciones entre diversos retardos<sup>3</sup>.
  - Algunas de estas condiciones son fácilmente contrastables. Por ejemplo, existirán indicios claros de no estacionariedad cuando una serie temporal muestre:
    - Una tendencia creciente (*esperanza no constante*).
    - Fluctuaciones de amplitud creciente en el tiempo (*varianza no constante*).

## 1.2. Ejemplos de procesos estocásticos (Ruido blanco, Paseo aleatorio, AR y MA)

- Podemos mencionar los siguientes **ejemplos de procesos estocásticos**:

- i) Ruido blanco: Se dice que un proceso es puramente aleatorio si tiene media cero, una varianza constante y no está serialmente correlacionado. Por tanto, en estos procesos conocer los valores pasados no nos proporciona información alguna sobre el futuro, ya que el proceso no tiene memoria.
- ii) Paseo aleatorio: El paseo aleatorio constituye el ejemplo de proceso no estacionario más conocido.
  - Paseo aleatorio sin deriva: Es un proceso donde el valor presente es igual al valor pasado más un proceso de ruido blanco.
    - En relación con su forma:
      - El proceso se define como  $y_t = y_{t-1} + \varepsilon_t$ , y puede ser descrito como  $y_T = y_0 + \sum_{t=0}^T \varepsilon_t$ .
      - Si tomamos esperanzas:  $E[y_T] = y_0$ .
      - Su varianza es:  $\text{var}[y_t] = t \cdot \sigma^2$ .
    - En relación con sus características:
      - Una de las características más importantes de estos procesos es la *persistencia de los shocks aleatorios*. Estos procesos tienen *memoria infinita*, lo que implica que el efecto de un shock no desaparece a lo largo del tiempo. A la suma de los términos error se le conoce como la tendencia estocástica.
      - Debido a que estas series no son estocásticas, deberemos realizar un tratamiento previo de los datos antes de proceder a su estimación.
    - Este proceso ha sido utilizado en la teoría económica para defender tesis como la *hipótesis de los mercados eficientes*: al analizar la evolución de las cotizaciones de los títulos de renta variable se constata que éstas siguen un paseo aleatorio. Autores como FAMA han argumentado que ello supone un ejemplo de la eficiencia de los mercados: las innovaciones se derivan de información que previamente no se encontraba disponible.
  - Paseo aleatorio con deriva: Estos procesos, además de tener una tendencia estocástica, tienen una tendencia determinista.
    - Por tanto, el proceso adopta la siguiente forma:
      - El proceso puede ser definido como:  $y_t = \delta + y_{t-1} + \varepsilon_t$ .
      - En estos procesos, la esperanza del mismo es igual a:  $E[y_t] = y_0 + t \cdot \delta$ .
    - De nuevo, para analizar estos datos, será necesario eliminar el componente tendencial de la serie.
- iii) AR(1) (generalizable a AR(p)): Diremos que una serie  $y_t$  sigue un **proceso autorregresivo** de primer orden si ha sido generada por:  $y_t = c + \varphi \cdot y_{t-1} + \varepsilon_t$ , donde  $-1 < \varphi < 1$  y el componente  $\varepsilon_t$  es un proceso de ruido blanco.
  - Estos procesos presentarán una media igual a:

$$E[y_t] = E[y_{t-1}] \Rightarrow \mu = c + \varphi \cdot \mu \Rightarrow \mu = c/(1 - \varphi)$$

<sup>3</sup> Nótese que la estacionariedad estricta implica la estacionariedad débil, pero, en general, la dirección de esta implicación no funciona al revés. Existe, sin embargo, un caso especial: si la distribución conjunta de las variables aleatorias sigue una distribución normal multivariante (que se caracteriza sólo por sus dos primeros momentos), entonces la estacionariedad débil implica también estacionariedad estricta.

- Para un proceso AR(1) con media cero ( $\mu = c = 0$ ), su varianza será igual a:

$$E[y_t^2] = \varphi^2 \cdot E[y_t^2] + 2 \cdot \varphi \cdot E[y_{t-1} \cdot \varepsilon_t] + E[\varepsilon_t^2] \Rightarrow \sigma_y^2 = \varphi^2 \cdot \sigma_y^2 + \sigma^2 \Rightarrow \sigma_y^2 = \sigma^2 / (1 - \varphi^2)$$

iv) MA(1) (generalizable a MA(q)): Se define el **proceso de media móvil** de orden 1 añadiendo a un proceso de ruido blanco una dependencia del valor actual de la serie de la última innovación ocurrida. De esta manera, el proceso será una combinación lineal de las 2 últimas innovaciones, de acuerdo con la ecuación:  $y_t = \varepsilon_t + \theta \cdot \varepsilon_{t-1}$ .

- Este proceso es la suma de 2 procesos estacionarios, por lo que el resultado siempre será estacionario.
- Si suponemos que  $-1 < \theta < 1$ , entonces diremos que el proceso es invertible y que tiene la propiedad de que el efecto de los valores pasados decrece con el tiempo.

- La invertibilidad puede demostrarse sustituyendo  $\varepsilon_{t-1}$  por  $y_{t-1}$ :

$$\left. \begin{aligned} y_t &= \varepsilon_t + \theta \cdot \varepsilon_{t-1} \\ y_{t-1} &= \varepsilon_{t-1} + \theta \cdot \varepsilon_{t-2} \Rightarrow \varepsilon_{t-1} = y_{t-1} - \theta \cdot \varepsilon_{t-2} \end{aligned} \right\} \Rightarrow y_t = \varepsilon_t + \theta \cdot (y_{t-1} - \theta \cdot \varepsilon_{t-2})$$

$$y_t = \varepsilon_t + \theta \cdot (y_{t-1} - \theta \cdot (y_{t-2} - \theta \cdot \varepsilon_{t-3}))$$

$$y_t = \varepsilon_t + \theta \cdot y_{t-1} - \theta^2 \cdot y_{t-2} + \theta^3 \cdot y_{t-3} - \dots$$

- Si continuamos la inducción hacia atrás:

$$y_t = \varepsilon_t - \sum_{i=1}^{t-1} (-\theta)^i \cdot y_{t-i}$$

- De esta manera observamos 2 cosas:
    - Que hemos conseguido invertir un proceso MA y representarlo en forma de proceso AR.
    - Que, a medida que retrocedemos en el tiempo, el efecto de las innovaciones pasadas en el proceso de ruido blanco sobre la variable  $y_t$  se reduce hasta converger a 0.
- Este no es más que un ejemplo particular del *Teorema de la descomposición de Wold*, que pasamos a ver.

### 1.3. Procesos ARMA como aproximación de la representación de Wold

#### 1.3.1. Teorema de la descomposición de Wold

- El **Teorema de la descomposición de Wold** establece que *todo proceso estacionario débil, de media finita, que no contenga componentes deterministas, puede escribirse bajo una representación de media móvil MA( $\infty$ ), es decir, una combinación lineal de retardos de un ruido blanco.*

- Analíticamente:

$$y_t = \sum_{i=0}^{+\infty} \psi_i \cdot \varepsilon_{t-i} \text{ con } \psi_0 = 1 \text{ y } \sum_{i=0}^{+\infty} \psi_i^2 < \infty$$

- Utilizando la notación del polinomio de operador de retardos:

$$y_t = \psi_\infty(L) \cdot \varepsilon_t$$

- Este resultado es importante, pues implica que siempre existirá una forma de especificar el modelo. El problema es que podríamos necesitar estimar un gran número de parámetros (¡en muchas ocasiones un número infinito!).



### 1.3.2. Procesos ARMA

- Los **procesos ARMA** son una aproximación a la representación de Wold que es más parsimoniosa (es decir, requiere menos parámetros).
  - Bajo condiciones muy generales, el polinomio de retardos infinitos de Wold ( $\psi_\infty(L)$ ) puede ser aproximado como el cociente de 2 polinomios con un n° finito de retardos. Así, tendríamos:

$$y_t \simeq \underbrace{\frac{\theta_q(L)}{\varphi_p(L)}}_{\simeq \psi_\infty(L)} \cdot \varepsilon_t$$

- Por tanto, admitiendo que también puede existir una constante, los procesos ARMA pueden representarse:

$$\begin{aligned}\varphi_p(L) \cdot y_t &\simeq \theta_q(L) \cdot \varepsilon_t \\ \varphi_p(L) \cdot y_t &= \alpha + \theta_q(L) \cdot \varepsilon_t\end{aligned}$$

donde:

- $\alpha$  es una constante.
- $\varphi_p(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$  es el polinomio de retardos del componente AR.
- $\theta_q(L) = 1 - \theta_1 L - \dots - \theta_q L^q$  es el polinomio de retardos del componente MA.
- En notación normal:

$$y_t - \varphi_1 \cdot y_{t-1} - \dots - \varphi_p \cdot y_{t-p} = \alpha + \varepsilon_t + \theta_1 \cdot \varepsilon_{t-1} + \dots + \theta_q \cdot \varepsilon_{t-q}$$

$$y_t = \alpha + \underbrace{\varphi_1 \cdot y_{t-1} + \dots + \varphi_p \cdot y_{t-p}}_{AR(p)} + \varepsilon_t + \underbrace{\theta_1 \cdot \varepsilon_{t-1} + \dots + \theta_q \cdot \varepsilon_{t-q}}_{MA(q)}$$

- Para que el proceso sea *estacionario*, se requiere sólo que las raíces del polinomio del componente AR queden fuera del círculo unidad.
- Además, para que sea *invertible* (i.e. que admita representación AR), se requiere que las raíces del componente MA también queden fuera del círculo unidad.

### 1.4. Procesos integrados

- Cuando un proceso no es estacionario, es habitual que su primera diferencia sí lo sea. Es decir:
  - $y_t$  no es estacionario.
  - $\Delta y_t = y_t - y_{t-1}$  es estacionario [siendo  $\Delta$  el operador de diferencias de orden 1].
    - Decimos entonces que el proceso  $y_t$  está **integrado** de orden 1, lo que se representa  $I(1)$ .
- Generalizando, un proceso  $y_t$  está integrado de orden  $d$ , es decir, es  $I(d)$ , cuando el proceso  $\Delta^d y_t$  (diferenciado  $d$  veces), es un proceso estacionario<sup>4</sup>.

### 1.5. Conclusión: Procesos ARIMA

- Por tanto, un proceso  $ARIMA(p,d,q)$  es un proceso estocástico que, cuando es diferenciado  $d$  veces, adquiere la forma de un proceso estocástico estacionario con forma  $ARMA(p,q)$ .
  - Así, los modelos  $ARIMA(p,d,q)$  adoptan la siguiente forma general:

$$\varphi_p(L) \cdot \Delta^d y_t = \alpha + \theta_q(L) \cdot \varepsilon_t$$

## 2. METODOLOGÍA PARA EL ANÁLISIS DE PROCESOS ARIMA DE BOX Y JENKINS (1970)

- Una vez estudiadas las distintas representaciones teóricas de los procesos estocásticos, podemos estudiar *cómo encontrar aquella representación que mejor se aproxima al verdadero proceso generador de datos que ha dado lugar a una determinada serie temporal observada*.
  - Para ello, seguimos el procedimiento iterativo de BOX y JENKINS ya comentado en 3 etapas:
    - Identificación.
    - Estimación.

<sup>4</sup> Nótese que un proceso estacionario está integrado de orden 0, es decir, es  $I(0)$ .

## 3) Verificación.

## 2.1. Identificación

- En la primera etapa de **identificación**, tendremos que:
  - Primero, determinar la estructura no estacionaria de la serie (si existe) para transformarla en una serie estacionaria.
  - Segundo, determinar la estructura ARMA de la serie estacionaria.

2.1.1. Transformaciones para lograr la estacionariedad

- Para obtener un proceso estacionario (débil), puede ser necesario:
  - Transformar la serie para que tenga varianza constante.
  - Diferenciar la serie para que tenga media constante.

1) Transformación para estabilizar la varianzaa) Heterocedasticidad dependiente del nivel de la serie

- En las series reales, es frecuente observar que la variabilidad es mayor cuando la serie toma valores altos que cuando toma valores bajos.
  - **Detección:** ¿Cómo detectar este problema? Una primera aproximación consiste en realizar un gráfico que relacione una medida de variabilidad (como la desviación típica) con una medida del nivel (como la media local)<sup>5</sup>. Es lo que se conoce como test rango-media.
  - **Solución:**
    - En caso de que la variabilidad de la serie *aumente linealmente con el nivel de la serie*, podríamos conseguir un proceso con variabilidad constante tomando logaritmos. De esta manera, se obtiene una serie en la que la variabilidad no depende del nivel.
    - La transformación logarítmica es sólo un caso particular de la familia de transformaciones Box-Cox. Puede demostrarse que las variables originales pueden transformarse mediante:
 
$$y_t^* = g(y_t) = \frac{y_t^\lambda - 1}{\lambda}$$
      - Existiendo un valor de  $\lambda$  que permite que la desviación típica de la variable transformada  $y_t$  sea constante. Para el caso particular anterior, la transformación logarítmica correspondería al caso en el que  $\lambda$  tiende a 0.

b) Otros casos de heterocedasticidad

- Existen series heterocedásticas en las que la varianza cambia sin relación con el nivel. Por ejemplo:
  - Valores atípicos: Algunas observaciones tienen mucha mayor varianza que las restantes o aparecen cambios de varianza por tramos.
  - Heterocedasticidad condicional: La varianza de la serie puede ser constante, pero la varianza condicional no serlo. En este caso, habría que aplicar *modelos ARCH* (*modelos autorregresivos con heterocedasticidad condicional*).

2) Diferenciación para estabilizar la media

- Se aplicarán diferencias regulares y/o estacionales según lo que la serie requiera, para tener una *media estable*.
  - a) Orden de diferenciación regular
- A la hora de tomar la decisión de aplicar o no diferencias:
  - Es útil comenzar analizando el gráfico de la propia serie
    - Si se observa que ésta muestra tendencia o cambios en el nivel de la media, *diferenciaremos* para transformarla en estacionaria.

<sup>5</sup> Por ejemplo, si disponemos de datos mensuales, podemos calcular las medias y desviaciones típicas de los datos de cada año y comparar ambos en un gráfico.

- Si el gráfico de la serie presentara *dudas*, conviene estudiar la Función de Autocorrelación Simple (FAS).
  - Una serie no estacionaria debe mostrar autocorrelaciones positivas con decrecimiento lento y lineal. En este sentido, si la FAS no se amortigua para retardos altos (mayores de 15 o 20), es necesario diferenciar para obtener un proceso estacionario.
- ¿Cuántas diferencias es conveniente realizar a una serie si se tienen dudas sobre su estacionariedad?
  - Si el objetivo es la predicción, ante la duda, *lo más recomendable es diferenciar*, dado que las consecuencias negativas de sobre-diferenciar son mucho menores que las de infra-diferenciar:
    - Si incluimos una diferencia de menos, y suponemos que la serie es estacionaria cuando en realidad no lo es, los *errores de predicción a medio plazo suelen ser enormes*, ya que la predicción de una serie estacionaria a medio plazo es su media, mientras que una serie no estacionaria puede alejarse indefinidamente de dicho valor y el error de predicción no está acotado.
    - Si incluimos una diferencia de más, ello implicará añadir un término de media móvil no invertible, por lo que *se perderá eficiencia al tener que estimar un parámetro de más*. Sin embargo, las *estimaciones serán consistentes* y el *efecto sobre las predicciones será pequeño*.
      - En definitiva, diferenciando, sea lo necesario o por exceso, siempre tendremos estimadores consistentes de los parámetros y errores de predicción acotados, mientras que no será el caso para series no diferenciadas suficientemente (y, por tanto, no estacionarias)<sup>6</sup>.

#### b) Orden de diferenciación estacional

- Si la serie presenta *componente estacional*, se aplicará una diferencia estacional para conseguir un proceso estacionario:  $\Delta_s = 1 - L^S$ .
- La **estacionalidad** puede detectarse:
  - En el gráfico de la serie, presentando una pauta repetida de periodo s. Por ejemplo, con estacionalidad mensual, veremos que algunos meses son sistemáticamente altos y otros bajos.
  - En la Función de Autocorrelación Simple (FAS), que presentará coeficientes positivos que decrecen lentamente en los retardos  $s, 2s, 3s \dots$
- En principio, sería teóricamente posible que una serie requiera más de una diferencia estacional, pero en la práctica, esto es extremadamente raro, por lo que la decisión se reduce en realidad a aplicar o no una diferencia estacional.

#### 2.1.2. Especificación de la estructura ARMA

- Una vez hemos logrado la estacionariedad de la serie, el siguiente paso en la identificación es especificar la estructura ARMA. Esto se realizará comparando las *estimaciones de la FAS y la FAP* con *funciones teóricas* de procesos  $ARMA(p,q)$ .
  - Para ello, en primer lugar, cabe distinguir entre:
    - Función de Autocorrelación Simple (FAS): La FAS nos proporciona la correlación que existe entre las observaciones separadas por  $k$  periodos (*autocorrelación de orden  $k$* ) y se definirá como:  $\rho_k = \frac{cov(y_t, y_{t-k})}{cov(y_t, y_t)} = \frac{cov(y_t, y_{t-k})}{var(y_t)}$ . Para un proceso  $AR(1)$ , la FAS cae de manera geométrica, de tal forma que:  $\rho_k = \phi^k$ .
    - Función de Autocorrelación Parcial (FAP): Ésta se define como el coeficiente de correlación entre observaciones separadas  $k$  periodos cuando eliminamos de la relación entre ambas la dependencia lineal debida a los valores intermedios. Es análogo al coeficiente que se obtendría en una regresión múltiple. Por tanto, medirá la tasa de cambio en el valor medio

<sup>6</sup> Ahora bien, en algunos estudios empíricos, el objetivo no es la predicción, sino simplemente saber si una serie es o no estacionaria. Este es el caso, por ejemplo, de los estudios empíricos sobre la *hipótesis de los mercados eficientes* de EUGENE FAMA (1970). En este caso, es necesario aplicar contrastos de raíces unitarias como el de Dickey-Fuller.

de la variable regresada ante un cambio unitario en la  $k$ -ésima regresora, manteniendo constante la influencia del resto de variables regresoras:  $y_t = \varphi_1 \cdot y_{t-1} + \varphi_k \cdot y_{t-k} \Rightarrow \rho_k^p = \varphi_k$ .

- La siguiente tabla muestra las propiedades teóricas de la FAS y la FAP en distintos procesos estocásticos:

	FAS	FAP
$AR(p)$	Muchos coeficientes no nulos	Primeros $p$ no nulos, resto 0
$MA(q)$	Primeros $q$ no nulos, resto 0	Muchos coeficientes no nulos
$ARMA(p,q)$	Muchos coeficientes no nulos	Muchos coeficientes no nulos

- Por tanto, a la hora de identificar la especificación de la estructura ARMA:
  - Cuando la serie temporal observada siga un proceso AR o MA puro, en general *debería ser posible identificar fácilmente el orden de estos procesos* a partir de las estimaciones muestrales de la FAS y la FAP.
  - Sin embargo, *en casos más generales*, la estructura compleja de los procesos ARMA hace que, en la práctica, sea *difícil identificar* los órdenes  $p$  y  $q$ .
    - Afortunadamente, no es necesario en la etapa de identificación decidir cuál es el orden del modelo, sino que basta con *seleccionar un conjunto de modelos ARMA que parezcan adecuados* para representar los rasgos principales de la serie. Posteriormente, estimaremos estos modelos y seleccionaremos el más adecuado.
    - En general, se seguirán una serie de *reglas heurísticas* para seleccionar los modelos iniciales. Deben considerarse aspectos como la significatividad de los coeficientes, pautas de decrecimiento (geométricas, sinusoidales...) y las interacciones alrededor de los retardos estacionales para confirmar la concordancia entre las partes regular y estacional<sup>7</sup>.

## 2.2. Estimación

- Llamemos  $\delta$  al vector de parámetros desconocidos del modelo ARMA:

$$\delta' = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \alpha, \sigma^2)$$

¿Cómo obtener una estimación de los  $n$  ( $= p + q + 2$ ) parámetros desconocidos que contiene  $\delta$ ?

- En esta etapa, debe considerarse que, a diferencia de lo que ocurría en el MLG, *deberá abandonarse el supuesto de regresores deterministas* [ver tema 5.B.2], puesto que las variables explicativas serán retardos de la propia variable endógena o del término de error del modelo y, por consiguiente, serán variables aleatorias.
  - Puede demostrarse que, cuando se tengan términos MA, las *estimaciones MCO*, podrían ser inadecuadas. Por esta razón, en general, la literatura enfocada específicamente al análisis de series temporales propone, directamente, la *estimación por máxima verosimilitud* o la *bayesiana* como los métodos más apropiados para la estimación de cualquier modelo  $ARMA(p,q)$ .

### 2.2.1. Estimación por MCO (para procesos AR puros)

- Con procesos AR puros, si se suponen perturbaciones no autocorrelacionadas y estacionariedad del modelo, se cumple siempre la condición de ortogonalidad:

$$E[y_{t-k} \cdot \varepsilon_t] = 0 \quad \forall k > 0$$

Es decir, las variables explicativas (i.e. los retados de  $y$ ), están incorrelacionadas con el error del periodo. Se cumple el supuesto de exogeneidad débil.

- Por tanto, *el estimador MCO sería consistente* en este caso. En este caso, la estimación por máxima verosimilitud condicionada sería muy similar al método de MCO.

<sup>7</sup> En la práctica, la mayoría de las series reales se aproximan bien con modelos  $ARMA$  de ordenes  $p$ ,  $d$  y  $q$  menores que 3 y, en la parte estacional, con órdenes  $P$  y  $Q$  menores que 2.

Además, existen procedimientos automáticos de selección de modelos, como el implantado en el programa TRAMO, que evitan la etapa de identificación y estiman todos los modelos posibles dentro de un subconjunto. El modelo final se obtiene con criterios de selección de modelo (AIC, BIC...).

- Ahora bien, si existen componentes MA, la condición de ortogonalidad dejaría de cumplirse y el estimador MCO ya no sería consistente.

### 2.2.2. Estimación por MV (para procesos MA puros y ARMA)

- En procesos MA puros o modelos ARMA, será preferible la estimación MV condicional y MV exacta.

#### Procedimiento de MV

- El método de MV consiste en encontrar el vector  $\delta$  que maximiza la probabilidad conjunta de la muestra.
  - Hay que proceder en 3 etapas:
    - i. Especificar una distribución para las innovaciones del proceso  $\varepsilon_t$ . Lo más habitual es establecer que  $\varepsilon_t$  es un ruido blanco:  $\varepsilon_t \stackrel{iid}{\sim} N(0, \sigma^2)$ .
    - ii. Calcular la función de verosimilitud.
    - iii. Obtener los valores de  $\delta$  que maximizan la función de verosimilitud. Generalmente, no existe una solución analítica al problema (excepto en el caso en el que la serie sigue un proceso AR puro, en cuyo caso, podemos aplicar MCO). Es necesario aplicar algoritmos numéricos para encontrar una solución al problema.
      - En esencia, estos algoritmos prueban distintos valores de  $\delta$  y evalúan la función de verosimilitud con esos valores. Van ajustando estos valores hasta lograr que la función de verosimilitud converja a un máximo. Este procedimiento puede acelerarse significativamente si el algoritmo comienza en valores próximos a los verdaderos, motivo por el cual, suele comenzarse con los valores obtenidos en una estimación preliminar sencilla (p.ej. MCO).

#### Complejidad

- El principal problema es que, en estos casos, estimar por máxima verosimilitud condicional es más complejo por 2 razones fundamentales:
  - La función de verosimilitud no siempre es lineal en los parámetros, lo cual implica expresiones muy pesadas de calcular.
  - El procedimiento de condicionar a ciertos valores iniciales resulta notoriamente más difícil. En concreto, el cálculo de las esperanzas y varianzas condicionadas será más complejo. Aunque, por ejemplo, para un proceso MA( $q$ ) condicionar bajo la hipótesis de que los primeros  $q$  valores de  $\varepsilon$  son iguales a 0 sería un enfoque simple y bueno.
- En el caso de la estimación MV exacta, se deberá calcular la función de verosimilitud exacta, para lo cual es habitual utilizar el filtro de Kalman, un algoritmo recursivo muy potente y eficiente.

#### Propiedades de los estimadores del método de máxima verosimilitud

- Si el modelo sigue un proceso ARMA estacionario, carente de factores comunes en su parte AR y MA, los estimadores de MV serán:
  - Consistentes.
  - Eficientes para muestras grandes.
  - Se distribuirán asintóticamente como una normal.

### 2.2.3. Método Generalizado de los Momentos (MGM)

- Por último, cabría considerar el método generalizado de los momentos (MGM) ampliamente desarrollado en HANSEN (1982). Puede demostrarse que muchos de los estimadores habituales son casos particulares de este método (incluso, en determinadas ocasiones, el de MV).
  - Las grandes *ventajas* del MGM son:
    - Logra obtener estimaciones consistentes; y
    - Sólo requiere especificar ciertas condiciones de los momentos (condiciones de ortogonalidad) y no toda la función de verosimilitud.



- Ahora bien, en la práctica, este método es *poco utilizado* porque, al no hacer uso de toda la información muestral disponible, *las estimaciones no son eficientes*.

#### 2.2.4. Enfoque bayesiano

- Para la estimación MV, además de suponer que las innovaciones siguen un proceso de ruido blanco distribuyéndose normalmente, asumíamos que no se contaba a priori con información alguna. Precisamente, el *enfoque bayesiano* cobra gran protagonismo dando opciones a que las *estimaciones sean mejoradas para el caso en el que el investigador cuente con información válida a priori*.
  - De este modo, se podrían obtener errores estándar no tan grandes como los que se presume que se tendrían con la estimación MV, lo cual resulta muy atractivo para hacer inferencia estadística.

### 2.3. Verificación

- Si el modelo ha sido correctamente ajustado, entonces los residuos deberían comportarse de manera similar a las verdaderas innovaciones del modelo (que son ruido blanco normalmente distribuido).
  - Por tanto, esta etapa requiere comprobar que:
    - i)  $E[\varepsilon_t] = 0 \quad \forall t = 1, 2, \dots$
    - ii)  $Var[\varepsilon_t] = \sigma^2 \quad \forall t = 1, 2, \dots$
    - iii)  $Cov[\varepsilon_t, \varepsilon_{t-k}] = 0 \quad \forall k = 1, 2, \dots$
    - iv)  $\varepsilon_t \sim N(0, \sigma^2)$
  - Estas propiedades deberán cumplirse no sólo para las distribuciones marginales, sino también para las condicionadas a cualquier conjunto de información de valores pasados de la serie:
    - $E[\varepsilon_t | y_{t-1}, \dots, y_1] = E[\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1] = 0$
    - $Var[\varepsilon_t | y_{t-1}, \dots, y_1] = Var[\varepsilon_t | \varepsilon_{t-1}, \dots, \varepsilon_1] = \sigma^2$

#### 2.3.1. Contraste de media cero (i.e. $E[\varepsilon_t] = 0 \quad \forall t = 1, 2, \dots$ )

- Comencemos por verificar la primera condición (**media cero de los errores**, i.e.  $E[\varepsilon_t] = 0 \quad \forall t = 1, 2, \dots$ ):
  - Para procesos AR puros, estimados por MCO, la hipótesis de media igual a cero de los residuos es una *condición impuesta*.
  - Sin embargo, para modelos estimados por MV exacta, esta restricción no existe y, por tanto, *puede realizarse un contraste sobre la hipótesis nula de que la media de los errores es nula*.
    - Téngase  $\sqrt{T} \cdot \frac{\bar{\varepsilon}}{\hat{\sigma}}$  donde  $\bar{\varepsilon}$  es la media de los residuos obtenidos y  $\hat{\sigma}$  el estimador de la desviación típica.
    - Si tal cociente es significativamente grande con respecto a la distribución  $N(0,1)$  concluiremos que  $E[\varepsilon_t] \neq 0$ .

#### 2.3.2. Contraste de homocedasticidad (i.e. $Var[\varepsilon_t] = \sigma^2 \quad \forall t = 1, 2, \dots$ )

- Pasemos a la segunda condición (**homocedasticidad de los errores**, i.e.  $Var[\varepsilon_t] = \sigma^2 \quad \forall t = 1, 2, \dots$ ):
  - Una primera aproximación para comprobar la estabilidad de la varianza marginal de los residuos es estudiar el *gráfico* de residuos a lo largo del tiempo.
  - Si hay sospechas de un cambio de varianza a partir de un determinado momento del tiempo ( $t = n_1$ ), puede *dividirse el intervalo muestral en 2 partes* para realizar un *contraste de varianzas* a partir del siguiente estadístico, que se distribuye, aproximadamente como una F con  $n_1$  y  $T - n_1$  grados de libertad.  $S_i^2$  es la varianza residual en el tramo  $i$  de  $n_i$  observaciones:

$$\frac{\sum_{t=1}^{n_1} \hat{\varepsilon}_t^2 / n_1}{\sum_{t=n_1+1}^T \hat{\varepsilon}_t^2 / (T - n_1)} = \frac{S_1^2}{S_2^2} \sim F_{n_1, T-n_1}$$

- La idea del contraste es que:

- Si la hipótesis nula de homocedasticidad es cierta, la varianza debería ser similar para ambos periodos. Por tanto, si el cociente  $S_1^2 / S_2^2$  toma un valor de uno aproximadamente, lo más probable es que no exista cambio en la varianza.



- En caso contrario, el cociente tomaría un valor lo suficientemente alejado de la unidad como para poder concluir que dicho cambio de varianza existe y ha sido detectado.

### 2.3.3. Contrastes de ausencia de autocorrelación residual (i.e. $Cov[\varepsilon_t, \varepsilon_{t-k}] = 0 \ \forall k = 1, 2, \dots$ )

- En relación con la tercera condición (**ausencia de autocorrelación residual**, i.e.  $Cov[\varepsilon_t, \varepsilon_{t-k}] = 0 \ \forall k = 1, 2, \dots$ ), se deberá contrastar si los residuos estimados están incorrelados.

#### 1) Análisis de la Función de Autocorrelación y del correlograma

- Una serie estacionaria mostrará que los valores de la FAS y la FAP se encuentran dentro de los intervalos de confianza para diferentes intervalos de tiempo.
  - Por tanto, si la FAS se revela como similar al correlograma de una serie de ruido blanco, podemos decir que dicha serie temporal quizá sea estacionaria.
  - Por el contrario, si analizamos el correlograma de una serie no estacionaria como un paseo aleatorio, la FAS mostrará unos coeficientes de correlación muy altos y decrecientes de manera muy lenta y progresiva. Por otra parte, la FAP muestra un primer coeficiente muy alto y significativo, seguido por el resto de coeficientes con una significatividad nula.

#### 2) Significatividad estadística de los coeficientes de autocorrelación

- El análisis de la FAS y la FAP nos puede arrojar fuertes intuiciones acerca de la significatividad de los coeficientes. Ahora bien, si deseamos realizar un análisis riguroso deberemos hacer un contraste estadístico. Distinguimos **2 contrastes** para la significatividad estadística de los coeficientes de autocorrelación:

- 1) Contraste de la significatividad de un único coeficiente: BARLETT demostró que si una serie es puramente aleatoria (i.e. si es una muestra de ruido blanco) y si la muestra es grande, los coeficientes de autocorrelación parcial son aproximadamente:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k} (\hat{\varepsilon}_t - \bar{\varepsilon})(\hat{\varepsilon}_{t+k} - \bar{\varepsilon})}{\sum_{t=1}^{T-K} (\hat{\varepsilon}_t - \bar{\varepsilon})^2} \sim N(0, 1/n)$$

Es decir, en muestras grandes, los coeficientes de autocorrelación muestrales están normalmente distribuidos y tienen media cero y varianza inversamente proporcional al tamaño de la muestra.

- En consecuencia, podremos calcular con intervalo de confianza los coeficientes y ver si contienen o no el cero.
  - Será preciso examinar la significatividad de tales coeficientes para un determinado número  $h$  de retardos iniciales, estableciendo unos límites de confianza. Un valor próximo a dichos límites debe considerarse un indicio de que el modelo no es el adecuado.
- 2) Contraste de Ljung-Box: Se trata de un contraste global para comprobar que los  $h$  primeros coeficientes son 0 ( $h$  deberá ser suficientemente grande). El contraste se plantea de la siguiente manera:
    - $H_0$ :  $h$  autocorrelaciones = 0
    - $H_1$ :  $h$  autocorrelaciones  $\neq 0$
    - El estadístico empleado en este test es:

$$Q(h) = n \cdot (n + 2) \cdot \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n - k} \sim \chi_{h-n}^2$$

donde  $h$  son el número de coeficientes que forman parte del sumatorio y  $n$  es el número de parámetros estimados para la estructura ARMA regular y estacional.

- En caso de que el estadístico  $Q$  sea mayor al valor crítico marcado por la distribución chi-cuadrado para el nivel de confianza elegido, rechazaremos la hipótesis nula de ausencia de autocorrelación de los residuos y deberemos revisar la identificación del modelo.

### 2.3.4. Contraste de normalidad (i.e. $\varepsilon_t \sim N(0, \sigma^2)$ )

- El cumplimiento de las 3 primeras condiciones garantiza la estacionariedad del proceso y es fundamental para que el modelo sea adecuado. Sin embargo, queda pendiente de comprobar la última condición (**normalidad de los residuos**, i.e.  $\varepsilon_t \sim N(0, \sigma^2)$ ), lo cual será *conveniente*, pues indica que *la correlación implica independencia* y, por tanto, que *no estamos dejando información alguna sin modelizar*:

- Un contraste sencillo y eficaz para comprobar que los residuos se distribuyen normalmente es calcular sus momentos de tercer y cuarto orden con respecto a la media, es decir, los coeficientes de asimetría ( $\alpha_3$ ) y de curtosis ( $\alpha_4$ ), respectivamente:

$$\alpha_3 = \frac{(\hat{\varepsilon}_t - \underline{\varepsilon})^3}{\hat{\sigma}^3} \quad ; \quad \alpha_4 = \frac{(\hat{\varepsilon}_t - \underline{\varepsilon})^4}{\hat{\sigma}^4}$$

- Bajo la hipótesis nula, el siguiente estadístico  $X = [T \cdot \alpha_3^2 / 6] + [T \cdot (\alpha_4 - 3)^2 / 24]$  sigue una distribución  $\chi^2_2$  con 2 grados de libertad.

### 2.3.5. Otros contrastes

#### Prueba de raíz unitaria

- Otra prueba sobre la estacionariedad es la **prueba de la raíz unitaria**.
  - Partimos de un proceso estocástico donde dudamos que siga o no un paseo aleatorio con raíz unitaria:  $y_t = \rho \cdot y_{t-1} + \varepsilon_t$ , con  $\rho \in [-1, 1]$ . Si queremos contrastar si el coeficiente es igual a 1 o no, bajo los supuestos del MLG podríamos estimar  $\Delta y_t = (\rho - 1) \cdot y_{t-1} + \varepsilon_t$  y analizar si el coeficiente del regresor es distinto de 0.
    - Por desgracia, la hipótesis nula de que el coeficiente estimado sea nulo no sigue una distribución  $t$ , ni siquiera en muestras grandes, es decir, no tiene una distribución normal asintótica. ¿Cuál es la alternativa?
  - Dickey-Fuller Test: DICKEY y FULLER probaron que, según la hipótesis nula de que el coeficiente sea cero, el valor estimado  $t$  del coeficiente sigue el estadístico *tau*. Estos autores lo calcularon en base a unas simulaciones de Monte Carlo.
    - Para realizar el *test* es necesario discernir con antelación si el proceso analizado es *a)* un proceso aleatorio, *b)* un paseo aleatorio con deriva; o *c)* un paseo aleatorio con deriva y una tendencia determinista.
    - Esta diferencia resulta fundamental, ya que los valores críticos de la prueba *tau* diferirán en función de las 3 especificaciones señaladas.
  - Prueba Dickey-Fuller aumentada: Este análisis será utilizado para comprobar la existencia de raíz unitaria en condiciones más generales. En particular, este test permitirá que, a parte de la variable del retardo anterior, también entren como variables explicativas retardos anteriores:

$$\Delta y_t = \beta_0 + \beta_1 \cdot t + \rho \cdot y_{t-1} + \sum \alpha_i \cdot y_{t-i} + \varepsilon_t$$

#### Criterios de información (AIC, BIC)

- En su día, el procedimiento en 3 etapas de BOX y JENKINS fue un avance importante, pues la estimación de los parámetros de un modelo ARMA requería mucho tiempo de cálculo, por lo que era conveniente asegurarse de que el modelo a estimar podía ser el adecuado.
- En la actualidad, la estimación de un modelo ARIMA por MV es inmediata, por lo que es más simple estimar todos los modelos que consideramos posibles para explicar la serie y, después, elegir entre todos ellos el que consideremos mejor con criterios automáticos.
  - En concreto, destacan el criterio AIC (de Akaike) y el criterio BIC. Ambos se basan en la misma idea: no es posible utilizar una medida habitual de bondad de ajuste, ya que entonces tenderemos a escoger siempre los modelos con mayor nº de parámetros (que tienen siempre un mayor  $R^2$ ). Por el contrario, con estos criterios *se calculará el valor esperado de la verosimilitud*

para cada uno de los modelos. Es decir, el valor que esperamos obtener para la verosimilitud sobre muchas realizaciones del proceso si en cada una estimamos los parámetros por MV.

- Puede demostrarse que *el criterio AIC es eficiente*: cuando los datos han sido generados por un modelo que puede ser de orden infinito, el predictor seleccionado es el de menor error de predicción esperado.
- Por el contrario, *el criterio BIC es consistente*: cuando los datos han sido generados por un modelo ARIMA, el BIC selecciona el orden adecuado del modelo con probabilidad 1.
- En cualquier caso, las 3 etapas propuestas por BOX y JENKINS son útiles desde un punto pedagógico para comprender las implicaciones de cada modelo.

### 3. PREDICCIÓN

- Una vez estimado adecuadamente el modelo ARIMA a partir de datos de series temporales para una muestra de tamaño  $T$ , el investigador puede utilizar dicho modelo para realizar predicciones, es decir, para estimar los valores de la variable de interés en periodos posteriores a  $T$  (fuera de la muestra).

#### 3.1. Predictor óptimo

##### 3.1.1. Concepto de predictor

- Llamaremos  $\widehat{y}_T(k)$  a un predictor del verdadero valor  $y_{T+k}$  obtenido como función de los  $T$  valores observados. Tenemos que:
  - $T$  es el origen de la predicción en  $T$ .
  - $k$  es el horizonte de la predicción.

##### 3.1.2. Predictor óptimo (criterio de minimización del ECMP)

- ¿Qué criterio se utiliza para seleccionar el mejor predictor? Naturalmente, queremos que la predicción esté lo más cerca posible del verdadero valor que queremos prever. En este, sentido, definimos el error de predicción como la diferencia entre estos dos valores:

$$e_T(k) = y_{T+k} - \widehat{y}_T(k)$$

- En predicción con modelos ARIMA, el criterio más habitual de selección de predictores es el de minimizar el error cuadrático medio de predicción (ECMP).

$$\widehat{y}_T(k) = \operatorname{argmin}(\text{ECMP}) = \operatorname{argmin}(E[e_T(k)^2 | Y_T])$$

- En los años 30, WIENER y KOLMOGOROV demostraron que el predictor óptimo (es decir, el que minimiza el ECMP) es la esperanza de la variable futura condicionada a los datos observados:

$$\widehat{y}_T(k) = E[y_{T+k} | Y_T]$$

#### 3.2. Cálculo de las predicciones

- Supongamos que conocemos todos los parámetros de un proceso  $\text{ARIMA}(p, q, d)$ :

$$\varphi_p(L) \nabla^d y_t = \alpha + \theta_q(L) \cdot \varepsilon_t$$

- Como conocemos los parámetros del modelo, si disponemos de una serie temporal de tamaño  $T$ , entonces podemos obtener también todas las innovaciones  $\varepsilon_1, \dots, \varepsilon_T$ , fijando unos valores iniciales para las primeras innovaciones (supondremos generalmente que son iguales a su esperanza, es decir, nulas). En resumen, suponemos en adelante que conocemos todos los valores  $y_1, \dots, y_T$ , así como todas las innovaciones  $\varepsilon_1, \dots, \varepsilon_T$ .
- Llamamos  $\varphi_h(L) = \varphi_p(L) \nabla^d$  al operador de orden  $h = p + d$  que se obtiene multiplicando el polinomio del operador autorregresivo y el operador de diferencias. Tenemos entonces que el valor de una variable  $y_{T+k}$  generada por el proceso ARIMA puede escribirse como:

$$y_{T+k} = \alpha + \varphi_1 \cdot y_{T+k-1} + \dots + \varphi_h \cdot y_{T+k-h} + \varepsilon_{T+k} - \theta_1 \cdot \varepsilon_{T+k-1} - \dots - \theta_q \cdot \varepsilon_{T+k-q}$$

- El predictor óptimo será la esperanza de la anterior expresión condicionada a la información observada hasta  $T$ :

$$\widehat{y}_T(k) = \alpha + \varphi_1 \cdot \widehat{y}_T(k-1) + \dots + \varphi_h \cdot \widehat{y}_T(k-h) + E[\varepsilon_{T+k} | Y_T] - \dots - \theta_q \cdot E[\varepsilon_{T+k-q} | Y_T]$$

- Por tanto, habrá elementos que conocemos en  $T$  y otros que todavía no son conocidos. En concreto:
  - $E[\varepsilon_{T+k-i}|Y_T]$  será un valor conocido en  $T$  si  $i > k$ .
  - $\hat{y}_T(k-i) = E[y_{T+k-i}|Y_T]$  será un valor conocido en  $T$  si  $i > k$ .
- Cuando estos valores no sean conocidos:
  - La esperanza de las innovaciones futuras condicionada a la historia de la serie es igual a su esperanza absoluta, es decir, 0 (ya que  $\varepsilon$  sigue un proceso de Márkov).
  - La esperanza de las de las predicciones  $y_{T+k-i}$  puede ser calculada de manera recursiva.
- De esta manera, podemos obtener la función de predicción.
- Puede demostrarse que la función de predicción de un modelo ARIMA tiene las siguientes características:
  - Los *operadores estacionarios* (AR y MA) determinan la *predicción a corto plazo*.
  - Los *operadores no estacionarios* (las diferencias y la constante) determinan la *predicción a largo plazo*.

### 3.3. Varianza de la predicción e intervalos de predicción

- Una vez obtenida una estimación puntual del predictor óptimo, es necesario construir un intervalo de predicción que nos permita conocer en qué rango puede llegar a situarse finalmente la variable futura con una determinada probabilidad.
  - Para ello, primero es necesario calcular la varianza de la predicción, que puede demostrarse que es igual a:
 
$$\text{Var}(e_T(k)) = \sigma^2 \cdot (1 + \psi_1^2 + \dots + \psi_{k-1}^2)$$
 donde los coeficientes  $\psi_i$  son los de un proceso ARIMA en su representación como un proceso MA( $\infty$ ) por la descomposición de Wold, de forma que:
 
$$y_t = \psi(L) \cdot \varepsilon_t$$
  - Puede deducirse que la incertidumbre de predicción difiere mucho según el proceso sea o no estacionario:
    - Si es estacionario, la varianza de la predicción a largo plazo converge a un valor constante, por lo que la incertidumbre está acotada (nótese que la predicción a largo plazo será la media del proceso).
    - Si no es estacionario, la serie del paréntesis no es convergente y la incertidumbre de la predicción a largo plazo crece sin límite hasta infinito. A largo plazo, no se puede prever el comportamiento de un proceso no estacionario.
  - Asumiendo normalidad para las innovaciones, pueden calcularse los intervalos de predicción:

$$IP = \hat{y}_T(k) \pm \lambda_{\alpha/2} \cdot \sqrt{\text{Var}(e_T(k))}$$

siendo  $\lambda_{\alpha/2}$  los percentiles de la distribución normal estándar. Para un nivel de confianza  $(1 - \alpha) = 95\%$ , utilizaríamos  $\lambda_{\alpha/2} = 1,96$ .

## CONCLUSIÓN

- Los modelos ARIMA son la base para estudio de los procesos estocásticos con datos de series temporales univariantes. A lo largo de la exposición, hemos visto cómo, a partir del procedimiento iterativo en 3 etapas de BOX y JENKINS, podemos ajustar un modelo ARIMA a partir de datos de series temporales.
- Dicha estimación puede servirnos para realizar predicciones, lo cual puede ser extremadamente útil en multitud de aplicaciones económicas, como el diseño de la política económica, las finanzas o la planificación de inversiones, entre otros.

## Bibliografía

### Anexos

#### A.1. Anexo 1: Procesos estocásticos puros (AR, MA)

Los modelos ARIMA son combinaciones de procesos AR y procesos MA, que además son integrados.

##### Procesos AR

Los procesos autorregresivos (AR) son los primeros procesos estacionarios que se estudiaron.

En el caso más sencillo, decimos que una serie temporal sigue un proceso autorregresivo de primer orden o un **AR(1)** si ha sido generada por:

$$y_t = \alpha + \Phi_1 y_{t-1} + \varepsilon_t \quad \text{para } t = 1 \dots T$$

Es decir, cada valor depende de una constante y de su propio valor retardado<sup>8</sup>.

Generalizando para p retardos, obtenemos un **proceso AR(p)** cuando la serie ha sido generada por:

$$y_t = \alpha + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \dots + \Phi_p y_{t-p} + \varepsilon_t$$

Utilizando la notación del operador de retardos:

$$\Phi_p(L) \cdot y_t = \alpha + \varepsilon_t$$

donde,

- $\Phi_p(L)$  es el polinomio de operadores de retardos de grado p, de manera que:

$$\Phi_p(L) = 1 - \Phi_1 L - \Phi_2 L^2 \dots - \Phi_p L^p$$

- L es el operador de retardo, de manera que:

$$L^k \cdot y_t = y_{t-k}$$

Llamaremos la ecuación característica a  $\Phi_p(L) = 0$ . Esta ecuación tendrá p raíces  $G_i^{-1}$ .

Condición de estacionariedad: el proceso es estacionario si  $|G_i| < 1$  para todo  $i = 1 \dots p$ . O lo que es lo mismo, las raíces del polinomio deben quedar fuera del círculo unidad.

##### Procesos MA

En el caso más sencillo, decimos que una serie temporal sigue un proceso autorregresivo de primer orden o un **MA(1)** si ha sido generada por:

$$y_t = \alpha - \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad \text{para } t = 1 \dots T$$

Generalizando para q retardos, tenemos un proceso **MA(q)**:

$$y_t = \alpha - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

Utilizando la notación del operador de retardos:

$$y_t = \alpha + \theta_q(L) \cdot \varepsilon_t$$

donde,

- $\theta_q(L)$  es el polinomio de operadores de retardos de grado p, de manera que:

$$\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 \dots - \theta_q L^q$$

Todos los procesos MA puros son siempre estacionarios, ya que son combinación lineal de las innovaciones (que a su vez son procesos estacionarios).

Sin embargo, pueden ser o no invertibles en función de las raíces unitarias del polinomio de retardos. La condición de invertibilidad es que las raíces del polinomio MA queden fuera del círculo unidad.

#### A.2. Anexo 2: Funciones FAC, FAS y FAP

Unos estadísticos fundamentales en la especificación de modelos univariantes son las funciones de autocovarianza, de autocorrelación simple y de autocorrelación parcial del proceso estocástico. Pasamos a definirlos.

**Función de autocovarianza (fac)**: es una función que describe las covarianzas entre dos variables del proceso en dos instantes cualesquiera. Es decir:

$$\gamma(t, t+k) = \text{Cov}(y_t, y_{t+k})$$

Uno de los inconvenientes de la fac es que depende de las dimensiones de la variable, por lo que es más conveniente la siguiente función, que normaliza dividiendo por las desviaciones típicas de las variables.

<sup>8</sup> Mediante iteración, podemos demostrar que el proceso AR(1) puede ser representado como suma de innovaciones, es decir, con forma de media móvil de orden infinito.

**Función de autocorrelación simple (fas)**: es una función que describe las correlaciones entre dos variables del proceso en dos instantes cualesquiera. Es decir:

$$\rho(t, t+k) = \frac{\text{Cov}(y_t, y_{t+k})}{\sqrt{\text{Var}(y_t)}\sqrt{\text{Var}(y_{t+k})}}$$

**Función de autocorrelación parcial (fap)**: es una función que describe las correlaciones entre dos variables del proceso en dos instantes cualesquiera ajustada por el efecto de los retardos intermedios.

El gran interés de un proceso estocástico estacionario reside en que las fac, fas y fap son independientes del tiempo t, por lo que puede omitirse dicho argumento temporal. Lo que es crucial es que dicha invarianza permite la estimación muestral de tales funciones, del modo que analizamos a continuación.

### A.3. Anexo 3: Test Durbin-Watson (contraste de autocorrelación)

Es el test más habitual cuando se sospecha que el término de error del modelo sigue un esquema autorregresivo de primer orden AR(1).

El contraste puede exponerse de tal manera que:

$H_0$ : ausencia de autocorrelación residual

$H_1$ : autocorrelación residual de primer orden  $\tilde{\rho} \neq 0$

El estadístico se define de la siguiente manera:

$$d = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=2}^T \hat{\varepsilon}_t^2}$$

Para muestras suficientemente grandes, el estadístico puede expresarse como:

$$d \approx 2 \cdot (1 - \tilde{\rho})$$

donde  $\tilde{\rho} = \frac{\sum_{t=2}^T (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=2}^T \hat{\varepsilon}_t^2}$  es el grado de correlación entre residuos consecutivos.

La idea del contraste es la siguiente:

- El coeficiente de correlación ( $\tilde{\rho}$ ) puede tomar valores comprendidos entre -1 y 1 (autocorrelación negativa o positiva), por lo que el estadístico  $d$  los tomará entre 0 y 4.
- El estadístico toma el valor  $d = 2$  cuando  $\tilde{\rho} = 0$  (correlación nula). En este caso, no se rechaza la hipótesis nula de ausencia de correlación residual de primer orden.
- A medida que  $d$  se aleje de 2, se tenderá a rechazar más fácilmente la hipótesis nula. Por ejemplo, en los casos extremos:
  - $d = 0$  cuando  $\tilde{\rho} = 1$ . Hay evidencia de autocorrelación residual positiva.
  - $d = 4$  cuando  $\tilde{\rho} = -1$ . Hay evidencia de autocorrelación residual negativa.

En resumen, se rechazará la hipótesis nula de ausencia de correlación serial cuando el estadístico  $d$  tome valores extremos (cerca de 0 o de 4).

A pesar de su popularidad, este contraste presenta ciertas limitaciones prácticas:

- Para determinados valores de  $d$ , el investigador no podrá tomar una decisión basada en este contraste<sup>9</sup>.
- Las variables explicativas se suponen deterministas (poco realista).
- No permite detectar esquemas de autocorrelación generados por procesos autorregresivos de orden mayor que 1.
- Es inaplicable a modelos dinámicos que incluyan retardos de la variable endógena como variable explicativa.

En vista de tales desventajas, se han planteado contrastes alternativos como el siguiente.

<sup>9</sup> La distribución de probabilidad de este estadístico es difícil de derivar y depende de forma compleja de  $X$ . No habrá, a diferencia de la mayoría de los contrastes, un único valor crítico para este test. DURBIN y WATSON hallaron un límite inferior ( $d_L$ ) y otro superior ( $d_U$ ), tales que, si  $d$  caía entre dichos límites, el contraste no determina si existe o no autocorrelación de primer orden (el contraste no concluye nada). Estos límites dependen del  $n^\circ$  de observaciones de la muestra y del  $n^\circ$  de variables explicativas.



## 5.B.7 : EL MODELO DE REGRESIÓN DINÁMICA UNIECUACIONAL: IDENTIFICACIÓN, ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

### ▪ **Contextualización:**

- En Econometría aplicada, se trabaja con 3 tipos de datos muestrales:
  - Datos de sección cruzada: Conjunto de datos recogidos observando diversas unidades económicas (como familias, empresas, ciudades...) en un mismo instante de tiempo o haciendo abstracción de la dimensión temporal.
  - Datos de series temporales: Conjunto ordenado de datos sobre una unidad económica a lo largo del tiempo.
  - Datos de panel (combinación de los 2 anteriores): Datos de sección cruzada observados a lo largo de varios periodos de tiempo.
- Esta exposición se centra en el análisis de **series temporales**. Podemos distinguir 2 grandes **tipos de modelos dinámicos uniecuacionales**:
  - Modelos univariantes: La evolución de una variable se explica utilizando sólo su propio pasado, a través de modelos ARIMA, donde la evolución de una variable se explica utilizando su propio pasado.
  - Modelos multivariantes: Modelos con variables explicativas.
- En esta exposición, nos vamos a centrar en los **modelos multivariantes**. En concreto, estudiaremos un grupo modelos que se conocen en las ciencias físicas como *modelos de función de transferencia* y en las ciencias económicas como *modelos de regresión dinámica*.
  - Estos modelos permiten considerar varias series temporales simultáneamente, lo cual es interesante por 2 motivos:
    - Para responder a *preguntas causales*. Por ejemplo, ¿cuál es la senda esperada de consumo de gasolina si los precios del petróleo aumentan un 10 %?
    - Para *mejorar nuestras predicciones al introducir más información que el propio pasado de la variable dependiente*. Por ejemplo, los shocks de los precios del petróleo pueden ser útiles para explicar la serie de consumo de gasolina.
  - Existen varios modelos económicos que se pueden representar como modelos de regresión dinámicos. Ejemplos:
    - Modelos de expectativas adaptativas (HEA).
    - Modelos de ajuste parcial (p.ej. teoría del acelerador flexible).
    - Modelos de optimización dinámica.

### ▪ **Problemática (Preguntas clave):**

- Para demostrar la consistencia de MCO es fundamental que las variables regresadas sean **estacionarias**. Por ello, primeramente, definiremos *qué son las series estacionarias*.
- Posteriormente, nos centraremos en **modelos de regresión entre variables estacionarias**. Diferenciaremos entre los casos en que:
  - *En las variables explicativas no hay valores retardados de la endógena*.
  - *Existen retardos de la endógena como variables explicativas*.
- Finalmente, veremos también brevemente cuáles son las **consecuencias de regresar variables no estacionarias** (*regresiones espurias y cointegración*).

▪ **Estructura:**

- 1. CUESTIONES PREVIAS: SERIES TEMPORALES ESTACIONARIAS**

1.1. Definición de procesos estocásticos y series temporales

1.2. Estacionariedad (estricta y débil)

1.3. Ejemplos de procesos estocásticos (Ruido blanco, Paseo aleatorio, AR y MA)

**2. MODELOS ARDL**

2.1. Idea

2.1.1. Formulación general

2.1.2. Casos particulares

2.1.3. Medidas descriptivas

2.1.4. Representación  $DL(\infty)$

2.2. Modelos de retardos distribuidos (sin retardo de la endógena) ( $DL(r)$ )

2.2.1. Identificación

Forma funcional de los modelos  $DL(r)$

Razones por las que un modelo econométrico podría incluir retardos de las variables explicativas

2.2.2. Estimación

Dificultades de MCO (orden infinito y multicolinealidad)

Solución: modelo truncado de Koyck (retardo geométrico)

2.3. Modelos con retardo de la endógena ( $ARDL(p,r)$ )

2.3.1. Identificación

2.3.2. Estimación

Método de variables instrumentales

Mínimos Cuadrados en 2 etapas (MC2E) para situaciones con sobreidentificación

2.3.3. Verificación (test de HAUSMAN y WU)

2.3.4. Predicción

2.3.5. Aplicación práctica: expectativas racionales

**3. MODELOS DINÁMICOS CON VARIABLES NO ESTACIONARIAS**

3.1. Relaciones espurias (GRANGER y NEWBOLD, 1974)

3.2. Cointegración

## 1. CUESTIONES PREVIAS: SERIES TEMPORALES ESTACIONARIAS

Este apartado brevemente.

### 1.1. Definición de procesos estocásticos y series temporales

- Llamamos **proceso estocástico** a un conjunto de variables aleatorias  $\{y_t\}$ . En nuestro caso, este conjunto es ordenado y corresponde a los instantes temporales (días, meses, años, etc.). Para cada valor  $t$ , está definida una variable aleatoria (i.e.  $y_1, y_2, \dots$ ). Los valores observados de las variables aleatorias en estos distintos instantes conforman una *serie temporal*.
  - En el análisis estadístico, se suele interpretar la serie temporal como una realización del proceso estocástico, es decir, una muestra. En este punto, es importante distinguir 2 casos:
    - En *situaciones experimentales*, es posible obtener varias series temporales (varias muestras) del mismo proceso estocástico.
      - Por ejemplo, podemos estudiar cómo evoluciona el tamaño de una planta en los meses 1, 2, ..., 12 desde que se planta la semilla ( $t = 0$ ) y volver a repetir el experimento en las mismas condiciones.
    - Sin embargo, en la mayoría de *aplicaciones económicas*, sólo podemos observar una realización del proceso.
      - Por ejemplo, la serie de crecimiento anual del PIB es una realización concreta de un proceso estocástico, pero no podemos volver atrás en el tiempo para observar otra realización.
        - En este último caso, el objetivo será estimar las características “transversales” del proceso (medias, varianzas, etc.) a partir de su evolución “longitudinal”. Para que ello sea posible, es necesario suponer que las propiedades transversales (distribución de cada variable en cada instante) son estables a lo largo del tiempo. Aquí entra el concepto de **estacionariedad**.

### 1.2. Estacionariedad (estricta y débil)

- Podemos distinguir 2 tipos de **estacionariedad**:
  - Estacionariedad en sentido estricto: Un proceso estocástico  $\{y_t\}$  es estacionario en sentido estricto si para toda  $m$ -tupla  $(t_1, t_2, \dots, t_m)$  y todo entero  $k$ , el vector de variables  $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$  tiene la misma distribución de probabilidad conjunta que el vector  $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$ .
    - En otras palabras, la función de distribución conjunta de cualquier conjunto de variables no se modifica si trasladamos las variables en el tiempo  $k$  periodos. Esto implica que todos los momentos que caracterizan a los 2 conjuntos son iguales (esperanza, varianza/covarianza, asimetría, curtosis...).
    - La estacionariedad estricta es una condición muy fuerte y excesiva para nuestras necesidades prácticas. Por ello, generalmente, nos conformamos con un concepto menos exigente.
  - Estacionariedad en sentido débil (o de 2º orden): Se produce cuando todos los momentos de 1ª y 2ª orden del proceso estocástico son invariantes en el tiempo. Estos momentos incluyen la esperanza matemática y la varianza de las variables  $y_t$ , pero también las covarianzas y correlaciones entre diversos retardos<sup>3</sup>.
    - Algunas de estas condiciones son fácilmente contrastables. Por ejemplo, existirán indicios claros de no estacionariedad cuando una serie temporal muestre:
      - Una tendencia creciente (*esperanza no constante*).
      - Fluctuaciones de amplitud creciente en el tiempo (*varianza no constante*).

<sup>3</sup> Nótese que la estacionariedad estricta implica la estacionariedad débil, pero, en general, la dirección de esta implicación no funciona al revés. Existe, sin embargo, un caso especial: si la distribución conjunta de las variables aleatorias sigue una distribución normal multivariante (que se caracteriza sólo por sus dos primeros momentos), entonces la estacionariedad débil implica también estacionariedad estricta.

### 1.3. Ejemplos de procesos estocásticos (Ruido blanco, Paseo aleatorio, AR y MA)

▪ Podemos mencionar los siguientes **ejemplos de procesos estocásticos**:

- Ruido blanco: Se dice que un proceso es puramente aleatorio si tiene media cero, una varianza constante y no está serialmente correlacionado. Por tanto, en estos procesos conocer los valores pasados no nos proporciona información alguna sobre el futuro, ya que el proceso no tiene memoria.
- Paseo aleatorio: El paseo aleatorio constituye el ejemplo de proceso no estacionario más conocido.
  - Paseo aleatorio sin deriva: Es un proceso donde el valor presente es igual al valor pasado más un proceso de ruido blanco.
    - En relación con su forma:
      - El proceso se define como  $y_t = y_{t-1} + \varepsilon_t$ , y puede ser descrito como  $y_T = y_0 + \sum_{t=0}^T \varepsilon_t$ .
      - Si tomamos esperanzas:  $E[y_t] = y_0$ .
      - Su varianza es:  $\text{var}[y_t] = t \cdot \sigma^2$ .
    - En relación con sus características:
      - Una de las características más importantes de estos procesos es la *persistencia de los shocks aleatorios*. Estos procesos tienen *memoria infinita*, lo que implica que el efecto de un shock no desaparece a lo largo del tiempo. A la suma de los términos error se le conoce como la tendencia estocástica.
      - Debido a que estas series no son estocásticas, deberemos realizar un tratamiento previo de los datos antes de proceder a su estimación.
    - Este proceso ha sido utilizado en la teoría económica para defender tesis como la *hipótesis de los mercados eficientes*: al analizar la evolución de las cotizaciones de los títulos de renta variable se constata que éstas siguen un paseo aleatorio. Autores como FAMA han argumentado que ello supone un ejemplo de la eficiencia de los mercados: las innovaciones se derivan de información que previamente no se encontraba disponible.
  - Paseo aleatorio con deriva: Estos procesos, además de tener una tendencia estocástica, tienen una tendencia determinista.
    - Por tanto, el proceso adopta la siguiente forma:
      - El proceso puede ser definido como:  $y_t = \delta + y_{t-1} + \varepsilon_t$ .
      - En estos procesos, la esperanza del mismo es igual a:  $E[y_t] = y_0 + t \cdot \delta$ .
    - De nuevo, para analizar estos datos, será necesario eliminar el componente tendencial de la serie.
- AR(1) (generalizable a AR(p)): Diremos que una serie  $z_t$  sigue un proceso autorregresivo de primer orden si ha sido generada por:  $y_t = c + \varphi \cdot y_{t-1} + \varepsilon_t$ , donde  $-1 < \varphi < 1$  y el componente  $\varepsilon_t$  es un proceso de ruido blanco.
  - Estos procesos presentarán una media igual a:
 
$$E[y_t] = E[y_{t-1}] \Rightarrow \mu = c + \varphi \cdot \mu \Rightarrow \mu = c/(1 - \varphi)$$
  - Para un proceso AR(1) con media cero ( $c = 0$ ), su varianza será igual a:
 
$$E[y_t^2] = \varphi^2 \cdot E[y_t^2] + 2 \cdot \varphi \cdot E[y_{t-1} \cdot \varepsilon_t] + E[\varepsilon_t^2] \Rightarrow \sigma_y^2 = \varphi^2 \cdot \sigma_y^2 + \sigma^2 \Rightarrow \sigma_y^2 = \sigma^2/(1 - \varphi^2)$$
- MA(1) (generalizable a MA(q)): Se define el proceso de media móvil de orden 1 añadiendo a un proceso de ruido blanco una dependencia del valor actual de la serie de la última innovación ocurrida. De esta manera, el proceso será una combinación lineal de las 2 últimas innovaciones, de acuerdo con la ecuación:  $y_t = \varepsilon_t + \theta \cdot \varepsilon_{t-1}$ .
  - Este proceso es la suma de 2 procesos estacionarios, por lo que el resultado siempre será estacionario.

- Si suponemos que  $-1 < \theta < 1$ , entonces diremos que el proceso es invertible y que tiene la propiedad de que el efecto de los valores pasados decrece con el tiempo.

- La invertibilidad puede demostrarse sustituyendo  $\varepsilon_{t-1}$  por  $y_{t-1}$ :

$$\left. \begin{aligned} y_t &= \varepsilon_t + \theta \cdot \varepsilon_{t-1} \\ y_{t-1} &= \varepsilon_{t-1} + \theta \cdot \varepsilon_{t-2} \Rightarrow \varepsilon_{t-1} = y_{t-1} - \theta \cdot \varepsilon_{t-2} \end{aligned} \right\} \Rightarrow y_t = \varepsilon_t + \theta \cdot (y_{t-1} - \theta \cdot \varepsilon_{t-2})$$

$$y_t = \varepsilon_t + \theta \cdot (y_{t-1} - \theta \cdot (y_{t-2} - \theta \cdot \varepsilon_{t-3}))$$

$$y_t = \varepsilon_t + \theta \cdot y_{t-1} - \theta^2 \cdot y_{t-2} + \theta^3 \cdot y_{t-3} - \dots$$

- Si continuamos la inducción hacia atrás:

$$y_t = \varepsilon_t - \sum_{i=1}^{t-1} (-\theta)^i \cdot y_{t-i}$$

- De esta manera observamos 2 cosas:
  - Que hemos conseguido invertir un proceso MA y representarlo en forma de proceso AR.
  - Que, a medida que retrocedemos en el tiempo, el efecto de las innovaciones pasadas en el proceso de ruido blanco sobre la variable  $y_t$  se reduce hasta converger a 0.

## 2. MODELOS ARDL

### 2.1. Idea

#### 2.1.1. Formulación general

- La formulación más general de los modelos de regresión dinámicos uniecuacionales es la de los **modelos ARDL (autoregressive distributed lag)**, que permiten la inclusión de retardos de todas las variables.

- Suponiendo por sencillez que no existe constante, un modelo ARDL(p,r), tiene la siguiente forma:

$$y_t = \alpha_1 \cdot y_{t-1} + \dots + \alpha_p \cdot y_{t-p} + \beta_0 \cdot x_t + \beta_1 \cdot x_{t-1} + \dots + \beta_r \cdot x_{t-r} + \varepsilon_t$$

$$y_t = \sum_{i=1}^p \alpha_i \cdot y_{t-i} + \sum_{j=0}^r \beta_j \cdot x_{t-j} + \varepsilon_t$$

- Utilizando la notación compacta del operador de retardos, podemos re-expresar la anterior ecuación del siguiente modo:

$$A_p(L) \cdot y_t = B_r(L) \cdot x_t + \varepsilon_t$$

donde:

- $A_p(L) = 1 - \alpha_1 \cdot L - \alpha_2 \cdot L^2 - \dots - \alpha_p \cdot L^p$
- $B_r(L) = \beta_0 - \beta_1 \cdot L - \beta_2 \cdot L^2 - \dots - \beta_r \cdot L^r$

#### 2.1.2. Casos particulares

- Dentro de la formulación general, podemos identificar los siguientes **casos particulares**:

- Si  $\beta_0 = 0$  y  $r = 0$ , se reduce a un modelo autorregresivo AR(p):

$$y_t = \sum_{i=1}^p \alpha_i \cdot y_{t-i} + \varepsilon_t$$

- Este tipo de modelos univariantes son estudiados en otra parte del temario dedicada a modelos univariantes tipo ARIMA. Se pueden estimar por el procedimiento iterativo en 3 etapas planteado por BOX y JENKINS (1970): identificación, estimación, verificación [ver tema 5.B.6].

- Si  $p = 0$ , tenemos un modelo de retardos distribuidos, es decir, un modelo DL(r):

$$y_t = \sum_{j=0}^r \beta_j \cdot x_{t-j} + \varepsilon_t$$



- Si  $r = 1$  y  $p = 1$ , obtenemos un modelo ARDL(1,1), que es frecuentemente utilizado en la literatura de series temporales:

$$y_t = c + \alpha_1 \cdot y_{t-1} + \beta_0 \cdot x_t + \beta_1 \cdot x_{t-1} + \varepsilon_t$$

- Este modelo tiene una importante representación alternativa conocida como *modelo de corrección del error*.

### 2.1.3. Medidas descriptivas

- El modelo ARDL es interesante para medir los efectos dinámicos de una variable  $x$  sobre otra variable  $y$ .
  - Por ejemplo, para medir los efectos del gasto público sobre el nivel de actividad a lo largo del tiempo. En este sentido, debemos introducir una serie de conceptos previos que permiten describir las propiedades de un modelo de retardos distribuidos:

#### ○ Multiplicadores:

- *Multiplicador de impacto (contemporáneo)*. Cuantifica el efecto de una variación unitaria de la variable exógena en el periodo actual sobre la variable endógena en el mismo periodo:

$$m_0 = \frac{\partial y_t}{\partial x_t} = \delta_0 = \beta_0$$

- *Multiplicador de retardo  $j^{\text{º}}$* . Cuantifica el efecto de una variación unitaria de la variable exógena en el periodo  $t - j$ , sobre la variable endógena en el periodo  $t$ :

$$m_j = \frac{\partial y_t}{\partial x_{t-j}} = \delta_j \neq \beta_j$$

- *Multiplicador total<sup>5</sup> (de largo plazo)*. Es la suma de todos los multiplicadores:

$$m_T = \sum_{j=0}^{+\infty} m_j$$

#### ○ Retardos:

- *Retardo medio*. Se define como la media ponderada por el retardo de todos los coeficientes del polinomio DL:

$$q_{\text{medio}} = \frac{\sum_{j=1}^{\infty} j \cdot \delta_j}{\sum_{j=1}^{\infty} \delta_j}$$

Mide el tiempo medio que un cambio en  $x$  tarda en producir modificaciones sobre  $y$ . La idea es informar sobre si el impacto de la variación de la variable exógena sobre la variable endógena es concentrado o diluido en el tiempo.

- *Retardo mediano*. Es el instante en el que se alcanza el 50 % del impacto total que se produce en  $y_t$  debido a una variación en  $x_t$ :

$$q_{\text{mediano}} = \min \left\{ q \left| \frac{\sum_{j=1}^q m_j}{\sum_{j=1}^{\infty} m_j} \geq 0,5 \right. \right\}$$

Permite responder a una pregunta como la siguiente: si aumentamos temporalmente la oferta monetaria hoy, ¿en qué momento se habrá producido la mitad de la subida de precios asociada a dicho estímulo monetario?

- Funciones de respuesta. Finalmente, partiendo de los anteriores multiplicadores, podemos construir 2 funciones:

- Una *función de respuesta al impulso* muestra la sucesión de los efectos sobre  $y_t, y_{t+1} \dots$  ante una variación unitaria de la variable explicativa  $x$  en el periodo  $t$ , que se revierte en periodos siguientes. Sería, por tanto, la sucesión de multiplicadores de retardo  $j$ , desde  $j = 0: \delta_0, \delta_1, \delta_2$

<sup>4</sup> En este caso el multiplicador no coincidirá con el estimador, ya que existe una dependencia implícita de las variables dependientes retardadas.

<sup>5</sup> Para que el modelo tenga sentido económico, el multiplicador total debe ser finito. Esto ocurrirá cuando el proceso sea estable.

- Una *función de respuesta al escalón* muestra la sucesión de estos efectos, pero en el caso de que la variación unitaria de la variable  $x_t$  sea permanente. Sería, por tanto, la sucesión de multiplicadores totales hasta cada periodo:  $\delta_0, (\delta_0 + \delta_1), (\delta_0 + \delta_1 + \delta_2) \dots$ 
  - Para que un modelo sea **estable**, se debe cumplir que:
    - La función de respuesta al impulso tiende a 0 (ante una variación puntual en  $x$ , la variable  $y$  retorna a su valor de equilibrio).
    - La función de respuesta al escalón se estabiliza (ante una variación permanente en  $x$ , la variable  $y$  evoluciona hacia su nuevo valor de equilibrio).
  - Para que esto se dé, es necesario que todas las raíces del término AR estén fuera del círculo unidad (condición de estacionariedad).

#### 2.1.4. Representación $DL(\infty)$

- Nótese que los modelos  $ARDL(p,r)$  pueden ser transformados en un modelo de retardos distribuidos  $DL(\infty)$ <sup>6</sup>:

$$\begin{aligned} A_p(L) \cdot y_t &= B_r(L) \cdot x_t + \varepsilon_t \\ y_t &= \frac{B_r(L)}{A_p(L)} \cdot x_t + \frac{1}{A_p(L)} \cdot \varepsilon_t \\ y_t &= D_\infty(L) \cdot x_t + u_t \\ y_t &= \delta_0 \cdot x_t + \delta_1 \cdot x_{t-1} + \delta_2 \cdot x_{t-2} + \dots + u_t \end{aligned}$$

donde:

- $u_t$  ya no es ruido blanco (ahora está autocorrelacionado).
- El cociente  $\frac{B_r(L)}{A_p(L)}$  se conoce como ganancia.
- Si intentamos la *estimación de estos modelos con el método de MCO*, vamos a encontrar una serie de problemas al incumplirse alguno de los supuestos “ideales” para que las estimaciones obtenidas sean “buenas” (*insesgadez, mínima varianza, consistencia...*).
  - Por ello, será necesario analizar la problemática que se presenta en cada caso para **aplicar el método de estimación más adecuado**.
  - Vamos a comenzar por el caso más sencillo e iremos añadiendo complejidad.

## 2.2. Modelos de retardos distribuidos (sin retardo de la endógena) ( $DL(r)$ )

### 2.2.1. Identificación

#### Forma funcional de los modelos $DL(r)$

- Una primera aproximación para mejorar del modelo de regresión clásico consiste en considerar retardos en las variables explicativas:

$$\begin{aligned} y_t &= \alpha + \beta_0 \cdot x_t + \beta_1 \cdot x_{t-1} + \dots + \beta_r \cdot x_{t-r} + \varepsilon_t \\ y_t &= \alpha + \sum_{i=0}^r \beta_i \cdot x_{t-i} + \varepsilon_t \end{aligned}$$

- Esta especificación es conocida como modelo de retardos distribuidos (*distributed lag*) de orden  $r$ , que se representa como  $DL(r)$ .
- Suponemos que:
  - 1) Existe una única variable explicativa, pero su efecto está distribuido en varios retardos (el modelo se podría generalizar a más variables).
  - 2) El término de error  $\varepsilon_t$  cumple una serie de propiedades:
    - a)  $\varepsilon_t$  es independiente de  $X$ .
    - b)  $E[\varepsilon_t] = 0$ .
    - c)  $Var[\varepsilon_t] = \sigma^2$  para todo  $t$  (homocedasticidad o varianza constante del error).

<sup>6</sup> Los **modelos de retardos distribuidos infinitos** ( $DL(\infty)$ ) relacionan la variable endógena con valores actuales y pasados de las variables explicativas, donde las variables explicativas retardadas se extienden a un pasado **infinito**.

### Razones por las que un modelo econométrico podría incluir retardos de las variables explicativas

- Existen múltiples razones por las que un modelo econométrico podría incluir retardos de las variables explicativas:
  1. Razones psicológicas: Pueden existir retardos como resultado de la fuerza del hábito (inercia).
    - Por ejemplo, la gente no cambia sus hábitos de consumo de forma inmediata tras un cambio en el contexto económico agregado, sino que van realizando un ajuste progresivo hasta el nuevo valor de equilibrio. Ello puede deberse a unas expectativas adaptativas, a una racionalidad limitada o a incertidumbre respecto a la naturaleza del cambio (transitorio o permanente).
  2. Razones tecnológicas: En muchas ocasiones, existen limitaciones físicas que impiden que la actividad económica se ajuste inmediatamente a las nuevas condiciones de mercado.
    - Por ejemplo, en el contexto de la inversión, ante una caída relativa del coste del capital, la curva de demanda de capital de la empresa será más flexible a largo plazo que a corto plazo. Ello puede deberse a la existencia de costes de ajuste en la función de inversión, tal y como modelizan EISNER y STROTZ. Por tanto, el ajuste del capital en el momento presente puede venir explicado por cambios en los tipos de interés que se produjeron en periodos anteriores.
  3. Razones institucionales: Las obligaciones contractuales también pueden impedir que las empresas cambien de una fuente de trabajo o de unas materias primas a otras. Otro ejemplo es la consideración de la política fiscal ante la coyuntura económica: debido al proceso legislativo subyacente a la aprobación de los presupuestos, la política fiscal aplicada actualmente puede ser la respuesta a los shocks económicos experimentados hace varios periodos.
- Por todas estas razones el retardo desempeña un papel central en la economía. Esto se refleja en la metodología económica del corto plazo y el largo plazo. Por esta razón se dice que las elasticidades precio-ingreso de corto plazo suelen ser menores que las elasticidades correspondientes a largo plazo, o que la propensión marginal al consumo en el corto plazo es menor a la del largo plazo.

#### 2.2.2. Estimación

### Dificultades de MCO (orden infinito y multicolinealidad)<sup>7</sup>

- Dado el modelo  $DL(r)$ , una de las primeras opciones en su estimación puede ser utilizar MCO e introducir retardos de manera ad hoc o específica para la base de datos con la que se esté trabajando.
  - No obstante, esto puede generar un conjunto de **problemas**:
    - i) *No existe una guía, a priori, sobre la longitud máxima de retardos a introducir.*
    - ii) *A medida que se estiman los retardos sucesivos, se van perdiendo grados de libertad, lo que debilita la inferencia estadística.*
    - iii) *En la información de series temporales económicas, los valores sucesivos tienden a estar altamente correlacionados. Nos enfrentaremos entonces a problemas de multicolinealidad.*
  - Para resolver estos problemas, se imponen hipótesis a priori sobre la forma de las ponderaciones. Se trata de cambiar la parametrización del modelo, de modo que aparezca un número finito de parámetros.

### Solución: modelo truncado de Koyck (retardo geométrico)

- KOYCK propuso un método ingenioso de estimación de los modelos de retardos distribuidos. Supongamos que comenzamos con un modelo  $DL(\infty)$ . KOYCK realiza un supuesto (que se ha convertido en muy habitual), consistente en que las ponderaciones de las variables explicativas retardadas son todas positivas ( $\beta_i > 0 \forall i$ ) y se reducen geoméricamente a lo largo del tiempo de acuerdo al siguiente esquema:

$$\beta_i = \beta_0 \cdot \lambda^i, \text{ donde } i \text{ es el retardo y } \lambda \in (0,1)$$

- El supuesto  $\lambda \in (0,1)$  es razonable, pues evita una especificación en la que el pasado lejano de  $x$  tenga un impacto más importante sobre  $y$  que los valores más recientes de  $x$ .

<sup>7</sup> Con los supuestos planteados anteriormente, no debería haber problemas para reproducir los resultados del modelo de regresión clásico, pues las distintas variables explicativas son todas deterministas.

Sin embargo, en la práctica, se plantean 2 posibles **dificultades**:

- Orden infinito: Cuando  $r = \infty$  (es decir, la estructura de retardos es de orden infinito) es imposible estimar directamente el modelo, pues no tendríamos observaciones suficientes para ello (nos quedaríamos sin grados de libertad).
- Multicolinealidad: Los retardos consecutivos de una variable económica tienden a estar correlacionados entre sí. Cuanto mayor sea la correlación entre los retardos de  $x$ , más importante será la presencia de multicolinealidad aproximada en el modelo de regresión.

– Por tanto, podemos reescribir la ecuación inicial del siguiente modo:

$$y_t = \alpha + \sum_{i=0}^{\infty} \beta_0 \cdot \lambda^i \cdot x_{t-i} + \varepsilon_t$$

$$y_t = \alpha + \beta_0 \cdot \sum_{i=0}^{\infty} \lambda^i \cdot x_{t-i} + \varepsilon_t$$

○ En este modelo, el multiplicador de largo plazo sería:

$$m_{\infty} = \sum_{i=0}^{\infty} \beta_0 \cdot \lambda^i = \frac{\beta_0}{(1 - \lambda)}$$

– Si al proceso  $DL(\infty)$  le restamos  $\lambda \cdot y_{t-1}$  se obtiene:

$$y_t - \lambda \cdot y_{t-1} = \alpha \cdot (1 - \lambda) + \beta_0 \cdot \sum_{i=0}^{\infty} \lambda^i \cdot x_{t-i} - \beta_0 \cdot \sum_{i=1}^{\infty} \lambda^i \cdot x_{t-i} + \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$$

$$y_t - \lambda \cdot y_{t-1} = \alpha \cdot (1 - \lambda) + \beta_0 \cdot x_t + \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$$

$$y_t = \alpha \cdot (1 - \lambda) + \beta_0 \cdot x_t + \lambda \cdot y_{t-1} + \underbrace{\varepsilon_t - \lambda \cdot \varepsilon_{t-1}}_{=v_t}$$

donde  $v_t$  sigue un proceso  $MA(1)$ .

○ El proceso descrito se conoce como **transformación de Koyck**. Al comparar la última ecuación con el  $DL(\infty)$  vemos la enorme simplificación que se ha realizado. En vez de tener que estimar  $\infty$  incógnitas, ahora sólo encontramos 3:  $\alpha, \beta_0, \lambda$ .

#### ▪ Características de la transformación de Koyck:

- 1) Transforma un modelo de retardos distribuidos a un modelo regresivo.
- 2) Es probable que la aparición del retardo de  $y$  cree algunos problemas estadísticos. Estamos introduciendo una variable explicativa estocástica, mientras que MCO parte del supuesto de variables explicativas deterministas. A su vez, debemos conocer si  $y_{t-1}$  está correlacionada con el término de error.
- 3) En el modelo original, el término de error era un paseo aleatorio, mientras que en el modelo final será un  $MA(1)$ . En consecuencia, los diferentes componentes  $v_t$  se encontrarán serialmente correlados.

#### ▪ Aplicación del modelo de Koyck (HEA):

– Aunque el modelo de Koyck resulta claro, es ad-hoc (ya que se obtiene a través de un proceso meramente algebraico) y se encuentra desprovisto de cualquier soporte teórico. Para otorgarle una aplicación teórica, fundamentaremos en él el *modelo de expectativas adaptativas*:

○ Partimos del siguiente proceso:

$$y_t = \beta_0 + \beta_1 \cdot x_t^* + u_t$$

donde la variable explicativa  $x^*$  no es directamente observable, por lo que podemos proponer la siguiente hipótesis sobre la manera en la que se conforman las expectativas:  $x_t^* - x_{t-1}^* = \gamma \cdot (x_t - x_{t-1}^*)$ , donde  $\gamma \in (0,1)$  es el coeficiente de expectativas.

○ Esto es lo que se conoce como la hipótesis de expectativas adaptativas (HEA), popularizada por autores como FRIEDMAN y CAGAN. De esta forma, los agentes adaptarán sus expectativas a la luz de la experiencia pasada, aprendiendo de sus errores.

○ Una manera de reescribir el proceso es:  $x_t^* = \gamma \cdot x_t + (1 - \gamma) \cdot x_{t-1}^*$ .

○ Si introducimos esta ecuación en el proceso obtenemos:

$$y_t = \beta_0 + \beta_1 \cdot (\gamma \cdot x_t + (1 - \gamma) \cdot x_{t-1}^*) + u_t$$

$$y_t = \beta_0 + \beta_1 \cdot x_t + \beta_1 \cdot (1 - \gamma) \cdot x_{t-1}^* + u_t$$

○ Retardando el proceso un período, multiplicándolo por  $(1 - \gamma)$  y restando el producto:

$$y_t = \gamma \cdot \beta_0 + \gamma \cdot \beta_1 \cdot x_t + (1 - \gamma) \cdot y_{t-1} + v_t$$

donde  $v_t = u_t - (1 - \gamma) \cdot u_{t-1}$

– Valoración de las expectativas adaptativas en los modelos econométricos:

○ *Aspectos positivos*: Proporciona un medio relativamente simple de diseñar modelos de expectativas en la teoría econométrica a la vez que formula un comportamiento por parte de los agentes razonable, dado que se asume que los agentes aprenden e sus errores y conforman sus expectativas de una forma *backward looking*.

- *Aspectos negativos*: Hasta la introducción de la hipótesis de las expectativas racionales (HER), la HEA fue dominante. No obstante, con los desarrollos realizados por MUTH, LUCAS o SARGENT, se ponen de relieve los fallos de los modelos con expectativas adaptativas. Bajo estas, los agentes podrían ser sistemáticamente engañados. Además, los agentes seguían un comportamiento *backward-looking* en vez de *forward-looking*, lo que puede ser inconsistente con el supuesto de racionalidad. Por lo tanto, puede considerarse como un avance la integración en la modelización económica y econométrica la existencia de agentes que conforman sus expectativas utilizando de manera eficiente toda la información disponible en el momento  $T$ .

## 2.3. Modelos con retardo de la endógena (ARDL( $p, r$ ))

### 2.3.1. Identificación

- Si aparecen valores retardados de la variable endógena como variables explicativas, entonces dejaría de cumplirse uno de los supuestos bajo los que desarrollamos las teorías de estimación e inferencia del modelo econométrico, pues algunas de las variables explicativas serían ahora variables estocásticas (ya que  $y_t$  lo es). Sin embargo:

- Si el término de error no tiene autocorrelación (p.ej. en modelos AR), entonces el problema de estimación no es muy importante. MCO dejará de ser insesgado, pero será consistente.
- Si el término de error sí tiene autocorrelación, entonces el problema de estimar con MCO es más grave, pues conduce a estimadores inconsistentes. En este caso, es recomendable la estimación por variables instrumentales.

#### ▪ Consecuencias cuando el error está autocorrelacionado

- Supongamos el siguiente modelo ARDL(1,0) en el que un retardo de la variable endógena actúa como variable explicativa y en el que el error del modelo está autocorrelacionado:

$$y_t = \beta_0 + \beta_1 \cdot y_{t-1} + \beta_2 \cdot x_t + \varepsilon_t$$

$$\varepsilon_t = \rho \cdot \varepsilon_{t-1} + u_t$$

- Iterando hacia atrás, podemos comprobar que:

$$y_{t-1} = \beta_0 + \beta_1 \cdot y_{t-2} + \beta_2 \cdot x_{t-1} + \varepsilon_{t-1}$$

- Por tanto, en la primera ecuación, en la que  $y_{t-1}$  actúa como variable explicativa, vemos que esta variable está correlacionada con el error (ambos contienen  $\varepsilon_{t-1}$ ):

$$E[y_{t-1} \cdot \varepsilon_t] \neq 0$$

- Esto supone una violación de la condición de ortogonalidad. Entonces, el estimador  $\hat{\beta}$  de MCO es, no solamente sesgado, sino también inconsistente.

- Si llamamos  $X$  a la matriz que contiene las variables explicativas, podemos comprobar:

$$plim \hat{\beta} = \beta + plim \left( \frac{X'X}{N} \right)^{-1} \cdot plim \left( \frac{X'\varepsilon}{N} \right)$$

$$plim \hat{\beta} = \beta + \sum x'x^{-1} plim \cdot \begin{pmatrix} E[y_{t-1}] \cdot \varepsilon \\ \vdots \\ E[x_t \cdot \varepsilon] \end{pmatrix}$$

- Por tanto, en general, todos los estimadores del vector  $\hat{\beta}$  serán **inconsistentes** (salvo cuando la matriz  $\sum x'x^{-1}$  tenga elementos 0 en la columna  $k$ ). Es decir, en general, la endogeneidad de un solo regresor, provoca que el estimador MCO sea inconsistente no sólo para el coeficiente del regresor endógeno, sino para todos los  $K$  coeficientes. La gravedad de este fenómeno dependerá de la correlación entre el regresor endógeno y los otros regresores.

### 2.3.2. Estimación

#### Método de variables instrumentales

- El **método de variables instrumentales (VI)** es el más utilizado para realizar estimaciones en situaciones muy diversas en las que las variables explicativas pueden presentar endogeneidad.
  - Supongamos que existe una matriz  $Z$  de dimensión  $N \times K$  (misma dimensión que  $X$ ) que cumple las siguientes 3 propiedades:

$$\text{plim} \left( \frac{Z' \varepsilon}{N} \right) = 0$$

$$\text{plim} \left( \frac{Z' X}{N} \right) = Q_{ZX}, \text{ donde } Q_{ZX} \text{ existe y es una matriz no singular}$$

$$\text{plim} \left( \frac{Z' Z}{N} \right) = Q_{ZZ}, \text{ donde } Q_{ZZ} \text{ existe y es una matriz no singular}$$

- Intuitivamente, las variables que componen la matriz  $Z$  cumplen:

- Condición de exogeneidad: No están correlacionadas con el error del modelo.
- Condición de relevancia: Están correlacionadas con  $X$ .
- La tercera condición es trivial y garantiza que se puede obtener el estimador de VI obligando a que los instrumentos sean linealmente independientes entre sí (*para evitar multicolinealidad*).

- Transformamos el modelo original pre-multiplicando por la matriz  $Z$  traspuesta:

$$y = X \cdot \beta + \varepsilon$$

$$Z' \cdot y = Z' \cdot X \cdot \beta + Z' \cdot \varepsilon$$

$$y^* = X^* \beta + \varepsilon^*$$

- Utilizamos el método MCO para estimar  $\beta$  en el nuevo modelo transformado:

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

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

- Que es el **estimador VI**. Haciendo uso de las condiciones impuestas anteriormente, podemos demostrar que este estimador es consistente.

- La principal **dificultad** del método VI, por tanto, consiste en *seleccionar adecuadamente las variables instrumentales  $z$  que componen la matriz  $Z$* . En este sentido, se pueden plantear 2 *dificultades*:
  - Condición de exogeneidad: El verdadero error del modelo no es observable, por lo que es difícil estar seguro de que los instrumentos realmente no están correlacionados en el límite con los errores.
  - Condición de relevancia: Los instrumentos de las variables endógenas<sup>8</sup> deben estar muy correlacionados con las variables explicativas  $x$  o, de lo contrario, el estimador tendría una varianza muy elevada.

- Cuando existe un número mayor de instrumentos que de variables explicativas, entonces decimos que estamos en una situación de *sobreidentificación*. Existirán varias regresiones posibles con el método de variables instrumentales, todas consistentes, pero con distinta eficiencia (varianza).

#### Mínimos Cuadrados en 2 etapas (MC2E) para situaciones con sobreidentificación

- El método de **mínimos cuadrados bietápicos** o **en 2 etapas (MC2E)** ayuda a realizar una estimación con variables instrumentales en situaciones de sobreidentificación. En particular, *el estimador MC2E combina de manera eficiente la información de múltiples instrumentos* para regresiones sobreidentificadas.
  - El método de MC2E se estructura en 2 etapas:
    - En una 1ª etapa, se determinan cuáles son las variables explicativas que son exógenas y cuáles son endógenas. A continuación, se construirán regresiones auxiliares de cada variable del modelo sobre las variables exógenas<sup>9</sup>. Obtenemos así el vector de variables explicadas.

<sup>8</sup> Nótese que los mejores instrumentos de las variables exógenas son ellos mismos, pues cumplen perfectamente las 3 condiciones.

<sup>9</sup> Con estas regresiones auxiliares, se obtienen estimaciones de las variables explicativas endógenas sin recurrir a la información muestral disponible de dichas variables endógenas.



- En la 2ª etapa, se sustituirán en el modelo original las variables explicativas endógenas por sus estimaciones. Esta ecuación será estimada por MCO mediante el sistema de ecuaciones normales, ya conocido.
- Este método podrá emplearse para la estimación de cualquier ecuación que esté exactamente identificada o sobreidentificada, siendo el segundo caso el que hace que este método tenga mayor interés.
- La **expresión del estimador MC2E** es:

$$\widehat{\beta}_{MC2E} = [X'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'y]$$

- En el caso de que la ecuación esté exactamente identificada (matriz  $Z$  de tamaño igual a  $X$ ) vemos que esta expresión se reduce al caso más sencillo visto antes de VI (i.e.  $\widehat{\beta} = (Z'X)^{-1}Z'y$ ).
- Puede demostrarse que el estimador MC2E es el **estimador lineal de VI eficiente** (es decir, el que tiene una mínima matriz de covarianzas).

### 2.3.3. Verificación (test de HAUSMAN y WU)

- La aparición de retardos de la variable endógena como variables explicativas en presencia de autocorrelación es un caso muy habitual en el que se debe utilizar un procedimiento de VI para obtener estimaciones consistentes.
  - Sin embargo, es aconsejable verificar que las propiedades de exogeneidad del resto de variables sí se satisfacen, pues, de lo contrario, obtendríamos igualmente estimadores inconsistentes.
- El **problema** es que *la exogeneidad no puede ser testada directamente* porque:
  - La condición  $E[\varepsilon \cdot x] = 0$  no puede ser observada directamente (el parámetro poblacional  $\varepsilon$  es desconocido).
  - Utilizar los residuos de MCO también es inútil, puesto que, por construcción, se cumple siempre que  $E[\widehat{\varepsilon} \cdot x] = 0$ , incluso en presencia de endogeneidad.
- **¿Cuál es entonces la solución?** Se puede comprobar la exogeneidad utilizando información adicional, en concreto, variables instrumentales. Ésta es la idea del test de HAUSMAN y WU.
  - Supongamos un modelo en el que se duda de la exogeneidad de  $r$  variables. El contraste consiste en:
    - Estimar el modelo por MCO y obtener la suma residual  $SR_0$ .
    - Estimar el modelo por VI y obtener la suma residual  $SR_1$ .
  - El estadístico a estimar es:  $(\widehat{\beta}_{MCO} - \widehat{\beta}_{VI})' \cdot [var(\widehat{\beta}_{VI}) - var(\widehat{\beta}_{MCO})] \cdot (\widehat{\beta}_{MCO} - \widehat{\beta}_{VI}) = \frac{SR_0 - SR_1}{\sigma^2} \sim \chi_r^2$ 
    - Se distribuye como una chi-cuadrado con  $r$  grados de libertad bajo la hipótesis nula de que todas las variables explicativas del modelo original son exógenas. Un valor elevado del estadístico rebatiría tal supuesto y mostraría la necesidad de utilizar un procedimiento de estimación de variables instrumentales.
  - **Intuición:** La lógica del test de Hausman es la siguiente:
    - Bajo la hipótesis nula, tanto el estimador de MCO como el de VI son consistentes (aunque este último no eficiente). Por tanto, la diferencia  $d = \widehat{\beta}_{MCO} - \widehat{\beta}_{VI}$  converge en probabilidad a 0.
    - Bajo la hipótesis alternativa, el estimador de MCO será inconsistente, por lo que dicha diferencia será no nula.

### 2.3.4. Predicción

- Una vez estimado adecuadamente el modelo a partir de datos de series temporales para una muestra de tamaño  $T$ , el investigador puede utilizar dicho modelo para realizar **predicciones**, es decir, para estimar los valores de la variable de interés en periodos posteriores a  $T$  (fuera de la muestra).

- Las predicciones de un modelo de regresión dinámica se calculan con los mismos principios que en el caso univariante. La utilidad de estos modelos para la predicción depende del intervalo entre observaciones:
  - Si el intervalo es grande (p.ej. datos anuales), la relación entre la variable explicativa y la respuesta no suele contener retardos, por lo que, para prever  $y_t$  es necesario conocer previamente  $x_t$ . Por tanto, es necesario construir un modelo univariante de  $x_t$  para prever dicha variable. En este caso, es posible que la mejora en la predicción de  $y_{t+k}$  que aporte el modelo dinámico basado en la predicción de  $x_t$  no sea muy grande respecto al modelo univariante para  $y_t$ <sup>10</sup>.
  - Si el periodo es corto (p.ej. datos diarios o por minutos), es frecuente que existan retardos en la relación entre las variables, de manera que  $x_{t-k}$  afecta a  $y_t$ . En este caso, diremos que la variable  $x_t$  es un indicador adelantado de la respuesta en  $t + k$ . Como el  $x_t$  es conocido en el origen de la predicción, el modelo de regresión dinámica suele proporcionar mejores predicciones que el univariante.
- ¿Qué criterio se utiliza para seleccionar el mejor predictor? Naturalmente, queremos que la predicción esté lo más cerca posible del verdadero valor que queremos prever. En este, sentido, definimos el error de predicción como la diferencia entre estos dos valores:

$$e_T(k) = y_{T+k} - \hat{y}_T(k)$$

- En predicción con modelos ARIMA, el criterio más habitual de selección de predictores es el de minimizar el error cuadrático medio de predicción (ECMP).

$$\hat{y}_T(k) = \operatorname{argmin}(ECMP) = \operatorname{argmin}(E[e_T(k)^2 | Y_T])$$

- En los años 30, WIENER y KOLMOGOROV demostraron que el predictor óptimo (es decir, el que minimiza el ECMP) es la esperanza de la variable futura condicionada a los datos observados:

$$\hat{y}_T(k) = E[y_{T+k} | Y_T]$$

### 2.3.5. Aplicación práctica: expectativas racionales

- Los modelos de expectativas racionales son una clase de *modelos dinámicos con gran importancia en la literatura económica y econométrica*. Se trata de modelos en los que la variable endógena depende, entre otros factores, de las expectativas que los agentes económicos tienen actualmente (o tuvieron en el pasado) de los valores futuros de alguna variable exógena o de la propia variable endógena.
  - Supongamos que queremos explicar la inversión empresarial en función de los tipos de interés. Algunos modelos teóricos, como el de JORGENSON, señalan que la inversión dependerá de los tipos de interés actuales, pero también de las expectativas de tipos futuros. En coherencia con tales teorías, podíamos especificar un modelo como el siguiente:

$$y_t = \beta_0 + \beta_1 \cdot x_t + \beta_2 \cdot x_{t+1}^e + \varepsilon_t$$

donde:

- $y_t$  es la inversión en  $t$ .
- $x_t$  es el tipo de interés en  $t$ .
- $x_{t+1}^e$  son las expectativas que los agentes tienen en el momento  $t$  acerca del valor del tipo de interés en el periodo  $t + 1$ .
- $\varepsilon_t$  es una perturbación que sigue un proceso de ruido blanco.
- Como la expectativa  $x_{t+1}^e$  no es observable, podemos adoptar uno de los siguientes 2 enfoques:
  - Hipótesis de expectativas adaptativas (HEA): Suponer que los agentes forman sus expectativas a partir de la evolución de la variable en periodos anteriores. El modelo pasaría a ser un modelo de retardos distribuidos como el analizado anteriormente.
  - Hipótesis de expectativas racionales (HER): Suponer que los agentes explotan de forma óptima toda la información de la que disponen, sin establecer una forma funcional concreta.
- En este último caso, sabemos que la predicción óptima (racional), que minimiza el error cuadrático medio de predicción (ECMP) es la esperanza condicionada a toda la información disponible en  $t$ :

$$x_{t+1}^e = E[x_{t+1} | \Omega_t]$$

- El error racional de la predicción un periodo hacia delante ( $u_t$ ) será necesariamente un ruido blanco, con media 0 y sin autocorrelación, pues, de lo contrario, el agente no estaría aprovechando óptimamente toda la información. Tenemos, por tanto, que:

$$x_{t+1}^e = x_{t+1} + u_t$$

<sup>10</sup> Sin embargo, el modelo puede ser muy adecuado para la predicción condicionar de la respuesta en función de posibles valores de la variable explicativa (p.ej. para el análisis de sensibilidad).

- Así, podemos reescribir el modelo original como:

$$y_t = \beta_0 + \beta_1 \cdot x_t + \beta_2 \cdot (x_{t+1} + u_t) + \varepsilon_t$$

$$y_t = \beta_0 + \beta_1 \cdot x_t + \beta_2 \cdot x_{t+1} + v_t$$

donde el nuevo error  $v_t = \beta_2 \cdot u_t + \varepsilon_t$  está libre de correlación (por ser la suma de 2 ruidos blancos).

- En definitiva:
  - El modelo tal y como lo acabamos de expresar puede ser *estimado eficientemente por un método habitual como MCO*.
  - Ahora bien, si en lugar de incorporar las expectativas un periodo hacia delante, introdujéramos *expectativas con horizontes más largos, puede demostrarse que el error del modelo sí estaría autocorrelacionado*. En estos casos, deberemos aplicar procedimientos de VI como el visto.

### 3. MODELOS DINÁMICOS CON VARIABLES NO ESTACIONARIAS

- Para aplicar métodos de estimación y de verificación estándar en un modelo multivariante, **se requiere que las variables sean estacionarias**, pues la mayoría de la teoría econométrica está basada en dicho supuesto.
- En caso de que las series fueran **no estacionarias**:
  - Podríamos tener problemas de *regresión espuria*: La regresión entre variables no estacionarias puede conducir a estimadores que parecen significativos pero que en realidad son inválidos.
  - Pero no necesariamente. Cuando dos o más variables están *cointegradas*, podríamos obtener estimadores súper-consistentes.

#### 3.1. Relaciones espurias (GRANGER y NEWBOLD, 1974)

- Supongamos 2 variables que siguen cada una un paseo aleatorio (por tanto, no estacionario):
 
$$y_t = y_{t-1} + u_t \quad \text{donde } u_t \stackrel{i.i.d.}{\sim} N(0,1)$$

$$x_t = x_{t-1} + v_t \quad \text{donde } v_t \stackrel{i.i.d.}{\sim} N(0,1)$$
- Si realizamos una regresión por MCO con la siguiente especificación:
 
$$y_t = \beta_0 + \beta_1 \cdot x_t + \varepsilon_t$$
  - Entonces es muy probable que obtengamos coeficientes de  $\beta_1$  distintos de 0 y aparentemente significativos según los métodos de contraste de hipótesis habituales.
    - Por ejemplo, tendremos que el estadístico  $t = \hat{\beta}_1 / \sqrt{\text{var}(\hat{\beta}_1)}$  tiene valores superiores a los valores críticos de la tabla  $t$  de Student para los niveles de significación normales.
  - Sin embargo, estas conclusiones son evidentemente erróneas, puesto que, por construcción, hemos supuesto que  $x_t$  e  $y_t$  son paseos aleatorios y, por tanto, independientes entre sí.
    - Es decir, en realidad no existe ninguna relación causal entre  $x_t$  e  $y_t$ .
    - Es lo que se conoce como **regresión espuria**.
    - Esto se debe, esencialmente, a que, cuando las series son no estacionarias, las distribuciones de los principales estadísticos habitualmente utilizados en contrastes de hipótesis dejan de tener distribuciones estándar (como  $t$ ,  $F$  o chi-cuadrado) y pasan a tener distribuciones que son funciones de un movimiento Browniano geométrico.
- Una posible **solución** para evitar estos problemas es diferenciar las series para convertirlas en estacionarias y después aplicar los métodos que hemos visto antes.
  - Sin embargo, esta alternativa, que fue popular hace unos años, tiene también sus inconvenientes. Es posible que, aunque las dos series sean no estacionarias, una de ellas explique totalmente el comportamiento no estacionario de la otra (es el caso de la cointegración). En este caso, tomar diferencias limitaría innecesariamente la información utilizada (reduciendo así la eficiencia).

### 3.2. Cointegración

- La **demostración de causalidad** en el ámbito de series temporales se ha revelado una cuestión compleja.
  - Uno de los instrumentos más utilizados para demostrar esta relación ha sido la demostración de la existencia de una relación de *cointegración* entre 2 variables, que además nos permite excluir la posibilidad de que la relación entre ambas sea espuria.
- Se dice que 2 series no estacionarias integradas de orden  $d_0$  ( $I(d_0)$ ) están *cointegradas* cuando existe alguna combinación lineal entre ambas que produce una serie cuyo orden de integración sea menor que  $d_0$ .
  - Es decir, si tenemos 2 series ( $x_t$  e  $y_t$ ) integradas de orden  $d_0$ , pero podemos construir una serie  $n_t = \alpha_1 \cdot x_t + \alpha_2 \cdot y_t$  integrada de orden  $d_1 < d_0$ , diremos que  $x_t$  e  $y_t$  están cointegradas.
    - A la combinación  $(\alpha_1, \alpha_2)$  se le denomina *relación de cointegración*.
  - Cuando existe una relación de cointegración, una de las dos variables explica parte de la tendencia de la otra. Es decir, la variable independiente explicará totalmente la no estacionariedad de la variable dependiente y la perturbación seguirá un proceso aleatorio.
- A menudo, la teoría económica sugiere la existencia de dichas relaciones de cointegración o de equilibrio a largo plazo (p.ej. la teoría de la paridad del poder adquisitivo sobre el tipo de cambio o la teoría cuantitativa del dinero).
  - La existencia de esta relación de largo plazo también tiene implicaciones para el comportamiento de las variables en el corto plazo, ya que tiene que haber algún mecanismo que reconduce a las variables hacia su relación de equilibrio a largo plazo. Este mecanismo se modeliza a través de lo que se conoce como “*mecanismo de corrección de errores*” (MCE).
- Tests de cointegración (sobre el error de la regresión, que debería seguir un paseo aleatorio)
  - **Dickey-Fuller:** Es una prueba de raíz unitaria que detecta estadísticamente la presencia de conducta tendencial estocástica en las series temporales de las variables mediante contraste de hipótesis. En otras palabras, nos permite saber si hay presencia significativa de tendencia en las series temporales de las variables. Por ejemplo, en un modelo AR(1) existirá tendencia cuando el primer regresor sea igual o cercano a 1. Esto es equivalente a no estacionariedad ya que, si el proceso estocástico fuera estable, este coeficiente sería menor que 1 o muy cercano a 0. Es la herramienta más fácil de usar comparado con otros contrastes más complejos que también prueban la presencia de tendencia en los datos.
  - **Augmented Dickey-Fuller:** Es un test exigente. Consiste en una versión del test Dickey-Fuller aplicable a un conjunto más amplio y complejo de modelos de series temporales.
  - Durbin Watson (CRDW: *cointegrating regression Durbin Watson test*)
  - Philips Ouliaris de autocorrelación.

#### Cointegración y MCE (GRANGER, 1983)<sup>11</sup>

- La idea de la causalidad de Granger se basa en que el tiempo avanza en una única dirección: un acontecimiento pasado puede propiciar sucesos presentes o futuros, pero no al contrario.
  - El test de Granger (**Premio Nobel en 2003 junto a ENGLE “por haber desarrollado métodos de análisis temporales con tendencias comunes”**) consiste en comprobar si los resultados de una variable sirven para predecir otra variable, si tiene carácter unidireccional o bidireccional. Para ello se tiene que comparar y deducir si el comportamiento actual y el pasado de una serie temporal  $A$  predice la conducta de una serie temporal  $B$ .
  - Si ocurre el hecho, se dice que el “resultado  $A$ ” causa en el sentido de Wiener-Granger el “resultado  $B$ ”; el comportamiento es unidireccional. Si sucede lo explicado y además el “resultado  $B$ ” predice el “resultado  $A$ ”, el comportamiento es bidireccional, entonces el “resultado  $A$ ” causa el “resultado  $B$ ”, y el “resultado  $B$ ” causa el “resultado  $A$ ”.

<sup>11</sup> En un sistema bivalente de 2 variables  $(y, x)$ , la variable  $x$  no causa a la variable  $y$  en el sentido de Granger si, para todo  $s > 0$ , el error cuadrático medio de la predicción (ECMP) de  $y_{t+s}$  dado  $(y_1, \dots, y_t)$  es el mismo que el ECMP de  $y_{t+s}$  dado  $(y_1, \dots, y_t, x_1, \dots, x_t)$ .

## CONCLUSIÓN

- **Hasta los años 70**, los MES fueron ampliamente utilizados en el análisis coyuntural y las tareas de predicción por parte de las autoridades económicas, en consonancia con las recomendaciones de la Cowles Commission. En concreto, se fueron desarrollando modelos macroeconómicos cada vez más complejos que reflejaban la interdependencia de las distintas variables económicas.
- Sin embargo, **desde mediados de los años 70**, la utilidad de estos modelos fue puesta en duda por varios autores, entre los que destaca la *crítica de LUCAS*<sup>12</sup> (1976) y la *crítica de SIMS*<sup>13</sup> (1980). Este último considera que la capacidad predictiva de los modelos MES era dudosa por varias deficiencias:
  - La validez de las restricciones de exclusión: restricciones lineales y otras técnicas para lograr la identificación de los modelos y obtener así una interpretación estructural eran arbitrarias y no siempre respaldadas por la teoría económica.
  - La clasificación de las variables en exógenas y endógenas también es arbitraria y obvia la incertidumbre del investigador sobre el verdadero sentido de la causalidad.
- Estos aspectos motivaron a SIMS a proponer una alternativa en la modelización de las relaciones entre variables. SIMS propone los métodos de **Vectores Autorregresivos (VAR)**, que son una generalización para el caso multivariante de los modelos dinámicos uniecuacionales. Este método permite evitar la imposición de restricciones subjetivas sobre los coeficientes del modelo y permite considerar relaciones dinámicas generales entre las variables.
- Valoración de los VAR:
  - Los defensores de los VAR destacan las siguientes virtudes del método:
    - i) Es un método simple, no es preciso preocuparse por determinar cuáles son las variables endógenas y cuales las exógenas.
    - ii) La estimación es simple, debido a que se usa el método de MCO para cada ecuación por separado.
    - iii) Las predicciones obtenidas mediante este método son en muchas ocasiones mejores que las obtenidas con los modelos de ecuaciones simultáneas más complejos.
  - A pesar de ello, estos modelos también hacen frente a una serie de **críticas**:
    1. A diferencia de los MES, un modelo VAR es ateórico, al utilizar menos información previa. En los MES, la exclusión o inclusión de ciertas variables desempeña un papel crucial en la identificación del modelo.
    2. Debido a su acento en la predicción, los modelos VAR son menos apropiados para el análisis de políticas económicas.
    3. El mayor desafío práctico en el diseño de los modelos VAR es seleccionar la longitud adecuada de los retardos. A menos que la muestra sea muy grande, la estimación de los parámetros consumirá muchos grados de libertad.
    4. En sentido estricto, en un modelo VAR de  $m$  variables, todas las  $m$  variables han de ser estacionarias. Si no es así se tendrá que transformar la información, usando por ejemplo primeras diferencias.
    5. Como con frecuencia es difícil interpretar los coeficientes individuales estimados en los modelos VAR, los económetras estiman las funciones de impulso-respuesta, que analizan la respuesta de una variable dependiente del sistema VAR ante shocks en los términos de error. Ello permite analizar las posibles relaciones causales entre variables.

<sup>12</sup> ROBERT LUCAS fue galardonado con el Premio Nobel de Economía en 1995.

<sup>13</sup> CHRISTOPHER ALBERT SIMS fue galardonado con el Premio Nobel de Economía en 2011 junto a SARGENT por “sus investigaciones empíricas sobre las relaciones de causalidad en el ámbito macroeconómico”.

### **Modelos VARMA (VAR, Sims, 1980)**

- Otro punto de vista sobre las series temporales multivariantes es la generalización de los procesos ARMA para varias series temporales. Es lo que se conoce como procesos de VARMA (vectores autorregresivos de media móvil), entre los que los más importantes en la literatura son los procesos VAR.
- Los procesos VAR fueron propuestos por Sims en 1980 como alternativa a los modelos de ecuaciones simultáneas. Su estudio queda fuera del alcance de este tema, pues el tema versa sobre modelos uniecuacionales.



## 5.B.8 : MODELOS DE COMPONENTES NO OBSERVABLES: TENDENCIA, CICLO Y ESTACIONALIDAD. APLICACIONES EN EL ANÁLISIS DE COYUNTURA, ESTIMACIÓN, VERIFICACIÓN Y PREDICCIÓN.

### INTRODUCCIÓN

#### ▪ **Enganche:**

- Cuando hacemos referencia a una disciplina, en este caso a la **economía**, es conveniente hacer una reflexión sobre 2 cuestiones:
  - ¿Qué fenómenos estudia? [*objeto*]
  - ¿Qué procedimientos se deben seguir para estudiar dichos fenómenos? [*método*]
- De acuerdo con BLAUG (1985), la ciencia económica tiene como **objeto el estudio de la actividad económica**, la cual se define por 3 rasgos fundamentales:
  - 1) La actividad económica se deriva siempre de la *existencia de un problema económico*: dados unos recursos escasos, se pretende con ellos satisfacer necesidades humanas ilimitadas.
  - 2) En toda actividad económica existe la *posibilidad de elección* respecto a la asignación de recursos (es posible optimizar).
  - 3) Toda actividad económica está sujeta a un *coste de oportunidad* en la elección de distribución de los recursos. Por tanto, la actividad económica buscará minimizar este coste.
- Ahora bien, ¿qué procedimientos se deben seguir para estudiar dichos fenómenos? Como ciencia social, la economía utiliza **métodos** propios de la ciencia, destacando el uso del método deductivo contrastable<sup>1</sup>. Éste se basa en el planteamiento de una serie de hipótesis a priori que, posteriormente, son contrastadas empíricamente.
  - En esta labor, resulta clave la **econometría**.
    - Econometría significa literalmente “*medición de la economía*”, de modo que permite encontrar respuestas cuantitativas a preguntas económicas.
    - Sin embargo, su alcance va más allá de la mera medición y podemos definirla como la disciplina híbrida que combina la *teoría económica* con los métodos de la *inferencia estadística*<sup>2</sup>.

#### ▪ **Relevancia:**

- El estudio de la **econometría** es clave para cualquier economista, pues permite alcanzar los siguientes 3 objetivos fundamentales:
  - Análisis descriptivo: Detectar y medir las relaciones entre variables y reconocer hechos estilizados.
  - Análisis causal: Comprobar la validez de las teorías económicas y evaluar la efectividad de las políticas económicas.
  - Predicción.

<sup>1</sup> La teoría económica produce proposiciones teóricas no ambiguas que establecen relaciones determinísticas entre diferentes variables económicas (funciones de demanda, de producción de consumo agregado, de inversión, etc.). A través de las mismas, se pueden identificar un conjunto de *variables dependientes* e *independientes* y, en la mayoría de los casos, se pueden enunciar un conjunto de afirmaciones en cuanto a la dirección de los efectos que producen las variaciones en las variables dependientes.

Ningún modelo es capaz de incorporar todos los elementos que afectan a las variables dependientes. En consecuencia, la variable dependiente observada reflejará variaciones derivadas de variables omitidas y de elementos aleatorios presentes en el proceso. Para poder realizar *inferencia estadística* y analizar la causalidad del modelo será necesario asumir un conjunto de supuestos en cuanto a estos elementos aleatorios o “errores”.

Uno de los objetivos de la econometría es contrastar la validez de las teorías económicas. De acuerdo con el método deductivo contrastable desarrollado por KARL POPPER, un modelo o teoría nunca podrá ser completamente confirmado. Por otro lado, una teoría determinística será invalidada con una sola observación contraria a la misma. La introducción de elementos estocásticos en el modelo lo transforma de una afirmación exacta a una descripción probabilística sobre los resultados esperados, por lo que sólo la preponderancia de evidencia contradictoria podrá invalidar un modelos probabilístico.

<sup>2</sup> De hecho, en la primera edición de la revista *Econometrica*, la *Econometric Society* enunció que “su principal objetivo sería el de promocionar los estudios que buscasen la unificación del binomio teórico-empírico en el estudio de los problemas económicos, a través del desarrollo de metodologías rigurosas y constructivas que permitiesen a la ciencia económica aproximarse al resto de ciencias naturales”.

“La dimensión cuantitativa de la economía está conformada por la ciencia económica, las matemáticas y la estadística. La experiencia nos muestra que cada una de estas perspectivas es necesaria pero no suficiente para la comprensión de las relaciones cuantitativas en las economías modernas. Es, por tanto, la combinación de las tres a través de la econometría la que permite obtener una comprensión profunda de las mismas”.

### ▪ Contextualización:

- En Econometría aplicada, se trabaja con 3 tipos de datos muestrales:
  - Datos de sección cruzada: Conjunto de datos recogidos observando diversas unidades económicas (como familias, empresas, ciudades...) en un mismo instante de tiempo o haciendo abstracción de la dimensión temporal.
  - Datos de series temporales: Conjunto ordenado de datos sobre una unidad económica a lo largo del tiempo.
  - Datos de panel (combinación de los 2 anteriores): Datos de sección cruzada observados a lo largo de varios periodos de tiempo.
- La mayoría de series temporales de variables económicas presentan características como *tendencia*, *ciclo* o *estacionalidad* que no podemos observar directamente de los datos. Habitualmente, una serie temporal se expresa como la suma de **3 componentes no observables**:
  - Un componente tendencial, que trata de captar la parte no estacionaria de la serie.
  - Un componente cíclico, que es un componente estacionario.
  - Además, en muchos casos, existe un 3<sup>er</sup> componente implícito que captura la variación estacional.
- Los componentes no observables de las series temporales son de especial interés para los economistas, especialmente en *macroeconomía*. En este sentido, los **modelos de componentes no observables (MCNO)** son una *herramienta extremadamente útil en múltiples aplicaciones tanto teóricas como de política económica*. Por ejemplo:
  - Contratación empírica de modelos de ciclos reales de la NMC.
  - Estimación de datos coyunturales en tiempo real<sup>3</sup>.
  - Completar series temporales en las que existe omisión de algunos datos.
  - Estimación de la inflación subyacente<sup>4</sup>, output gap, detección de cambios de tendencia en la actividad económica (*turning points*)...
    - El objetivo de esta exposición es:
      - Estudiar los métodos estadísticos que permiten separar estos componentes no observables; e
      - Ilustrar las principales aplicaciones de los modelos de componentes no observables en el *análisis de coyuntura, estimación, verificación y predicción*.
- Desde un punto de vista histórico,
  - Tradicionalmente, los modelos de componentes no observables desarrollados en los años 40 adoptaban una *visión determinista de dichos componentes*.
  - Esto cambió entre los años 60 y 80, cuando se popularizan *métodos de alisado basados en medias móviles*, con filtros tan conocidos como el *filtro Hodrick-Prescott (HP)*.
  - En la actualidad, los modernos modelos de componentes no observables (MCNO) asumen que cada componente sigue un proceso estocástico y que, cuando todos ellos se combinan, dan lugar a la serie observada.
    - En la práctica, existen programas informáticos como TRAMO (*Time series with ARIMA-based Missing observations and Outliers*) y SEATS (*Signal Extraction in ARIMA Time Series*) que permiten aplicar fácilmente estos procedimientos a series reales.

### ▪ Problemática (Preguntas clave):

–

<sup>3</sup> Conocer con rapidez datos como el PIB, el nivel de precios, empleo, etc. es fundamental para diseñar la política económica de corto plazo.

<sup>4</sup> Los datos de inflación observados son altamente volátiles. Las decisiones de política monetaria no pueden basarse en estos datos, puesto que ello implicaría cambios constantes de la orientación de la política monetaria, altamente distorsionantes y además imposibles desde un punto de vista operativo. Es necesario, por tanto, que las decisiones de política económica se basen en datos filtrados.

■ **Estructura:****1. MODELOS CON TENDENCIAS DETERMINISTAS Y ESTOCÁSTICAS: COMPONENTES Y DESCOMPOSICIÓN POR MODELOS DETERMINISTAS***1.1. Idea principal: Series con tendencias y patrones cíclicos**1.2. Componentes de la serie*

## 1.2.0. Descomposición general de una serie: Marco de análisis

## 1.2.1. Componente tendencial

Relevancia de la tendencia

Tipos de tendencias

Descripción de procesos: Ejemplos de tendencias deterministas y estocásticas

Eliminación del componente tendencial

## 1.2.2. Componente estacional

Concepto, tipos y ejemplos

Relevancia

Eliminación del componente estacional

## 1.2.3. Ciclo

*1.3. Valoración***2. MÉTODOS DE FILTRADO AD-HOC***2.1. Concepto**2.2. Ejemplo: filtro Hodrick-Prescott**2.3. Valoración de los filtros ad-hoc*

Ventajas

Limitaciones

**3. MÉTODOS MODERNOS DE COMPONENTES NO OBSERVABLES (BASADOS EN MODELOS)***3.1. Identificación: caracterización de los componentes*

## 3.1.1. Tendencia

## 3.1.2. Estacionalidad (ARIMA estacional multiplicativo)

## 3.1.3. Combinación con el ciclo estocástico

*3.2. Estimación: filtro de Kalman**3.3. Predicción***4. APLICACIONES PRÁCTICAS***4.1. Output gap y saldo ajustado al ciclo*

## 4.1.1. Idea: medición del efecto macroeconómico del presupuesto

## 4.1.2. Descomposición del déficit público

## 4.1.3. Cálculo del déficit estructural por el método indirecto

## 1) Estimación de los niveles potenciales de actividad

## 2) Obtención del saldo presupuestario ajustado al ciclo económico y el saldo estructural

*4.2. Otros*

## 1. MODELOS CON TENDENCIAS DETERMINISTAS Y ESTOCÁSTICAS: COMPONENTES Y DESCOMPOSICIÓN POR MODELOS DETERMINISTAS

### 1.1. Idea principal: Series con tendencias y patrones cíclicos

- Los modelos ARIMA nos permiten estimar de manera insesgada una gran variedad de procesos estocásticos que satisfacen las condiciones de estacionariedad y ergodicidad.
  - Desafortunadamente, una gran variedad de series temporales muestra tendencias o patrones de crecimiento, unidos a patrones cíclicos. La literatura macroeconómica ha buscado separar la información cíclica de la tendencial para su posterior análisis.
  - Los antecedentes los encontramos en la obra *Measuring Business Cycles* (1946) de BURNS y MITCHELL, con la que se inicia la literatura de los *ciclos económicos*.
    - En este tratado aparece una definición del ciclo de negocios que se ha hecho clásica<sup>5</sup> y que ha sido la base de su estudio empírico, basado en 2 etapas:
      - *Datación o fechado de los ciclos*:
        - BURNS y MITCHELL ofrecieron por primera vez una datación de los ciclos estadounidenses, mediante la identificación de los puntos de cambio de tendencia (*turning points*) de la actividad económica, de modo que el ciclo se mide de valle a valle.
        - Desde entonces, el *National Bureau of Economic Research* (NBER) estadounidense mediante su comité de expertos ha realizado su seguimiento para la economía estadounidense y define una *recesión* como un periodo de declive significativo de la producción, la renta, el empleo y el comercio, y marcado por una contracción generalizada en muchos sectores de la economía y que dura “más de unos pocos meses” (<https://www.nber.org/research/business-cycle-dating>)<sup>6</sup>.
      - *Medición de los ciclos o caracterización cuantitativa*: Los valores observados de las variables macroeconómicas de estudio son el resultado, grosso modo, de 3 componentes: *tendencial, cíclico y estacional*. Para estudiar los ciclos económicos es relevante aislar el componente cíclico como veremos a lo largo de la exposición.
        - El *componente estacional* se suele eliminar en primer lugar de modo que las series queden limpias de las fluctuaciones periódicas que se producen dentro del año. Este paso suele realizarse mediante técnicas estadísticas avanzadas<sup>7</sup>.
        - De la serie desestacionalizada se separaría el *componente cíclico* del *componente tendencial*. Como no se conoce con certeza el funcionamiento de las economías (de ahí el uso de diferentes modelos para captar tal funcionamiento y sobre los que discutiremos en las secciones siguientes), no existe un único método correcto para separar la tendencia del ciclo:
          - Una primera aproximación sería utilizar una regresión lineal simple por Mínimos Cuadrados Ordinarios. Sin embargo, esto supondría asumir que el componente tendencial es lineal, lo que implica aceptar que las economías están siempre en un hipotético estado estacionario (creciendo a una tasa constante). Este supuesto es muy

<sup>5</sup> “Los ciclos económicos son fluctuaciones en la **actividad agregada** de las **economías de mercado**: un ciclo consiste en **expansiones** que ocurren al mismo tiempo en muchas actividades económicas, seguido de **recesiones o contracciones** igualmente generalizadas, y reactivaciones que conducen a la fase de expansión del siguiente ciclo; esta secuencia de cambios es **recurrente, pero no periódica**; en cuanto a su duración, los ciclos económicos varían desde **más de 1 año** hasta 10 o 12 años; no son divisibles en ciclos más cortos de duración exacta”.

“Business cycles are a type of fluctuation found in the aggregate economic activity of nations that organize their work mainly in business enterprises: a cycle consists of expansions occurring at about the same time in many economic activities, followed by similarly general recessions, contractions, and revivals which merge into the expansion phase of the next cycle; this sequence of changes is recurrent but not periodic; in duration business cycles vary from more than one year to ten or twelve years; they are not divisible into shorter cycles of similar character with amplitudes approximating their own.”

<sup>6</sup> What is a recession? What is an expansion?

The NBER's traditional definition of a recession is that it is a significant decline in economic activity that is spread across the economy and that lasts more than a few months. The committee's view is that while each of the three criteria—depth, diffusion, and duration—needs to be met individually to some degree, extreme conditions revealed by one criterion may partially offset weaker indications from another. For example, in the case of the February 2020 peak in economic activity, we concluded that the drop in activity had been so great and so widely diffused throughout the economy that the downturn should be classified as a recession even if it proved to be quite brief. The committee subsequently determined that the trough occurred two months after the peak, in April 2020. An expansion is a period when the economy is not in a recession. Expansion is the normal state of the economy; most recessions are brief. However, the time that it takes for the economy to return to its previous peak level of activity may be quite extended.

<https://www.nber.org/research/business-cycle-dating/business-cycle-dating-procedure-frequently-asked-questions>

<sup>7</sup> Una de las herramientas más utilizadas internacionalmente son los programas TRAMO (*Time series Regression with ARIMA noise, Missing values and Outliers*) y SEATS (*Signal Extraction in ARIMA Time Series*), desarrollados por GÓMEZ y MARAVALL, del Banco de España.

restrictivo, y de hecho, la teoría moderna del crecimiento endógeno [ver tema 3.A.43] sugiere que la tasa de progreso técnico puede variar con la actividad de innovación endógena de las empresas (economía de las ideas) o con la acumulación de capital humano, por lo que esa tendencia no será lineal sino que presentará sus propias fluctuaciones.

→ Uno de los métodos más popularmente utilizados para separar ciclo de una tendencia no lineal es el filtro de HODRICK y PRESCOTT. El principal inconveniente de este método (y otros similares) es su carácter estrictamente estadístico y ateoórico<sup>8</sup>.

- En la práctica, debido a que las tendencias y los ciclos son componentes no observables, es necesario realizar un conjunto de *supuestos acerca de la naturaleza de la serie para el tratamiento de los datos*.

- La cuestión se complica aún más cuando realizamos una distinción entre *eliminación de la tendencia* y *filtrado* 2 conceptos que muchas veces han sido utilizados indistintamente dentro de la literatura:

- Eliminación de tendencia: Proceso a través del cual se busca convertir una serie temporal en estacionaria. La eliminación de la tendencia es necesaria si queremos calcular funciones de los segundos momentos de los datos, que pueden no existir si la serie no es estacionaria (p.ej. paseo aleatorio).
- Filtrado: Es un término mucho más amplio. En econometría, los filtros son operadores que eliminan ciertas frecuencias del espectro de datos. Por tanto, los filtros pueden ser usados para eliminar movimientos de baja frecuencia, para enfatizar la variabilidad de una frecuencia en particular o para suavizar la serie de movimientos idiosincráticos de alta frecuencia.

## 1.2. Componentes de la serie

### 1.2.0. Descomposición general de una serie: Marco de análisis

- Resulta útil representar los procesos de **series temporales** como casos particulares de una solución general de un **proceso estocástico en 3 partes**:

$$y_t = \underbrace{\mu_t}_{\text{tendencia}} + \underbrace{\psi_t}_{\text{componente estacional}} + \underbrace{\varepsilon_t}_{\text{ruido blanco}}$$

### 1.2.1. Componente tendencial

#### Relevancia de la tendencia

- La existencia o no de fluctuaciones alrededor de una tendencia de variables como el PIB resulta de gran importancia.
  - Por ejemplo, la Síntesis Neoclásica argumentaba que el crecimiento a largo plazo venía dado por factores exógenos de la oferta, mientras que las fluctuaciones se debían a factores de la demanda que debían ser estabilizados por el sector público.
  - Por otro lado, la Teoría del Ciclo Real argumentará que el ciclo no son fluctuaciones alrededor de la tendencia, sino que son shocks estocásticos tecnológicos sobre la propia tendencia. Autores como NELSON y PLOSSER (1982) encuentran que numerosas variables macroeconómicas tienen una raíz unitaria. Por tanto, los shocks no tendrán influencia temporal sobre la economía sino permanente.

<sup>8</sup> Por lo tanto, surge la cuestión de cómo valorar la bondad un determinado filtrado (en el caso del filtro HODRICK-PRESCOTT, cuál es el valor “correcto” del parámetro clave lambda). Y esta cuestión no es menor, pues el criterio de evaluación de los modernos modelos teóricos de ciclo va a ser precisamente el ajuste de las series generadas por los mismos a esas series de componente cíclico observadas en la realidad.

Las series resultantes del filtrado HODRICK-PRESCOTT cumple una serie de propiedades:

1. Un valle debe ser seguido por un pico, y un pico debe ser seguido por un valle;
2. La fase de expansión, así como la de contracción debe durar como mínimo 2 trimestres: este criterio implica que requerimos cierto grado de persistencia en el movimiento de la actividad económica;
3. Un ciclo completo de negocios debe contener al menos 5 trimestres: este criterio implica que las fluctuaciones que duran un año o menos no se consideran ciclos de negocios.



## Tipos de tendencias

- a) *Tendencia determinista*: Aquellos procesos cuya variación esperada en cada periodo contiene una expresión de un proceso no aleatorio tienen una tendencia determinista. Por ejemplo<sup>9</sup>:

$$y_t = y_0 + a_0 \cdot t + A(L) \cdot \varepsilon_t$$

- Debido a que en estos procesos sin tendencia estocástica los shocks tienen impacto temporal sobre la variable dependiente, de tal forma que las desviaciones con respecto a la tendencia son temporales, decimos que son estacionarios respecto a su tendencia.
- b) *Tendencia estocástica*: En aquellos procesos donde existe raíz unitaria, las diferentes innovaciones del proceso de ruido blanco van construyendo una tendencia cuyos efectos no se van diluyendo en el tiempo. Por tanto, presentarán una tendencia estocástica:

$$y_t = y_0 + \sum \varepsilon_t$$

## Descripción de procesos: Ejemplos de tendencias deterministas y estocásticas

- Podemos destacar los siguientes **ejemplos de procesos** para ilustrar la diferencia entre series estacionarias con tendencia determinista y series con tendencia estocástica (i.e. con raíz unitaria).

- a) *Ruido blanco*: Se dice que un proceso es puramente aleatorio si tiene media cero, una varianza constante y no está serialmente correlacionado. Por tanto, en estos procesos conocer los valores pasados no nos proporciona información alguna sobre el futuro, ya que el proceso no tiene memoria.

- b) *Paseo aleatorio*: El paseo aleatorio constituye el ejemplo de proceso no estacionario más conocido.

- *Paseo aleatorio sin deriva*: Es un proceso donde el valor presente es igual al valor pasado más un proceso de ruido blanco.

- En relación con su forma:

- El proceso se define como  $y_t = y_{t-1} + \varepsilon_t$  y puede ser descrito como  $y_T = y_0 + \sum_{t=0}^T \varepsilon_t$ .
- Si tomamos esperanzas:  $E[y_t] = y_0$ .
- Su varianza es:  $\text{var}[y_t] = t \cdot \sigma^2$ .

- En relación con sus características:

- Una de las características más importantes de estos procesos es la *persistencia de los shocks aleatorios*. Estos procesos tienen *memoria infinita*, lo que implica que el efecto de un shock no desaparece a lo largo del tiempo. A la suma de los términos error se le conoce como la tendencia estocástica.
- Debido a que estas series no son estocásticas, deberemos realizar un tratamiento previo de los datos antes de proceder a su estimación.

- Este proceso ha sido utilizado en la teoría económica para defender tesis como la *hipótesis de los mercados eficientes*: al analizar la evolución de las cotizaciones de los títulos de renta variable se constata que éstas siguen un paseo aleatorio. Autores como FAMA han argumentado que ello supone un ejemplo de la eficiencia de los mercados: las innovaciones se derivan de información que previamente no se encontraba disponible.

- *Paseo aleatorio con deriva*: Estos procesos, además de tener una tendencia estocástica, tienen una tendencia determinista.

- Por tanto, el proceso adopta la siguiente forma:

- El proceso puede ser definido como:  $y_t = \delta + y_{t-1} + \varepsilon_t$ .
- En estos procesos, la esperanza del mismo es igual a:  $E[y_t] = y_0 + t \cdot \delta$ .

- De nuevo, para analizar estos datos, será necesario eliminar el componente tendencial de la serie.

<sup>9</sup> Suponemos una serie que siempre cambia en la misma cantidad fija de un periodo a otro, es decir,  $\Delta y_t = a_0$ . La solución a esta ecuación diferencial lineal es:  $y_t = y_0 + a_0 \cdot t$ , es decir, una tendencia determinística con intercepto  $y_0$  y pendiente  $a_0$ . Añadiendo el componente estacionario  $A(L) \cdot \varepsilon_t$  a la tendencia se obtiene:  $y_t = y_0 + a_0 \cdot t + A(L) \cdot \varepsilon_t$ . Por lo tanto,  $y_t$  puede diferir de su tendencia en una cantidad  $A(L) \cdot \varepsilon_t$ . Dado que la desviación es estacionaria, la secuencia de  $y_t$  solo exhibirá desviaciones temporales de su tendencia. Por lo tanto, la predicción de  $y_{t+s}$  convergerá a la línea tendencial  $y_0 + a_0 \cdot (t + s)$ . Este tipo de modelo es denominado modelo con tendencia estacionaria.



- c) *AR(1)* (generalizable a *AR(p)*): Diremos que una serie  $y_t$  sigue un proceso autorregresivo de primer orden si ha sido generada por:  $y_t = c + \varphi \cdot y_{t-1} + \varepsilon_t$ , donde  $-1 < \varphi < 1$  y el componente  $\varepsilon_t$  es un proceso de ruido blanco.
- Estos procesos presentarán una media igual a:  

$$E[y_t] = E[y_{t-1}] \Rightarrow \mu = c + \varphi \cdot \mu \Rightarrow \mu = c/(1 - \varphi)$$
  - Para un proceso *AR(1)* con media cero ( $c = 0$ ), su varianza será igual a:  

$$E[y_t^2] = \varphi^2 \cdot E[y_{t-1}^2] + 2 \cdot \varphi \cdot E[y_{t-1} \cdot \varepsilon_t] + E[\varepsilon_t^2] \Rightarrow \sigma_y^2 = \varphi^2 \cdot \sigma_y^2 + \sigma^2 \Rightarrow \sigma_y^2 = \sigma^2/(1 - \varphi^2)$$
- d) *MA(1)* (generalizable a *MA(q)*): Se define el proceso de media móvil de orden 1 añadiendo a un proceso de ruido blanco una dependencia del valor actual de la serie de la última innovación ocurrida. De esta manera, el proceso será una combinación lineal de las 2 últimas innovaciones, de acuerdo con la ecuación:  $y_t = \varepsilon_t + \theta \cdot \varepsilon_{t-1}$ .
- Este proceso es la suma de 2 procesos estacionarios, por lo que el resultado siempre será estacionario.

- Por lo tanto, de este apartado inferimos que existen diferencias importantes entre las series temporales sin tendencia estocástica y las series con raíz unitaria.
  - Los shocks en series estacionarias (tengan o no tendencia) tendrán consecuencias temporales.
  - Sin embargo, en una serie con tendencia estocástica los shocks tendrán efectos permanentes.

### Eliminación del componente tendencial

- Los métodos usuales para eliminar tendencias son la *diferenciación* o la *eliminación de tendencias deterministas mediante modelos polinomiales*.

#### Diferenciación

- Tomemos un paseo aleatorio con deriva:

$$y_t = y_0 + \delta \cdot t + \sum \varepsilon_t + A(L) \cdot y_t + B(L) \cdot \varepsilon_t$$

- Si tomamos diferencias, obtenemos:

$$y_t - y_{t-1} = \Delta y_t = A(L) \cdot y_t + B(L) \cdot \varepsilon_t$$

- El resultado es una serie estacionaria que podrá ser estimada de manera consistente a través de un modelo *ARMA(p,q)*.

#### Modelos polinomiales para la eliminación de una tendencia determinista

- Sin embargo, en el caso de series estocásticas con tendencia el procedimiento de diferenciación para obtener una representación *ARMA* no siempre es posible, y puede ser necesario recurrir a **modelos polinomiales para la eliminación de una tendencia determinista**.

#### Identificación

- Como, habitualmente, el interés del investigador se centra únicamente en uno de los componentes, es útil reescribir la anterior ecuación como la suma de sólo 2 componentes. Así, tendremos:

$$y_t = \mu_t + \varepsilon_t$$

donde:

- $\mu_t$  es el nivel de la serie o componente tendencial, que se supone aquí una función conocida determinista del tiempo (depende del instante considerado  $t$  y de un vector de parámetros  $\beta$ ).
- $\varepsilon_t$  es la innovación o ruido, es decir, un componente aleatorio que recoge todos los demás efectos que actúan sobre la serie. Se supone que tiene una estructura estable a lo largo del tiempo (media cero, varianza constante y distribución normal e independencia entre periodos).
- Generalmente, el componente tendencial suele modelarse como un **modelo de tendencia polinómica**:  

$$\mu_t = f(t, \beta) = \beta_0 + \beta_1 \cdot t + \dots + \beta_r \cdot t^r$$
  - Tenemos, por tanto, una regresión en la que la variable endógena es la serie observada y las variables explicativas son distintas potencias del tiempo.
  - Por ejemplo, en los casos más sencillos:
    - Si  $r = 0$ , obtenemos una tendencia constante.
    - Si  $r = 1$ , obtenemos una tendencia lineal.
    - Si  $r = 2$ , obtenemos una tendencia parabólica.

- En la práctica, lo habitual es que  $r \leq 2$ . Supondremos ahora que estamos en el caso  $r = 1$  por sencillez expositiva.

### Estimación

- Las **estimaciones** se obtienen por el método de MCO (es decir, minimizando las diferencias entre los valores observados y los previstos a horizonte uno por el modelo). En definitiva, el problema consiste en encontrar el vector de estimadores  $\beta$  que minimiza la siguiente expresión:

$$SR = \sum_{t=1}^T (y_t - \mu_t)^2 = \sum_{t=1}^T (y_t - \hat{\beta}_0 - \hat{\beta}_1 \cdot t)^2$$

- CPO:

$$\frac{\partial SR}{\partial \beta_0} = 0 ; \quad \frac{\partial SR}{\partial \beta_1} = 0$$

- Suponiendo que la variable  $t$  está construida para tener media 0, tenemos que:

$$\hat{\beta}_0 = \bar{y}_t ; \quad \hat{\beta}_1 = \frac{\sum_{t=1}^T (y_t - \bar{y}_t) \cdot t}{\sum_{t=1}^T t^2}$$

donde  $\bar{y}_t$  es la media muestral.

### Predicción

- Con este modelo, la **predicción** en  $T$  de la serie para un periodo futuro  $T + k$ , se obtiene extrapolando el nivel de la serie, ya que la predicción óptima de la innovación es su esperanza (siempre nula):

$$y_t = \mu_t + \varepsilon_t$$

$$\hat{y}_T(k) = \mu_{T+k} + \underbrace{E[\varepsilon_t]}_{=0}$$

- Por tanto, la forma concreta del modelo determinista dependerá de cómo se modelice la tendencia.

### Limitaciones

- El ajuste de tendencias deterministas tiene varias **limitaciones**:
  - En la práctica, aunque las series con nivel constante ( $r = 0$ ) son frecuentes, es muy raro encontrar series reales con tendencia determinista polinómica con un orden superior a la unidad.
    - Una posibilidad sería intentar ajustar tendencias lineales por tramos (dividiendo la serie en tramos que tengan aproximadamente tendencia constante y ajustar cada tramo con un modelo distinto). Sin embargo, este enfoque es poco útil para la predicción, pues no sabemos cuántas observaciones pasadas utilizar para predecir el nivel futuro.
  - Otro inconveniente es que, como vemos por la expresión de  $\hat{\beta}_1$ , estas estimaciones atribuyen el mayor peso al crecimiento observado en el centro del periodo y el mínimo al crecimiento en los extremos.
    - Esta ponderación simétrica y centrada implica que la predicción da un peso mínimo al último dato observado (e igual al crecimiento más alejado en el tiempo). Esto es poco razonable en la mayoría de casos, donde suele existir una cierta “inercia”.

### 1.2.2. Componente estacional

#### Concepto, tipos y ejemplos

- Las **variaciones estacionales** se suelen definir como aquellas oscilaciones de periodo igual o inferior a un año generadoras de periodicidad regular, que pueden deberse, por ejemplo, a factores climatológicos o de calendario.
  - Así, en una serie temporal, pueden darse simultáneamente varios movimientos estacionales diferentes dependiendo del fenómeno estudiado y del periodo para el que tomemos los datos.
    - Por ejemplo:
      - Si consideramos el consumo de energía eléctrica en los hogares medido de forma mensual, observamos que los meses de otoño e invierno presentan un valor más alto que los de primavera y verano.

- Si estudiamos la misma serie, pero esta vez con datos horarios, veremos que surge un 2º movimiento estacional diario (con un máximo a últimas horas de la noche y un mínimo de madrugada).
- Por sencillez, nos limitaremos al caso en el que sólo existe un movimiento estacional, pero en realidad pueden presentarse 2 o más superpuestos.

### Relevancia

- El estudio del componente estacional es esencial en el análisis de coyuntura para captar el movimiento subyacente de la mayoría de variables económicas.
  - Por ejemplo, cuando se analizan la serie de la tasa de paro de la EPA, se observa que el dato del 3º trimestre es inferior en media al del resto de trimestres (debido a que, por la estructura económica española, se producen numerosas contrataciones temporales en verano, asociadas principalmente al turismo).
  - Por ello, si queremos evaluar cuál es la evolución del paro, no podemos comparar directamente el 3º trimestre con el 2º trimestre, pues ello conduciría a ser sistemáticamente optimistas.
  - Una alternativa habitual consiste en utilizar crecimientos interanuales (es decir, comparar con el mismo trimestre del año anterior).
    - Sin embargo, esto no deja de ser arbitrario, pues no deja de ser una ponderación particular del crecimiento intertrimestral de los últimos 4 trimestres:

$$(y_t - y_{t-4}) = (y_t - y_{t-1}) + (y_{t-1} - y_{t-2}) + (y_{t-2} - y_{t-3}) + (y_{t-3} - y_{t-4})$$

- Por tanto, volviendo a la ecuación antes expuesta, los datos observados se generan como suma de 3 factores:

$$y_t = \underbrace{\mu_t}_{\text{tendencia}} + \underbrace{\psi_t}_{\text{componente estacional}} + \underbrace{\varepsilon_t}_{\text{ruido blanco}}$$

donde  $\psi_t$  es el componente estacional y será necesario eliminarlo.

### Eliminación del componente estacional

- Los métodos clásicos suponen que tanto tendencia como estacionalidad son deterministas.
  - La estacionalidad se modeliza como una función periódica, que verifica la condición:

$$\psi_t = \psi_{t-s}$$

donde  $s$  es el periodo de la función, que depende del tipo de estacionalidad de los datos.

- Por ejemplo, una serie mensual con estacionalidad anual tiene periodo  $s = 12$  meses. Por tanto, los coeficientes estacionales se repetirán cada 12 meses.
- Suponiendo que la serie es mensual y la estacionalidad anual, el procedimiento para estimar el componente estacional determinista es el siguiente:

- Estimar la tendencia de la serie según el método antes explicado.
- Restar la tendencia de la serie original para obtener la “serie sin tendencia”:

$$y_t^* = y_t - \mu_t = \psi_t + \varepsilon_t$$

- A partir de esta serie sin tendencia, estimar los componentes estacionales  $\psi_1, \dots, \psi_{12}$  como la diferencia entre la media de los periodos estacionales y la media general. Por ejemplo, el componente estacional de junio ( $\psi_6$ ) sería la diferencia entre:

- La media de la serie sin tendencia para todos los meses de junio.
- La media general de toda la serie sin tendencia.

Al final, se puede comprobar que  $\sum_{t=1}^{12} \psi_t = 0$ .

- Otra alternativa para modelizar la estacionalidad es representarla a través de una función armónica de periodo  $s$ , como por ejemplo el seno o el coseno. En este sentido, FOURIER (matemático francés) demostró a principios del s. XIX que toda función periódica puede representarse como suma de funciones sinusoidales de distinta amplitud y frecuencia.

### 1.2.3. Ciclo

- Aunque existen también modelos que tratan de modelizar los **ciclos** como algo *determinista* (a través de funciones sinusoidales), este enfoque está hoy en día desfasado.
  - La interpretación más habitual es que el **componente cíclico** es simplemente el *residuo obtenido una vez se han eliminado de la serie los componentes tendencial y estacional*.

### 1.3. Valoración

- En conclusión, los **modelos deterministas**:
  - Pueden ser *útiles para la descripción simple* de las pautas de una serie.
  - Sin embargo, las *predicciones que proporcionan suelen tener un alto error*. Esto se debe a que, en una serie temporal, la observación hoy ( $y_t$ ) suele depender en general de sus valores previos ( $y_{t-1}, y_{t-2}, \dots$ ) pero esta dependencia suele ser más fuerte con los datos más recientes y más débil con los datos más alejados.
- Como las predicciones de los modelos de deterministas no utilizan esta propiedad, se propusieron como alternativa los *métodos de alisado*.

## 2. MÉTODOS DE FILTRADO AD-HOC

### 2.1. Concepto

- La idea de estos métodos es *permitir que los últimos datos de la serie tengan mayor peso* en las predicciones que los más antiguos. *Esto se logra permitiendo que los parámetros del modelo de tendencias deterministas no sean constantes, sino que puedan ir variando en el tiempo*.
  - Estos métodos se introdujeron en los *años 60* y se extendieron rápidamente por sus buenos resultados y facilidad de cálculo con las nuevas técnicas de computación de la época. Las primeras aproximaciones vienen de la mano de modelos como el de *alisado simple de HOLT* (1956).
  - Posteriormente, se desarrollan *modelos más finos* basados en la idea de *media móvil*.
    - En concreto, los 2 filtros ad-hoc más conocidos son el *filtro X11* y el *filtro de Hodrick-Prescott* (HP). Se trata de filtros de media móvil centrados y simétricos.
- Veamos cómo funciona el filtro de HODRICK y PRESCOTT (HP) propuesto por estos 2 autores en 1980.

### 2.2. Ejemplo: filtro Hodrick-Prescott

- Este filtro determina el componente cíclico ( $y_t - \mu_t$ ) y el tendencial ( $\mu_t$ ) de una serie desestacionalizada<sup>10</sup>  $y_t$ , minimizando la siguiente expresión:

$$\min_{\{\mu_t\}} HP = \sum_{t=1}^T \underbrace{(y_t - \mu_t)^2}_{\substack{\text{Componente} \\ \text{cíclico} \\ \text{elevado al cuadrado}}} + \lambda \cdot \sum_{t=2}^{T-1} \underbrace{\left[ (\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}) \right]^2}_{\substack{\text{Lo que varía la tasa de crecimiento} \\ \text{elevado al cuadrado}}}$$

donde  $y_t$  es la serie temporal observada desestacionalizada,  $\mu_t$  es el componente tendencial estimado y  $\lambda$  es un parámetro que mide la sensibilidad de la tendencia a las fluctuaciones a corto plazo (cuanto menor sea, más se ajusta la tendencia a la serie).

- El método consiste en minimizar la expresión anterior. Por lo tanto, buscamos minimizar una combinación de:
  - La distancia entre la serie y el componente tendencial y por ende el **componente cíclico elevado al cuadrado** (*primer sumatorio*); y
  - La **variación de la tasa de crecimiento tendencial elevado al cuadrado** (*segundo sumatorio*).

<sup>10</sup> El componente estacional se suele eliminar en primer lugar de modo que las series queden limpias de las fluctuaciones periódicas que se producen dentro del año. Este paso suele realizarse mediante técnicas estadísticas avanzadas.

Una de las herramientas más utilizadas internacionalmente son los programas TRAMO (*Time series Regression with ARIMA noise, Missing values and Outliers*) y SEATS (*Signal Extraction in ARIMA Time Series*), desarrollados por GÓMEZ y MARAVALL, del Banco de España.

- Dado que estos 2 objetivos están contrapuestos, el parámetro lambda,  $\lambda$ , es elegido por el analista para ponderar la importancia de cada uno. Para entenderlo mejor veamos 2 casos extremos:
  - Si  $\lambda = 0$ , la expresión a minimizar sería:

$$\min_{\{\mu_t\}} HP = \sum_{t=1}^T \underbrace{(y_t - \mu_t)^2}_{\substack{\text{Componente} \\ \text{cíclico} \\ \text{elevado al cuadrado}}} \Rightarrow \begin{cases} y_t = \mu_t \\ y_t - \mu_t = 0 \end{cases}$$

Es decir, se minimiza el componente cíclico elevado al cuadrado, que será igual a cero, y toda la serie será tendencial.

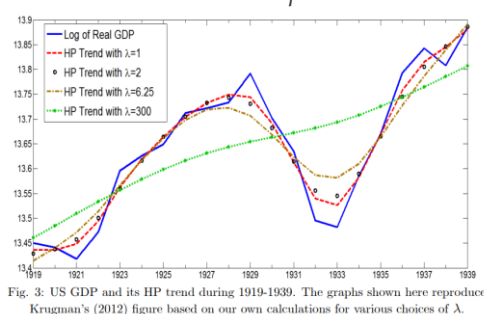
- Si  $\lambda = +\infty$ , la expresión a minimizar sería:

$$\min_{\{\mu_t\}} HP = \sum_{t=2}^{T-1} \underbrace{\left[ (\mu_{t+1} - \mu_t) - (\mu_t - \mu_{t-1}) \right]^2}_{\substack{\text{Lo que varía la tasa de crecimiento} \\ \text{elevado al cuadrado}}}$$

Es decir, se minimiza la variación del componente tendencial y este será, por lo tanto, lineal.

- La elección de  $\lambda$  es aleatoria, pero HODRICK y PRESCOTT estiman que, en la economía estadounidense, para datos trimestrales, un valor de  $\lambda = 1.600$  es razonable, bajo el supuesto de que cualquier perturbación que tiene efectos durante 8 o más años tiene carácter permanente. Para series mensuales se suele utilizar  $\lambda = 14.400$  y para series anuales se recomienda un valor  $\lambda = 100$ .
  - Se trata de convenciones basadas en la evidencia de que ese valor produce resultados razonables.

IMAGEN 1.– Filtro de Hodrick-Prescott para distintos valores de Lambda

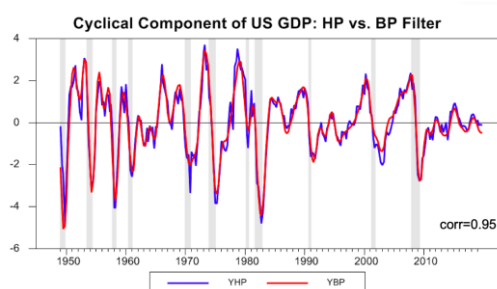


Fuente: Phillips, P. C. B. & Jin, S. (2021). Business Cycles, Trend Elimination, and the Hp Filter. *International Economic Review*, 62(2), 469-520.

<https://doi.org/10.1111/iere.12494>

- Finalmente, cabe mencionar el método *Band-Pass Filter*<sup>11</sup> que permite eliminar el componente tendencial (como el filtro HP) y además elimina las fluctuaciones de muy corto plazo, de modo que elige los coeficientes para que la variable filtrada capture sólo las fluctuaciones que tienen lugar en un intervalo determinado. En la práctica, el intervalo que se considera está comprendido entre 6 y 32 cuatrimestres.

IMAGEN 2.– Componente cíclico del PIB de EEUU (Hodrick-Prescott vs Band-Pass)



Fuente:

<sup>11</sup> [https://read.oecd-ilibrary.org/economics/a-note-on-band-pass-filters-based-on-the-hodrick-prescott-filter-and-the-oecd-system-of-composite-leading-indicators\\_jbcma-2011-5kg0pb01sbbt#page1](https://read.oecd-ilibrary.org/economics/a-note-on-band-pass-filters-based-on-the-hodrick-prescott-filter-and-the-oecd-system-of-composite-leading-indicators_jbcma-2011-5kg0pb01sbbt#page1)  
<https://www.jstor.org/stable/2646708>

### 2.3. Valoración de los filtros ad-hoc

#### Ventajas

- Simplifica la tarea de definir un componente: El componente es simplemente el resultado del filtro.
- Son sencillos y fáciles de usar. Ésta es una propiedad importante cuando se necesita estimar los componentes de un gran número de series muy heterogéneas, como sucede por ejemplo cuando las agencias estadísticas presentan datos desestacionalizados. Sin embargo, en la investigación de economía aplicada, cuando se trata de estudiar pocas series y cuidando aspectos metodológicos, esto no es necesariamente conveniente.

#### Limitaciones

- Problema del punto final: Los filtros simétricos (*two-sided*) no permiten estimar el componente  $\mu_t$  cuando  $t$  está cerca de los extremos de la serie (es decir, al principio o al final). Una solución propuesta consiste en truncar el filtro para esas observaciones de “punto final” con pesos *ad-hoc*. Así, para estimaciones recientes, habrá que utilizar filtros asimétricos que utilicen dichos pesos. A medida que pase el tiempo, y se vayan haciendo disponibles nuevas observaciones, se podrán recalcular estas estimaciones preliminares de los componentes con el filtro simétrico habitual. El problema es que, según simulaciones realizadas por MARAVALL, cuando utilizamos el filtro HP, las revisiones de la tendencia actual pueden tardar en estabilizarse hasta 10 años en el futuro. Esto hace poco práctico el filtro para el análisis de coyuntura, en el que queremos conocer rápidamente cómo nos situamos respecto a la tendencia. En definitiva, el filtro HP tiene periodos de revisión muy largos.
  - Riesgo de regresión espuria: Estos filtros son mecánicos y, por ello, pueden fácilmente encontrar tendencias y ciclos aunque el proceso estocástico que ha generado la serie sea en realidad un paseo aleatorio.
  - Ausencia de un modelo teórico: El hecho de que estos métodos carezcan de un modelo probabilístico subyacente hace más difícil saber en qué casos su uso es o no adecuado. En este sentido, por ejemplo, estos filtros no proveen datos sobre las propiedades del estimador (varianza, consistencia, etc.), lo que limita también su utilidad para labores de inferencia estadística y predicción.
- Para superar las limitaciones de esta “caja negra” de los filtros ad-hoc, en las 2 últimas décadas han surgido nuevos enfoques para la estimación de los componentes no observables, basados directamente en la estimación de modelos estocásticos del tipo ARIMA (*Autoregressive Integrated Moving Average*).

### 3. MÉTODOS MODERNOS DE COMPONENTES NO OBSERVABLES (BASADOS EN MODELOS)

- Los **modelos modernos de componentes no observables** (CNO) consideran que cada uno de los componentes sigue en realidad un proceso estocástico distinto y que, cuando todos ellos se agregan, dan lugar a la serie efectivamente observada.
- Hay que señalar que no existe una definición universal sobre la definición de una tendencia, un ciclo o un componente estacional. Por tanto, no hay tampoco una única forma de modelizar los CNO. Cuando se especifica el modelo general, existen, básicamente, 2 opciones:
  - Enfoque de Series Temporales Estructurales (STS, por sus siglas en inglés): Consiste en especificar directamente un modelo para cada componente basándose en conocimientos a priori sobre dicho componente. Es el enfoque más usado en la econometría aplicada.
  - Enfoque basado en modelos ARIMA (AMB): Como sólo existen observaciones sobre la serie agregada, la idea es proceder en 2 pasos:
    - Primero, estimar un modelo ARIMA para la serie observada.



- Segundo, derivar modelos para los CNO que, cuando sean agregados, sean compatibles con el primer modelo.

### 3.1. Identificación: caracterización de los componentes

- Es interesante revisar las principales especificaciones de los CNO más comunes, es decir: *tendencia, ciclo y estacionalidad*.

#### 3.1.1. Tendencia

- Consideremos una **tendencia determinista** lineal:

$$\mu_t = \beta_0 + \beta_1 \cdot t$$

- Entonces, tenemos que las 2 primeras diferencias de la tendencia son constantes:

$$\nabla \mu_t = \mu_t - \mu_{t-1} = \beta_1$$

$$\nabla^2 \mu_t = 0$$

- En palabras: “en un modelo determinista, la variación del componente tendencial de la serie es igual en todos los periodos y, por tanto, la variación de los cambios es 0”.

- Una **tendencia estocástica** no satisface estas últimas condiciones de manera exacta para cada periodo, sino que se asume que están perturbadas en cada periodo por un shock de media 0 y pequeña varianza. Por ejemplo, tendríamos que:

$$\nabla \mu_t = \mu_t - \mu_{t-1} = \beta_1 + u_t, \text{ con } u_t \text{ siendo ruido blanco iid}$$

- Es decir, con esta especificación, la tendencia sigue un paseo aleatorio con deriva (donde la deriva  $\beta_1$  es la “pendiente” de la tendencia).
- Siguiendo el enfoque STS de HARVEY y TODD (1983), podemos incluir un *grado adicional de complejidad*, haciendo que la pendiente también sea estocástica. Tendríamos entonces que:

$$\beta_1 = a_t, \text{ con } a_t \text{ siendo ruido blanco iid}$$

por lo que:  $\nabla^2 \mu_t = a_t$

- Generalizando, las tendencias estocásticas pueden modelizarse haciendo que la 1ª y la 2ª diferencia de la tendencia ( $\nabla^d \mu_t$ , siendo  $d$  igual a 1 o 2) siga algún proceso ARMA de orden bajo.

#### 3.1.2. Estacionalidad (ARIMA estacional multiplicativo)

- Un enfoque habitual en presencia de estacionalidad consiste en modelar de forma separada la dependencia regular y la estacional y construir el modelo incorporando ambas de forma multiplicativa. Se obtiene así el modelo estacional multiplicativo  $ARIMA(P,D,Q) \times (p,d,q)$ , que tiene la siguiente forma:

$$\Phi_P(L^s) \cdot \varphi_p(L) \cdot \nabla_s^D \cdot \nabla^d \cdot y_t = \Theta_Q(L^s) \cdot \theta_q(L) \cdot \varepsilon_t$$

- Suponiendo que tenemos datos mensuales y estacionalidad anual ( $s = 12$ ), la idea es la siguiente:

- Primero, estimar los modelos ARIMA para las 12 series anuales constituidas por los meses de enero, febrero... diciembre. Si suponemos que estas series tienen el mismo modelo común, entonces, para todo  $t$ :

$$(1 - \Phi_1 \cdot L^{12} - \dots - \Phi_P \cdot L^{12 \cdot P}) \cdot y_t = (1 - \Theta_1 \cdot L^{12} - \dots - \Theta_Q \cdot L^{12 \cdot Q}) \cdot u_t$$

- Segundo, como no se está teniendo en cuenta la dependencia entre un dato y el inmediatamente anterior, la estructura del error  $u_t$  no será un ruido blanco, por lo que podemos construir otro modelo ARIMA para él:

$$\varphi_p(L) \cdot \nabla^d \cdot u_t = \theta_q(L) \cdot \varepsilon_t$$

- Tercero, sustituyendo este modelo en el anterior, obtendríamos el modelo ARIMA estacional multiplicativo.

### 3.1.3. Combinación con el ciclo estocástico

- La tendencia estocástica puede ser combinada con un ciclo estocástico para conseguir un modelo como el siguiente:

$$y_t = \mu_t + \psi_t + \varepsilon_t$$

donde  $\psi_t$  es el componente cíclico. Y puede expresarse:

...

### 3.2. Estimación: filtro de Kalman

- El **filtro de Kalman** es un algoritmo sencillo de programar, eficiente computacionalmente, muy utilizado en la metodología STS tanto para la estimación como para la predicción.
  - El filtro de Kalman comienza *reformulando el modelo en un formato de estado-espacio*. Corre una serie de iteraciones tras establecer condiciones iniciales adecuadas.
    - En concreto, calcula el estimador que *minimiza el error cuadrático medio de predicción (ECMP) del vector estado*, basándose en observaciones pasadas.
  - A continuación, *calcula los errores de predicción, lo que permite construir la función de verosimilitud del modelo*. Esto se realiza recursivamente.
- Otra aproximación similar es el **filtro Wiener-Kolmogórov (WK)**.

### 3.3. Predicción

- La ventaja de los modelos de CNO basados en modelos es que permiten aplicar técnicas de **predicción** de modelos ARIMA a cada uno de los componentes estimados.
  - De este modo, teniendo en cuenta que conocemos las propiedades de los estimadores, podremos realizar intervalos de predicción más precisos que con los filtros ad-hoc.

## 4. APLICACIONES PRÁCTICAS

### 4.1. Output gap y saldo ajustado al ciclo

#### 4.1.1. Idea: medición del efecto macroeconómico del presupuesto

- Los efectos de la política fiscal sobre la demanda se deben tanto a la *política fiscal discrecional* como a los *estabilizadores automáticos*.
  - No obstante, a nivel empírico, la diferenciación y medición de ambos resulta una cuestión compleja, por lo que es necesario desarrollar un marco metodológico para la medición del efecto macroeconómico de ambos tipos de política fiscal [ver tema 4.B.5].
  - A continuación analizaremos la medición del denominado *efecto macroeconómico del presupuesto*, el cual ha cobrado gran importancia a raíz de la LOEPSF.
  - La **medición del efecto macroeconómico del presupuesto** trata de determinar la **orientación/tono** de la política fiscal, que viene dada exclusivamente por las actuaciones discrecionales, esto es, por el **saldo presupuestario estructural**, que no es directamente observable.
    - Por lo tanto, tenemos que llevar a cabo un ejercicio de descomposición del saldo presupuestario observable en *saldo cíclico* (i.e. efecto de los estabilizadores automáticos) y *saldo estructural*.

#### 4.1.2. Descomposición del déficit público

- El déficit público puede descomponerse en varios componentes. Concretamente, el déficit total ( $B$ ) es igual al déficit cíclico ( $DC$ ) más el déficit tendencial ( $DT$ ) más el déficit discrecional o impulso fiscal ( $DD$ ):

$$B_t = \underbrace{DC_t}_{\text{Déficit cíclico}} + \underbrace{DT_t}_{\text{Déficit tendencial}} + \underbrace{DD_t}_{\text{Déficit discrecional}}$$

*Déficit estructural*

- Déficit cíclico*:
  - Es el producido por oscilaciones del PIB respecto al PIB potencial.
  - Ejemplo: Aumento de las prestaciones por desempleo durante una fase de recesión.

- *Déficit tendencial*:
  - Es el producido por cambios en la tendencia del PIB potencial.
  - Ejemplo: Aumento de las prestaciones por desempleo en respuesta a un aumento del desempleo estructural.
- *Déficit discrecional*:
  - Es el producido por decisiones expresas de política fiscal de las autoridades.
  - Ejemplo: Aumento del consumo público o de la inversión pública.
- La suma del déficit tendencial y el discrecional, restándole las medidas *one-off*, es lo que se conoce como *déficit estructural*, que es la medida que se suele utilizar para determinar la orientación/tono de la política fiscal del Gobierno y, comparándola con el PIB, su carácter estabilizador o intensificador del ciclo económico.
  - ¿Cómo podemos estimar dicho déficit estructural?

#### 4.1.3. Cálculo del déficit estructural por el método indirecto

- Distinguimos **2 métodos** para la obtención del déficit estructural:
  - *Método directo*: Consiste en medir directamente el saldo estructural calculando el impacto sobre el presupuesto de las decisiones discrecionales de política fiscal.
    - Es poco utilizado por la complejidad para determinar qué ingresos y gastos son discrecionales y cuáles no.
  - *Método indirecto*: En vez de calcular el saldo estructural, se calcula el saldo cíclico, y se le resta al total, obteniéndose así el saldo cíclicamente ajustado como residuo. A dicho saldo cíclicamente ajustado se le restan las medidas *one-off* para obtener el déficit estructural.
- En la práctica, en España, se utiliza el **método indirecto**, debido a la dificultad de determinar qué parte de los ingresos y los gastos son directamente el resultado de las decisiones del sector público. La metodología seguida es la aplicada por la *Comisión Europea* y acordada en el seno del Grupo de Trabajo del Output Gap. El procedimiento es el siguiente:
  1. Definición de un nivel de actividad, unos gastos y unos ingresos de referencia.
  2. Asignación de elasticidades para asociar variaciones de los ingresos y gastos a las variaciones de la renta.

### 1) Estimación de los niveles potenciales de actividad

- Como consecuencia del comportamiento cíclico de las principales variables macroeconómicas, la teoría macroeconómica ha buscado descomponer el PIB observado en 2 componentes:
  1. *PIB potencial*: Evolución a largo plazo vinculada a factores asociados a la OA (tecnología, nivel y composición de la fuerza laboral).
  2. *Componente cíclico*: Evolución a corto plazo y medio plazo caracterizada por las fluctuaciones recurrentes, aunque no necesariamente periódicas, en torno al potencial. Este componente cíclico se corresponderá con la brecha de producción, output gap (OG), de tal forma que un OG positivo se interpretará como una situación de excesiva demanda y sobrecalentamiento de la economía, mientras que un OG negativo se asociará con una demanda insuficiente para alcanzar el pleno empleo de los recursos.
- Desarrollaremos el método utilizado por la Comisión Europea, usado como referencia en la actual LO 2/2012 de Estabilidad Presupuestaria y Sostenibilidad Financiera:
  1. Estimación de la función de producción y del nivel de actividad potencial: Se parte de una función de producción Cobb-Douglas donde el único elemento a estimar es la productividad total de los factores (PTF), que podrá ser derivada a través de una reorganización de los términos:

$$Y = A \cdot L^{0,65} \cdot K^{0,35}$$

siendo  $L$  representa las horas trabajadas.

A partir de la ecuación anterior, se estima el nivel de actividad potencial a través de los niveles potenciales de cada componente:

$$Y_{pot} = A_{pot} \cdot L_{pot}^{0,65} \cdot K_{pot}^{0,35}$$

- Estimación de las fluctuaciones en la PTF potencial:** Se asume que el componente cíclico de la PTF es un factor inobservable que sigue un proceso estocástico, donde el término error seguirá un proceso AR(1) (autorregresivo de primer orden).
- Evolución del mercado potencial de trabajo:** El nivel potencial de empleo ( $L_{pot}$ ) será igual al resultado de multiplicar el número de trabajadores en edad de trabajar (POB) por la tasa de participación potencial ( $ACT_{pot}$ ), por el número de horas potenciales trabajadas al año por trabajador ( $H_{pot}$ )<sup>12</sup> y la tasa de desempleo no aceleradora de la inflación salarial, NAWRU ( $\mu_{NAWRU}$ )<sup>13</sup>. Expresado en términos matemáticos obtenemos:

$$L_{pot} = POB \cdot ACT_{pot} \cdot (1 - \mu_{NAWRU}) \cdot H_{pot}$$

- Cabe destacar que la NAWRU no es el nivel de pleno empleo, sino el nivel de empleo compatible con la estabilidad de la inflación salarial.
  - También se incluye una curva de Phillips, que es en ocasiones criticada por su enfoque backward-looking, en vez del característico comportamiento forward-looking consistente con la literatura macroeconómica actual.
- Estimación del stock de capital potencial:** El stock de capital potencial se corresponde con el capital observado, aplicando la ecuación dinámica de acumulación del capital.

$$K_{e,t} = K_{t-1} \cdot (1 - \delta) + I_t$$

- Obtención del output gap (que representará el componente cíclico),** que puede ser estimado de la siguiente forma:

$$Y = Y_e + Y_c \Rightarrow \widetilde{OG} = \frac{Y_c/Y_e}{Y_e} \cdot 100$$

## 2) Obtención del saldo presupuestario ajustado al ciclo económico y el saldo estructural

- Una vez descompuesta la serie del PIB observado en sus componentes tendencial y potencial es posible *descomponer el saldo presupuestario observado* en una parte vinculada al ciclo y, como residuo, el componente estructural, independiente de las fluctuaciones a corto plazo y medio plazo de la actividad económica:
  1. Como paso preliminar, descontamos del saldo presupuestario total el gasto vinculado con el pago de intereses de la deuda. A esto se le denomina el *saldo presupuestario primario (SPP)*.
  2. A partir del SPP, podemos derivar el *saldo ajustado al ciclo (SAC)* restando al SPP el output gap multiplicado por la semielasticidad del saldo presupuestario primario respecto al OG ( $\eta$ )<sup>14</sup>. De esta forma obtenemos:

$$SAC = SPP - \eta \cdot OG$$

<sup>12</sup> La tasa de actividad potencial y el número de horas potenciales trabajadas al año por trabajador se obtienen aplicando un filtro Hodrick-Prescott a las tasas observadas.

<sup>13</sup> A su vez, para medir  $\mu_{NAWRU}$ , la Comisión Europea:

- Emplea una *curva de Phillips*.
- Busca una *especificación econométrica* que regresa la inflación salarial (definida como la variación periodo sobre periodo de los CLU) sobre (i) el componente cíclico de la tasa de desempleo (a veces con lags) y (ii) un vector de variables exógenas que actúan como factores de control para identificar desde un punto de vista estadístico la relación entre inflación salarial y desempleo cíclico. Las variables exógenas son específicas de cada país y para el caso de España son: el nivel de productividad del trabajo y su variación, el nivel de la relación real de intercambio y su variación y el nivel de participación de las rentas salariales en la renta nacional.
- El modelo se completa especificando la dinámica de la tasa de desempleo, tanto de su componente estructural como del cíclico. Para ello, se supone un modelo cuyo componente estructural es un paseo aleatorio. Este componente estructural o permanente del desempleo es el que se aplica para calcular el empleo potencial de la economía.

<sup>14</sup> La sensibilidad al ciclo económico del presupuesto se define como:

$$\underset{\text{Sensibilidad al ciclo económico}}{\eta} = \left( \sum_i T_{i,t} \cdot \epsilon_i - \sum_j G_{j,t} \cdot \epsilon_j \right)$$

3. Finalmente, otra corrección es descontar las medidas de políticas fiscales de carácter puntual con impacto en un solo ejercicio, como las ayudas a la reestructuración del sistema financiero, o ingresos obtenidos en amnistías fiscales. Estas son las medidas *one-off*. Como resultado obtenemos el *saldo estructural*.

- DEROOSE *et al* (2008) estiman la capacidad amortiguadora de los estabilizadores de diferentes economías. Los de la Eurozona tienen una capacidad estabilizadora de 0,48, frente a la capacidad menor de EEUU, de 0,34. La capacidad amortiguadora de los estabilizadores de España es inferior a la media de la UE, 0,39.

#### 4.2. Otros

- Los 2 enfoques más utilizados para ajustar de estacionalidad las series temporales son el ajuste estacional basado en modelos y el basado en métodos no paramétricos:
  - El ajuste estacional basado en modelos consiste en los siguientes pasos:
    - Primero, se construye un modelo para la serie que se quiere ajustar.
    - Seguidamente, se obtienen a partir de él otros modelos para los componentes; se emplea el filtro de Wiener-Kolmogórov para separarlos.
    - Finalmente, se vuelven a agregar excluyendo el componente estacional.
- La implementación más extendida de este método es el paquete TRAMO-SEATS, desarrollado por el BdE (A. MARAVALL). Este paquete incorpora 2 módulos:
  - TRAMO (*Time series with ARIMA-based Missing observations and Outliers*): Programa de identificación automática, estimación y verificación del modelo.
  - SEATS (*Signal Extraction in ARIMA Time Series*): Programa que lleva a cabo la separación de los componentes.
- Los métodos no paramétricos permiten descomponer la serie en CNO mediante un procedimiento iterativo basado en alisados sucesivos. Estos alisados se obtienen aplicando medias móviles.
  - Los programas de la familia del X-11, del *Bureau of Census de EE.UU.* son los que se emplean más a menudo para hacer este tipo de ajuste. A partir del X-12-ARIMA, se extienden las series mediante predicciones hechas con modelos ARIMA. De esta forma se reducen las revisiones al final de la serie.
- En la actualidad, se han desarrollado programas que permiten hacer el ajuste mediante los 2 enfoques (p.ej. el X-13 ARIMA-SEATS).
- Con respecto a su aplicación a las estadísticas españolas:
  - INE: En la actualidad, el estándar del INE es el ajuste estacional basado en modelos utilizando para ello la versión más actualizada de TRAMO-SEATS. Así, desde 2013, el INE ofrece:
    - Datos corregidos de efectos calendario y efectos estacionales.
    - Datos de paro desestacionalizados.
    - Inflación subyacente.
  - SEPE: Datos de paro registrado, Afiliaciones a la SS.

## CONCLUSIÓN

donde:

- $T_{i,t}$  representa el porcentaje de peso en el PIB nominal de cada uno de los 4 ingresos impositivos (IRPF, IS, indirectos y cotizaciones sociales) del año  $t$ .
- $\varepsilon_i$  es la elasticidad de cada tipo de impuesto respecto a la brecha de producción.
- $G_{j,t}$  representa el gasto en desempleo en porcentaje del PIB.
- $\varepsilon_j$  es la elasticidad respecto a la brecha de producción.

