

# TEMA 1

# Etapas del Proceso Científico.

# **Las etapas del proceso científico son cuatro:**

## **Primera, Formulación de premisas o supuestos.**

Un supuesto es un enunciado de carácter universal que establece la pauta de comportamiento supuestamente seguida por los agentes cuyo comportamiento se pretende explicar.

Es el punto de arranque de todo proceso de elaboración científica y en él se concreta el contenido que se transmitirá a otros enunciados en las siguientes etapas del proceso.

## **Segunda, Derivación de conclusiones.**

Las conclusiones se obtienen a partir de los supuestos mediante la aplicación de procedimientos deductivos o inferencias demostrativas.

Las conclusiones son enunciados de carácter universal y su contenido es el mismo que el de los supuestos aunque en una versión alternativa que las hace más útiles para el objetivo que se persigue.

## **Tercera, Obtención de Predicciones y Explicaciones.**

Las predicciones y explicaciones son enunciados de carácter singular que nos informan, para un tiempo y lugar concretos, sobre lo que la teoría dice acerca de cómo funciona la realidad.

Ambas se obtienen combinando las conclusiones con las llamadas condiciones iniciales o enunciados auxiliares que son enunciados singulares y que nos concretan el marco en el cual se supone que va a aplicarse la teoría.

# Validación según el enfoque de Mill-Robbins.

Para los autores en esta corriente, la ciencia económica descansa esencialmente en unas pocas proposiciones generales que son el resultado de la observación o la introspección y que cualquier hombre, tan pronto como oye de ellas, las admite como algo familiar, a partir de las cuales se derivan las conclusiones que serán verdad en ausencia de causas perturbadoras.

Podemos decir que, en Economía, tenemos acceso directo a los elementos últimos de nuestras generalizaciones.

Las teorías son deducciones a partir de una serie de postulados.

Y la mayor parte de estos postulados son todos ellos hipótesis con un contenido simple e indiscutible sobre la experiencia diaria acerca de cómo se administra la escasez de los bienes económicos.

No necesitamos ni experimentos controlados ni procedimientos de contraste de tipo estadístico-econométrico para establecer su validez: son algo tan familiar en nuestra experiencia diaria que basta que sean enunciados para que sean reconocidos como obvios.

En cuanto a los seguidores de esta línea, además del propio John Stuart Mill (1806–1873), destacan otros economistas del siglo XIX, como Ricardo, Cairnes y Senior, que también pertenecen a esta escuela.

El autor más importante de esta corriente fue Robbins quien, en 1932, escribió la obra “Un Ensayo sobre la Naturaleza y Significado de la Ciencia Económica”.

En fechas recientes los dos autores que han realizado aportaciones dentro de esta tradición han sido Hausman y Cartwright.

# Validación según el enfoque de Hutchison-Friedman.

Para los autores en esta línea, la validación de un sistema teórico hay que hacerla mediante el contraste empírico de alguna parte de ese sistema utilizando procedimientos de contraste de tipo estadístico-econométrico que sean replicables.

No vale la introspección ni la evidencia inmediata.

Hay que huir del componente subjetivo y tratar de utilizar procedimientos objetivos.

Hutchison escribe en 1938, con 28 años, un libro titulado “El Significado y los Postulados Básicos de la Teoría Económica” que es considerado por muchos como la primera formalización de esta segunda tradición.

Milton Friedman escribe en 1953 un trabajo titulado “La Metodología de la Economía Positiva” que es, sin duda, el trabajo de metodología más citado del siglo XX.

# Proceso Estacionario En Sentido Débil.

Decimos que una variable es estacionaria en sentido débil cuando se cumple lo siguiente:

$$E(y_t) = \mu \quad \forall t$$

$$Var(y_t) = \sigma^2 \quad \forall t$$

$$Cov(y_t, y_s) = \gamma_{t-s} \quad \forall t, s$$

Lo que significa esta caracterización de la estacionariedad es que todos los elementos del proceso, con independencia del periodo temporal en que nos encontremos, giran en torno a un mismo valor, con una dispersión en torno a ese valor medio que no cambia con el tiempo y con una dependencia temporal que tampoco cambia.

# Variable Integrada de orden 1.

Se dice que una serie es integrada de orden 1, y la denotaremos por  $y_t \sim I(1)$ , cuando no tiene ningún componente determinista y después de ser diferenciada una vez resulta una representación ARMA estacionaria e invertible.

# Tendencia Estocástica.

La tendencia estocástica es la suma de tantos ruidos blancos como indica el subíndice del elemento del proceso, es decir,  $y_t = \sum_{i=1}^t \varepsilon_i$ , en donde  $\varepsilon_i$ ,  $i = 1, 2, \dots, t$  son ruidos blancos.

Equivale a decir que el proceso es un paseo aleatorio sin deriva y con un valor de origen igual a cero; es decir

$$y_t = y_{t-1} + \varepsilon_t$$

$$y_o = 0$$



# Contraste de Dickey y Fuller.

Es un procedimiento que se utiliza para rechazar, en su caso, la hipótesis nula de que la serie tiene una raíz unitaria o, equivalentemente, que la serie es integrada de orden 1.

El estadístico de contraste es el t-ratio que se define como un cociente entre el estimador MCO del coeficiente de la variable retardada un periodo y la desviación típica de ese estimador.

Lo forma que adoptan estos estimadores depende de la forma que toman los elementos deterministas del modelo.

La distribución de probabilidad de ese estadístico no coincide con la t de student.

# Ideal Bacon –Descartes.

Es una línea de pensamiento, dentro de la Filosofía de la Ciencia, que trata de establecer criterios para preferir una teoría frente a otras.

Esta línea de pensamiento fue impulsada en la segunda mitad del siglo XX por el Neo-empirismo por autores como Lakatos, Radnitzky, Watkins,.... Los criterios tienen carácter bipolar: por un lado, trata de elegir las teorías que son informativas, profundas y arriesgadas; por otro, trata de mantener teorías que se ajusten a los hechos.

# Aspectos para saber como formular una predicción

A la hora de plantear un ejercicio de predicción es importante tener en cuenta los siguientes aspectos:

1. Quién va a utilizar la predicción y para que.
2. Que nivel de detalle se requiere
3. El horizonte temporal de predicción
4. Nivel de exactitud requerido
5. Recursos disponibles
6. Urgencia y frecuencia con las que se necesitan las predicciones.
7. Nivel de agregación deseado
8. Todos aquellos aspectos que pueden ser relevantes para incorporar la predicción en el proceso de toma de decisiones.

# Métodos de predicción. Tipos.

Podemos distinguir tres grandes grupos de métodos de predicción:

1. Métodos Subjetivos
2. Estudios de Mercado
3. Métodos Objetivos

## 1. Métodos Subjetivos.

Son métodos basados en la información privilegiada que algunos agentes tienen del fenómeno que se va a predecir. Dentro de estos métodos podemos distinguir

- Jurado de Opinión
- Método de Escenarios
- Método Delphi
- Analogía Histórica

## 2. Estudios de Mercado

En este caso, es necesario recoger la información que está dispersa en una población mediante técnicas de muestreo.

## 3. Métodos Objetivos

- Modelos univariantes de series temporales
- Modelos Causales Multivariantes

# Metapredicción.

Hace referencia a los siguientes puntos:

1. La incertidumbre respecto al futuro siempre va a estar presente en cualquier ejercicio predictivo. La finalidad de éste no es tanto eliminar dicha incertidumbre como minimizarla.
2. Los economistas tienen que ser conscientes del tipo de realidad que la Economía estudia y trata de proyectar. Hutchison (1977) nos sitúa en el lugar apropiado: “Hay, sin embargo, generalizaciones útiles en la economía y las ciencias sociales que son descritas mejor como tendencias ya que en general no son tan precisas ni tan contrastables como las leyes propiamente dichas. Tendencias, y no leyes, es lo que el material de la economía y las ciencias sociales parecen proporcionar o han proporcionado hasta el momento.... A falta de leyes, todo lo que los economistas tienen son tendencias y deben de procurar sacar el máximo de ellas”.
3. Cada situación hay que tratarla de forma específica y diferente. Hay que distinguir según sea el tipo de variable, el horizonte de predicción, la información disponible, la frecuencia y urgencia requeridas, los medios disponibles para calcular la predicción, etc.
4. Como ya hemos comentado, hay numerosos métodos de predicción y en cada situación hay que aplicar el que sea más apropiado.
5. Hay que ser cuidadosos sobre como se presentan los resultados de un ejercicio predictivo. En principio, es importante destacar que la predicción son dos números: el primero es el valor predicho y el segundo se refiere a la amplitud del intervalo formado por los valores que, en torno al valor predicho, se consideran como aceptables habiendo tomado un nivel de significación a priori.

# Experimentos Aleatorizados.

Se refiere a una situación en la que todos los elementos de una población se pueden asignar aleatoriamente a dos grupos, A y B.

Los elementos dentro de cada grupo son heterogéneos entre sí, pero podemos decir que, agregadamente, la composición de A es similar a la de B. A los componentes del grupo A se les somete a un estímulo que no reciben los de B.

En ambos grupos se mide el nivel alcanzado por una determinada variable que puede verse afectada por el estímulo y que llamamos variable respuesta. Si los niveles alcanzados por esa variable son muy diferentes en ambos grupos podemos decir que la variable estímulo causa a la variable respuesta.

Ejemplo: dos parcelas de terreno agrícola contiguas; a una de ellas se le aplica un fertilizante y a la otra no y se mide la diferencia en las cosechas.

# Modelo VAR(p).

El modelo VAR (p) es un modelo en el que todas las variables consideradas son función lineal de los p valores pasados de las variables. Para el caso de dos variables el modelo VAR(2) sería el siguiente:

$$y_t = \phi_{111}y_{t-1} + \phi_{112}y_{t-2} + \phi_{121}x_{t-1} + \phi_{122}x_{t-2} + u_{1t}$$

$$x_t = \phi_{211}y_{t-1} + \phi_{212}y_{t-2} + \phi_{221}x_{t-1} + \phi_{222}x_{t-2} + u_{2t}$$

$u_{1t}$  y  $u_{2t}$  son perturbaciones aleatorias.

# No causalidad en el sentido de Granger.

Decimos que  $x_t$  no causa a  $y_t$  en el sentido de Granger si la varianza del error de predicción de  $y_t$ , un periodo hacia delante, es la misma cuando se utiliza el pasado de la propia variable que cuando se utiliza el pasado de las dos variables. Dicho de otra manera,  $x_t$  no causa a  $y_t$  en el sentido de Granger, si la predicción de  $y_t$  no mejora si se añade el pasado de  $x_t$ .



# TEMA 2

# Función de Verosimilitud, Gradiente, Matriz de Información y Cota de Cramer-Rao.

La función de verosimilitud es la función de probabilidad suponiendo una muestra fija de forma que el valor de la función de verosimilitud cambia conforme lo hacen los parámetros del modelo.

El gradiente es el vector de las primeras derivadas del logaritmo de la función de verosimilitud respecto a cada uno de los elementos del vector de parámetros. Por lo tanto, el gradiente tiene tantos elementos como parámetros hay en el modelo.

La matriz de información es menos la esperanza matemática de la matriz de segundas derivadas. Es una matriz cuadrada con un orden igual al número de parámetros. La cota de Cramer-Rao es igual a la inversa de la matriz de información.

# Estimadores Máximo-Verosímiles(MV)

Los estimadores MV son los que maximizan la función de verosimilitud de la muestra o, equivalentemente, su logaritmo. Se obtienen igualando a cero los elementos del gradiente. En general, son estimadores consistentes y asintóticamente eficientes.

# Estimador Insesgado y Estimador Eficiente.

Un estimador es insesgado si su esperanza matemática coincide con el valor del parámetro que se quiere estimar. Un estimador es eficiente si tiene la menor varianza entre todos los estimadores propuestos. Dentro de la familia de estimadores insesgados, la varianza del estimador eficiente coincide con la cota de Cramer-Rao.

# Estimador Consistente y Asintóticamente Eficiente.

Un estimador es consistente si converge en probabilidad al parámetro que se quiere estimar o, lo que es lo mismo, si todos los posibles valores del estimador giran en torno al parámetro que se estima con una dispersión que va haciéndose menor conforme el tamaño muestral crece hasta llegar a ser cero cuando  $T \rightarrow \infty$ , y.

Suponer que normalizamos multiplicando por  $\sqrt{T}$  la diferencia entre el estimador y el valor del parámetro, y que la expresión normalizada converge a una distribución asintótica que tiene una varianza finita diferente de cero. Decimos que el estimador es asintóticamente eficiente si la varianza de la distribución asintótica es menor que la correspondiente a otros estimadores consistentes.

# Estimador Superconsistente e Hiperconsistente.

Un estimador es superconsistente si es consistente y si después de normalizar multiplicando por  $T$  la diferencia entre el estimador y el parámetro que se estima, se llega a una expresión normalizada que converge a una distribución asintótica con varianza finita diferente de cero. Un estimador es hiperconsistente si es consistente y la normalización requiere multiplicar por  $T^\nu$  con  $\nu > 1$ .

# Hipótesis relativas a la parte aleatoria del Modelo Lineal General (MLG).

Suponemos que los regresores son estrictamente exógenos, por lo que la esperanza matemática de las  $T$  perturbaciones aleatorias será cero. Lo escribimos como:

$$E(u_i) = 0 \quad \forall i = 1, 2, \dots, T$$

Además, añadimos otras dos hipótesis. La primera la de no autocorrelación que significa que cualquier perturbación aleatoria es independiente del resto de las perturbaciones. Lo escribimos como,

$$E(u_i u_j) = 0 \quad \forall i, j = 1, 2, \dots, T, \quad i \neq j$$

La segunda hipótesis es la de homoscedasticidad, que indica que las  $T$  perturbaciones aleatorias tienen la misma varianza, lo que escribimos como

$$E(u_i^2) = \sigma^2 \quad \forall i = 1, 2, \dots, T$$

# Matriz de varianzas y covarianzas de los estimadores MCO en un modelo con k regresores.

La matriz de varianzas y covarianzas viene dada por:

$$Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Es una matriz cuadrada de orden k y  $\sigma^2$  es la varianza de la perturbación aleatoria del modelo. Los elementos de la diagonal principal son las varianzas de los k estimadores y los elementos fuera de la diagonal principal son las covarianzas de las parejas de estimadores que se pueden formar con esos k estimadores.



# Vector de residuos MCO.

El vector de residuos MCO viene dado por

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

En donde M es una matriz simétrica, idempotente de rango T-k. Además cumple que  $MX=0$ , por lo que el vector de residuos puede escribirse como:

$$\hat{u} = M(X\beta + u) = Mu$$

# Estimadores Restringidos de $\beta$ .

Son los estimadores que se obtienen aplicando MCO al modelo restringido.

Estos estimadores son insesgados si se cumplen las restricciones pero no lo son si las restricciones no son ciertas.

Por otra parte, la varianza de los estimadores restringidos siempre es menor que la de los estimadores no restringidos.

Este resultado no depende del cumplimiento de las restricciones. Pero si que depende si comparamos la estimación de la varianza de los estimadores.

En este caso, la estimación puede ser igual, mayor o menor según sea el valor estimado de la varianza de las perturbaciones.

# Residuos restringidos. Estimador de la varianza de las perturbaciones.

Los residuos restringidos son:

$$\hat{u}_{Rt} = y_t - \hat{\beta}_R x_t$$

En donde  $\hat{\beta}_R$  es el estimador de  $\beta$  definido en el modelo restringido.

La suma de cuadrados de los residuos del modelo restringido siempre es mayor que la de los residuos del modelo sin restringir.

Sin embargo no puede decirse lo mismo respecto al estimador de la varianza de las perturbaciones; este estimador puede ser igual, mayor o menor que el del modelo restringido dependiendo de los grados de libertad que aparecen en el denominador.

# Método de Almon.

El método de Almon se ha propuesto para estimar un modelo en el que, como variables explicativas de un modelo lineal, aparece el valor contemporáneo de un regresor y un número finito de retardos de este regresor.

Si se aplican MCO, la multicolinealidad hace que esos estimadores tengan grandes varianzas y no sean buenos estimadores pese a ser insesgados.

La multicolinealidad solo se puede contrarrestar con información a priori que se añada a la información muestral.

En el método de Almon la información a priori consiste en suponer que los coeficientes de los retardos son un polinomio de orden reducido, digamos dos o tres, del subíndice.

Sustituyendo estos polinomios en el modelo original se obtiene un nuevo modelo con dos o tres parámetros en el que se reduce el problema de la multicolinealidad.

Pero no hay que olvidar que el problema se resuelve solo si la información a priori es válida.

# Método de Koyck.

El método de Koyck es un método que sirve para estimar un modelo en el que, como variables explicativas de un modelo lineal, aparece un regresor con un número infinito de retardos. Este modelo, sin información a priori, no se puede estimar porque el número de observaciones muestrales es finito.

La información a priori que propone Koyck es suponer que los coeficientes de la estructura de retardos siguen una función geométrica decreciente del tipo,

$$\beta_i = \beta(1 - \lambda)\lambda^i$$

Sustituyendo esta expresión en el modelo original se obtiene el siguiente modelo

$$y_t = \lambda y_{t-1} + \beta(1 - \lambda)x_t + u_t - \lambda u_{t-1}$$

En este modelo, solo hay dos parámetros de situación y el problema del número infinito de parámetros queda resuelto.

Pero aparecen otros problemas como el de la autocorrelación en la perturbación aleatoria y la presencia de la variable dependiente, retardada un periodo, como variable explicativa.

De todas formas, la solución solo es válida si la información a priori en torno a la función geométrica decreciente es válida.

# Tema 3

# Contraste de Hipótesis

El contraste de hipótesis es un procedimiento estadístico que sirve para decidir si una hipótesis nula se rechaza o no.

El proceso consta de tres fases:

1. Definición del estadístico de contraste.
2. Derivación de la distribución de probabilidad del estadístico de contraste bajo la hipótesis nula.
3. Determinación de la región crítica del contraste utilizando los cuantiles de la distribución obtenida en la fase 2 y el nivel de significación especificado a priori.

# Función de Potencia y Tamaño del Error Tipo 1.

La función de potencia de un contraste es una función que nos proporciona, para cada valor del parámetro, la probabilidad de rechazar la hipótesis nula.

$$Pot(\theta) = P(Rechazar H_0)$$

El tamaño del error tipo 1 es el valor que toma la función de potencia para el valor del parámetro que especifica la hipótesis nula.

$$Tam. EI = Pot(\theta_{H_0})$$

Alternativamente, podemos decir que es la probabilidad de rechazar la hipótesis nula cuando es cierta, es decir, es la probabilidad de tomar una decisión incorrecta cuando los datos son generados por la hipótesis nula. Se dice, también, que es el tamaño del contraste.

$$P(Rechazar H_0 | H_0) = P(H_A | H_0)$$



# Función de Potencia y Potencia de un contraste.

La función de potencia de un contraste es una función que nos proporciona, para cada valor del parámetro, la probabilidad de rechazar la hipótesis nula.

$$Pot(\theta) = P(\text{Rechazar } H_0)$$

La potencia de un contraste es el valor que toma la función de potencia para valores del parámetro que caen bajo la hipótesis alternativa.

$$Pot(\theta_{H_A})$$

Por lo tanto, la potencia es la probabilidad de tomar una decisión correcta cuando los datos son generados bajo la hipótesis alternativa.

$$Pot(\theta_{H_A}) = P(\text{Rechazar } H_0 | H_A) = P(H_A | H_A)$$

NOTA:

Notar que los terminos son equivalentes:

$$P(H_A | H_A) = 1 - P(H_0 | H_A)$$

# Contraste Uniformemente más Potente de tamaño $\varepsilon$ (UMP).

Decimos que un contraste es UMP de tamaño  $\varepsilon$  si cumple:

1. Tener un tamaño igual a  $\varepsilon$
2. Su función de potencia toma siempre un valor superior a la de cualquier otro contraste que tenga el mismo tamaño.

Podemos sintetizar diciendo que el contraste UMP es aquel que entre todos los contrastes que se equivocan de la misma manera bajo la hipótesis nula, es el que acierta más para todos los valores del parámetro bajo la hipótesis alternativa.

# Línea de Procedimientos de Contraste Admisibles (LPA)

La LPA se sitúa en un sistema de coordenadas en el que, en el eje de abscisas, se pone el tamaño del error tipo 1 y, en el de ordenadas, el tamaño del error tipo 2.

Cada punto de esta línea representa un procedimiento de contraste admisible en el sentido de que cualquier otro contraste que tenga el mismo tamaño del error tipo 1 tendrá un tamaño del error tipo 2 que será mayor que el que corresponde al contraste que representa el punto de la LPA. Una vez situados en la LPA, para determinar la región crítica solo queda especificar el nivel de significación.

# Lema de Neyman-Pearson.

El Lema dice así: Sea A un procedimiento de contraste tal que la hipótesis nula se rechaza si  $aL_0(y) \leq bL_1(y)$  y no se rechaza si  $aL_0(y) > bL_1(y)$  en donde a y b son dos constantes positivas cualesquiera, y  $L_0(y)$  y  $L_1(y)$  son, respectivamente, los valores tomados por la función de verosimilitud cuando el parámetro se sustituye por los valores bajo la hipótesis nula y bajo la alternativa.

Entonces, para cualquier otro contraste B se cumple que:

$$a\varepsilon(A) + b\delta(A) \leq a\varepsilon(B) + b\delta(B)$$

En donde  $\varepsilon()$  y  $\delta()$  son, respectivamente, los tamaños de los errores tipo 1 y tipo 2.

# Contraste de la Razón de Verosimilitud (LR).

Suponemos que la función de verosimilitud depende de un vector de  $k$  parámetros que llamamos  $\theta$ .

Se trata de contrastar que los elementos de este vector cumplen  $r$  restricciones. El estadístico del contraste LR es

$$LR = 2[l(\tilde{\theta}) - l(\tilde{\theta}_R)]$$

En donde  $l(\tilde{\theta})$  es el valor que toma el logaritmo de la función de verosimilitud sustituyendo los  $k$  parámetros por sus estimadores MV sin restricciones;  $l(\tilde{\theta}_R)$  es el valor que toma el logaritmo cuando los parámetros se sustituyen por los estimadores MV con restricciones.

La distribución de probabilidad de este estadístico bajo la hipótesis nula es

$$LR \approx \chi^2(r)$$

En donde  $\chi^2(r)$  es una chi-cuadrado con  $r$  grados de libertad.

La región crítica del contraste viene dada por:

$$LR > \chi^2_{\varepsilon}(r)$$

$$e^{i\pi} + 1 = 0$$

En donde  $\varepsilon$  es el nivel de significación adoptado a priori.

# Contraste de los Multiplicadores de Lagrange (LM).

Suponemos que la función de verosimilitud depende de un vector de  $k$  parámetros que llamamos  $\theta$ .

Se trata de contrastar que los elementos de este vector cumplen  $r$  restricciones.

El estadístico del contraste LM es:

$$LM = d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R)$$

En donde  $d(\tilde{\theta}_R)$  e  $I(\tilde{\theta}_R)^{-1}$  son, respectivamente, el gradiente y la matriz de información, ambos evaluados con los estimadores MV con restricciones.

La distribución de probabilidad de este estadístico bajo la hipótesis nula es:

$$LM \approx \chi^2(r)$$

En donde  $\chi^2(r)$  es una chi-cuadrado con  $r$  grados de libertad.

La región crítica del contraste viene dada por:

$$LM > \chi_{\varepsilon}^2(r)$$

En donde  $\varepsilon$  es el nivel de significación adoptado a priori.

# Exogeneidad Estricta.

Se dice que una variable es estrictamente exógena cuando la esperanza matemática de la perturbación aleatoria del modelo, condicionada a cualquier valor de la variable, es cero.

Las implicaciones de esta definición son dos: primero, que la esperanza no condicionada de la perturbación es cero y, segundo, que la covarianza entre la perturbación y la variable es cero.

Además, el cumplimiento de la exogeneidad estricta hace que el estimador MCO sea consistente lo que permite una interpretación causal de la estimación realizada.

# Variable Instrumental.

Es una variable correlacionada con un regresor endógeno (relevancia del instrumento) e incorrelacionada con el término de error de la regresión (exogeneidad del instrumento).

La variable instrumental se utiliza cuando en un modelo el regresor no es estrictamente exógeno y, como consecuencia, el estimador MCO no es consistente.



# Sesgo de Variable Omitida

Se refiere al caso en que el regresor de un modelo no es estrictamente exógeno como consecuencia de que ese regresor depende de una variable observable y de otros factores y la perturbación del modelo depende de esa misma variable observable y de otros factores independientes de los que determinan al regresor.

La consecuencia es que el regresor y la perturbación del modelo están correlacionados por lo que el estimador MCO no es consistente.

# Sesgo de Simultaneidad

Se refiere al caso en que el regresor de un modelo no es estrictamente exógeno como consecuencia de que ese regresor depende de un grupo de variables estrictamente exógenas y de otros factores correlacionados con la perturbación del modelo.

La consecuencia es que el regresor y la perturbación del modelo están correlacionados por lo que el estimador MCO no es consistente.

Hay que utilizar estimadores de variable instrumental o estimadores en dos etapas para lograr estimadores consistentes.

# Sesgo de Error de Observación

Se refiere al caso en que el regresor de un modelo no es estrictamente exógeno porque la variable dependiente depende de un regresor que no es observable o, lo es, pero se observa incorrectamente.

La estimación del parámetro se hace en un segundo modelo en el que la variable no observable se sustituye por otra que sí es observable.

Pero la no coincidencia entre las dos, observable y no observable, produce la no exogeneidad estricta del regresor observable.

La consecuencia es que el regresor observable y la perturbación del modelo están correlacionados por lo que el estimador MCO no es consistente.

# Estimador de Variable Instrumental

Este estimador se propone para aquellos casos en que el regresor ( $x_t$ ) no es estrictamente exógeno.

Se parte de una variable, que es la variable instrumental ( $z_t$ ), que está correlacionada con el regresor pero incorrelacionada con la perturbación aleatoria del modelo y se define la ecuación normal que puede escribirse como:

$$\sum z_t y_t = \hat{\beta}_{VI} \sum z_t x_t$$

A partir de esta relación se despeja el estimador de variable instrumental que es un estimador consistente.

# Estimador en dos etapas.

Este estimador se propone para aquellos casos en que el regresor ( $x_t$ ) no es estrictamente exógeno debido a que puede escribirse como una función lineal de un grupo de variables estrictamente exógenas y de una perturbación que está correlacionada con la perturbación del modelo original.

Esta correlación provoca que el estimador MCO sea inconsistente.

El estimador en dos etapas es un estimador consistente.

Se define así:

- En la primera etapa, se hace la regresión del regresor sobre el grupo de variables estrictamente exógenas de las que depende y se define el regresor estimado.
- En la segunda etapa, se sustituye el regresor por el regresor estimado y se aplican MCO.

El estimador resultante es el estimador en dos etapas y es consistente.

# Contraste de Hausman

Es un procedimiento propuesto para contrastar la hipótesis nula de exogeneidad estricta.

El estadístico del contraste es:

$$m = \frac{\hat{q}^2}{Var(\hat{q})}$$

El numerador es el cuadrado de  $\hat{q} = \hat{\beta}_1 - \hat{\beta}_0$  en donde  $\hat{\beta}_0$  es un estimador del parámetro del modelo que es consistente y asintóticamente eficiente bajo la hipótesis nula pero es inconsistente bajo la hipótesis alternativa y  $\hat{\beta}_1$  es un estimador consistente bajo ambas hipótesis pero no es eficiente bajo la hipótesis nula.

La distribución de probabilidad de  $m$  bajo la hipótesis nula es una chi- cuadrado con un grado de libertad.

La región crítica del contraste viene dada por:

$$m > \chi^2_{\varepsilon}(1)$$

En donde  $\varepsilon$  es el nivel de significación fijado a priori.

# Factor de Parsimonia del $\bar{R}^2$ .

La región crítica de este criterio cuando se comparan dos modelos anidados de  $k_1$  y  $k_2$  regresores, respectivamente, es:

$$\bar{R}_1^2 < \bar{R}_2^2$$

Sustituyendo en esta expresión la fórmula:

$$\bar{R}_i^2 = 1 - \frac{T-1}{T-k} \frac{\hat{u}_i' \hat{u}_i}{\sum (y_t - \bar{y})^2}$$

la región crítica puede escribirse como:

$$\hat{u}_1 \hat{u}_1 > \hat{u}_2 \hat{u}_2 \frac{T-k_1}{T-k_2}$$

El factor de penalización es el término que multiplica a la suma de cuadrados de los residuos del modelo amplio, es decir:

$$h(\bar{R}^2) = \frac{T-k_1}{T-k_2}$$

# Factor de Parsimonia del AIC.

La región crítica de este criterio cuando se comparan dos modelos anidados de  $k_1$  y  $k_2$  regresores, respectivamente, es:

$$AIC_1 > AIC_2$$

Sustituyendo en esta expresión la fórmula  $AIC_i = \ln \tilde{\sigma}_i^2 + \frac{2k_i}{T}$  la región crítica puede escribirse como:

$$\hat{u}'_1 \hat{u}_1 > \hat{u}'_2 \hat{u}_2 \exp\left(\frac{2(k_2 - k_1)}{T}\right)$$

El factor de penalización es el término que multiplica a la suma de cuadrados de los residuos del modelo amplio, es decir:

$$h(AIC) = \exp\left(\frac{2(k_2 - k_1)}{T}\right) \approx 1 + \frac{2(k_2 - k_1)}{T}$$



# Factor de Parsimonia del SBIC.

La región crítica de este criterio cuando se comparan dos modelos anidados de  $k_1$  y  $k_2$  regresores, respectivamente, es:

$$SBIC_1 > SBIC_2$$

Sustituyendo en esta expresión la fórmula  $SBIC_i = \tilde{\sigma}_i^2 + \frac{\ln T k_i}{T}$  la región crítica puede escribirse como:

$$\hat{u}'_1 \hat{u}_1 > \hat{u}'_2 \hat{u}_2 \exp\left(\frac{\ln T (k_2 - k_1)}{T}\right)$$

El factor de penalización es el término que multiplica a la suma de cuadrados de los residuos del modelo amplio, es decir

$$h(SBIC) = \exp\left(\frac{\ln T (k_2 - k_1)}{T}\right) \approx \frac{\ln T (k_2 - k_1)}{T}$$

# Tema 4

# Cointegración.

Se dice que dos variables están cointegradas cuando cumplen las dos condiciones siguientes: Primero, las dos variables tienen el mismo orden de integración; segundo, se puede encontrar una combinación lineal de las dos variables tal que el residuo resultante tenga un orden de integración inferior al de las dos variables.

Se dice también que dos variables están cointegradas cuando la tendencia estocástica de una de ellas es explicada por la tendencia estocástica de la otra.

# Contrastes de cointegración

Son procedimientos propuestos para contrastar la hipótesis nula de no cointegración frente a la alternativa de existencia de cointegración.

Hay dos grupos de contrastes: uniecuacionales y multiecuacionales.

Los primeros se basan en los residuos MCO de la relación de cointegración.

Dentro del primer grupo, los más conocidos son el contraste CRDW y el contraste Dickey-Fuller aplicado a los residuos.

En el segundo grupo el más conocido es el contraste de Johansen que es el procedimiento de la razón de verosimilitud aplicado al modelo VAR definido para todas las variables.

El estadístico de contraste del CRDW es el estadístico del contraste Durbin-Watson para contrastar la autocorrelación.

La distribución de probabilidad bajo la hipótesis nula es diferente a la obtenida por Durbin y Watson debido a la no estacionariedad bajo la hipótesis nula.

Por último, la región crítica del contraste esta definida por aquellos valores del estadístico CRDW que superan un valor que depende del número de variables incluidas en la relación de cointegración y del nivel de significación adoptado previamente.

# Correlación Espuria

La correlación espuria se refiere a aquellas situaciones en las que un uso incorrecto de ciertas técnicas estadísticas puede llevar a la conclusión de que dos variables están relacionadas entre sí cuando realmente no lo están.

Por eso se dice que “correlación no es causación” o que “la correlación está en los datos y la causación en la mente”.

Si, en el marco no estacionario, tenemos dos variables que son  $I(1)$  y han sido generadas independientemente una de la otra y hacemos la regresión con MCO entonces tanto el t-ratio como el coeficiente de determinación nos van a decir que las dos variables están fuertemente relacionadas cuando sabemos que no lo están.

El error, en este caso, reside en no haber contrastado previamente la hipótesis nula de no cointegración.

Si hubiéramos prestado atención a los residuos, la conclusión habría sido muy diferente.

# Modelo con Mecanismo de Corrección del Error (MCE).

El MCE es la versión restringida del modelo VAR (p) de variables en niveles que están cointegradas.

Incorporando en el modelo VAR las restricciones asociadas con la cointegración se llega al MCE.

En este modelo, cada uno de los incrementos contemporáneos de las variables se explica en función de los retardos de esos incrementos y del residuo de la relación de cointegración retardado un periodo.

Para el caso de dos variables, el modelo VAR (2) en niveles toma la forma siguiente:

$$y_t = \phi_{111}y_{t-1} + \phi_{112}y_{t-2} + \phi_{121}x_{t-1} + \phi_{122}x_{t-2} + u_{1t}$$

$$x_t = \phi_{211}y_{t-1} + \phi_{212}y_{t-2} + \phi_{221}x_{t-1} + \phi_{222}x_{t-2} + u_{2t}$$

El correspondiente MCE es:

$$\Delta y_t = \alpha_1(y_{t-1} - \beta x_{t-1}) + \gamma_{111}\Delta y_{t-1} + \gamma_{121}\Delta x_{t-1} + v_{1t}$$

$$\Delta x_t = \alpha_1(y_{t-1} - \beta x_{t-1}) + \gamma_{211}\Delta y_{t-1} + \gamma_{221}\Delta x_{t-1} + v_{2t}$$