

# Algunos problemas en el empleo de los métodos econométricos

**Angel Vilariño**

## **1. Los datos**

En las ciencias físicas es costumbre presentar los datos junto con una medida del error de observación, mientras que en las ciencias sociales está costumbre no es habitual. En muchas ocasiones ni se ha realizado el esfuerzo de estimar dicho error por lo que resulta imposible publicarlo. En economía todas las variables de interés se miden con error. El producto interior bruto, el consumo, la inversión, el stock de capital, la balanza de pagos, la tasa de paro, los agregados monetarios, la inflación,...En el caso de las variables financieras existe un grado mayor de precisión por tratarse de datos muy desagregados y observados directamente, como los precios de cierre diarios de una acción, el valor diario de apertura o cierre de un índice bursátil, o el valor de un tipo de cambio a una hora determinada. Esto supone una ventaja inicial respecto a otras variables, pero en muchos modelos las variables que intervienen son las expectativas y no los datos observados, por lo que volvemos a encontrarnos el problema de, en primer lugar definir los mecanismos de generación de expectativas, y a continuación los problemas derivados de la estimación de variables no observadas y la cuantificación de los errores de observación.

Si se publicarán los errores de estimación de las variables económicas, como por ejemplo, el PIB y la tasa de paro, muchas discusiones y reflexiones sobre la “bondad” o “maldad” de las cifras publicadas quedarían vacías de contenido. La publicación y el manejo de las variables económicas se realiza sin ningún tipo de reflexión sobre hasta donde llega la precisión de las cifras y se utilizan a veces décimas o centésimas, o unidades, que carecen de toda significación.

Para entender las causas de los errores de medición de las variables económicas es necesario conocer como se elaboran los datos, y en este sentido para muchas variables las encuestas juegan un papel esencial. Encuestas de salarios, de horas trabajadas

para elaborar índices de productividad, encuesta de presupuestos familiares para las ponderaciones del índice de precios y patrones de consumo, encuesta de población activa, encuesta de utilización de la capacidad productiva, además de formularios y registros de actividades, declaraciones de actividades y otros medios indirectos de obtención de información, como los estados financieros de las empresas, o la información que las entidades financieras suministran al Banco Central.

Algunos de los problemas que presentan las encuestas y se trasladan a los datos son:

- Diseño defectuoso de los cuestionarios
- Ocultación de información, mentiras
- Falta de preparación de los encuestadores (tiene relación con el coste elevado de la realización de buenas encuestas)
- Sesgos subyacentes (economía informal, actividades no legales)
- Errores en la manipulación de la información
- Frecuencia con la que se observan los hechos. Pueden mezclarse respuestas de diferentes fechas u obtenerse de unos datos primarios, datos de frecuencia menor con procedimientos ad-hoc como trimestralizar datos anuales.

## **2. Los modelos econométricos**

Un modelo econométrico es una construcción matemática que se rige por criterios de coherencia interna según las reglas formales de las matemáticas y de la estadística en particular. Pertenece al campo de la estadística matemática y en ese ámbito no existen modelos buenos o malos, sino consistentes o no, y todos son consistentes en el sentido de no contradicción interna. A partir de determinadas hipótesis se obtienen las propiedades del modelo, que son ciertas en el ámbito de las formas matemáticas.

El más sencillo de los modelos econométricos, el modelo clásico de regresión lineal (MCRL), está definido de forma esencial del modo siguiente:

Suponemos dos variables X e Y, y se postula una relación entre ambas definida por la ecuación:

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

$x_t$  e  $y_t$  representan los valores de X e Y en un evento genérico definido por el subíndice t. El subíndice t hace referencia tanto a una secuencia temporal, como a cualquier otro criterio de ordenación.

- a) Los parámetros  $\alpha$  y  $\beta$  son constantes
- b) La variable X no es una variable aleatoria.
- c)  $\varepsilon_t$  representa los valores de una variable aleatoria  $\varepsilon$  que tiene esperanza (media) cero y varianza  $\sigma_\varepsilon^2$  constante. La variable  $\varepsilon$  carece de autocorrelación. Lo anterior se expresa mediante  $E(\varepsilon_t) = 0$   $\text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 \quad \forall t$  y  $E(\varepsilon_t \varepsilon_{t-k}) = 0 \quad \forall t \neq k$   
Como X no es aleatoria se verifica que,  $E(x_t \varepsilon_t) = x_t E(\varepsilon_t) = 0$   
Un caso importante es cuando se relaja la hipótesis de que X no se aleatoria, pero se supone que X y  $\varepsilon$  son variables aleatorias independientes, en cuyo caso  $E(x_t \varepsilon_t) = E(x_t) E(\varepsilon_t) = 0$
- d) No falta ninguna variable en el modelo. A veces en un intento de “realismo” se dice que no es cierto que no existan otras variables influyentes, pero son muchas y de poca intensidad cada una, de tal modo que el resultado de la acción simultánea de todas ellas se puede sintetizar en la variable aleatoria  $\varepsilon$ . Entonces esta variable, el “ruido”, se puede interpretar como el entorno subyacente, todo lo que no está en la variable explicativa X.
- e) La variable X causa o explica a la variable Y. En la ecuación anterior carece de sentido intentar despejar la  $x_t$  en función de la  $y_t$ . No existe bidireccionalidad (*feedback*).

Este modelo puede considerarse el modelo canónico de la econometría, ya que sirve de referencia para el resto de los modelos. A medida que se relajan las hipótesis anteriores se obtienen nuevos modelos, cuyo análisis, estimación, propiedades y relaciones con el modelo canónico, son el contenido de gran parte de los textos de econometría.

Bajo las anteriores hipótesis la estadística matemática demuestra que si disponemos de una muestra aleatoria de  $N$  valores de  $x_t$  e  $y_t$ , podemos obtener estimadores con propiedades deseables<sup>1</sup> para  $\alpha$ ,  $\beta$  y  $\sigma_\varepsilon^2$ , que son los parámetros desconocidos.

Las expresiones de los estimadores MCO (mínimos cuadrados ordinarios) se muestran a continuación:

$$\hat{\beta} = \frac{\sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad \hat{\sigma}_\varepsilon^2 = \frac{\sum_{t=1}^N (y_t - \hat{\alpha} - \hat{\beta}x_t)^2}{N - 2}$$

$$\bar{x} = \frac{\sum_{t=1}^N x_t}{N} \quad \bar{y} = \frac{\sum_{t=1}^N y_t}{N}$$

y  $N$  el número de observaciones de las variables  $X$  e  $Y$ .

Una vez estimado el modelo se define una nueva variable  $\hat{\varepsilon}$ , el residuo de la ecuación, que tiene un papel central en el contraste de la validez del modelo.

$$\hat{\varepsilon}_t = y_t - \hat{\alpha} - \hat{\beta}x_t$$

Además es de amplia utilización el estadístico  $R^2$ , definido en general por la expresión

$$R^2 = \frac{\sum_{t=1}^N \hat{\varepsilon}_t^2}{\sum_{t=1}^N (y_t - \bar{y})^2}$$

---

<sup>1</sup> Propiedades de insesgadez y varianza mínima entre todos los estimadores lineales, entre las principales

en un modelo con término constante ( $\alpha \neq 0$ ) se demuestra que el estadístico  $R^2$  coincide con el cuadrado del coeficiente de correlación lineal entre la variable  $Y$  y la variable  $\hat{Y}$  definida por

$$\hat{y}_t = \hat{\alpha} + \hat{\beta}x_t, \text{ es decir } R^2 = \rho^2_{Y\hat{Y}}.$$

En el límite si  $\hat{Y}$  coincidiera con  $Y$  el coeficiente de correlación sería la unidad, o por el contrario si entre  $Y$  e  $\hat{Y}$  no existe relación, el coeficiente de correlación es nulo. En ese sentido se dice que  $R^2$  mide la “bondad del ajuste” (de  $\hat{Y}$ , valor estimado, a  $Y$ , valor muestral).

El estimador MCO es también un estándar para los procedimientos de estimación, al igual que el MCRL lo es para el resto de los modelos. Muchas técnicas de estimación son versiones modificadas de MCO, debido a la naturaleza del modelo, o procedimientos específicos para reconducir la situación a un formato MCO.

## **Simulación de un modelo correctamente especificado y resultados de la estimación**

Podemos entender mejor los problemas de la estimación e interpretación de un modelo econométrico realizando un ejercicio de simulación del modelo y estimándolo con los datos generados.

Suponemos que conocemos el modelo “verdadero”, tanto su especificación funcional, como los valores de los parámetros. Suponemos que la relación entre una variable  $X$  (causa) y otra  $Y$  (efecto) está determinada por la ecuación,

$$y_t = 2 + 5x_t + \varepsilon_t \quad E(\varepsilon_t) = 0 \quad \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = 1$$

Generamos  $N = 100$  valores independientes de una variable aleatoria normal con media cero y varianza  $\sigma_\varepsilon^2 = 1$ . La variable  $X$  toma un conjunto de valores controlados y

de este modo podemos generar los valores de la variable Y. Con los valores de X e Y generados por el modelo procedemos a estimar los parámetros  $\alpha$  y  $\beta$ , que sabemos valen 2 y 5 respectivamente. Los resultados de la estimación MCO se presentan en el cuadro 1. Se han realizado 10 simulaciones para reforzar la idea de que los resultados de la estimación del modelo utilizando los datos generados por el propio modelo, dependen de cada una de las muestras utilizadas. En este caso 10 muestras formadas por 100 pares de observaciones de X e Y cada una.

Cuadro 1. Resultados de la estimación del modelo

Muestra	N	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\varepsilon^2$	$R^2$
1	100	2,215	4,667	0,979	0,687
2	100	1,942	5,226	1,089	0,690
3	100	2,204	4,577	0,951	0,691
4	100	1,803	5,461	1,128	0,693
5	100	2,113	4,854	1,153	0,631
6	100	1,865	5,470	0,971	0,754
7	100	2,074	4,774	0,979	0,697
8	100	2,109	4,610	0,988	0,678
9	100	2,471	4,305	1,105	0,595
10	100	1,944	5,049	0,833	0,780
Medias		2,074	4,899	1,018	

Podemos observar como en cada una de las estimaciones realizadas los valores de los parámetros pueden apartarse sensiblemente del valor verdadero. Si pudiéramos realizar muchas veces la estimación del modelo (en economía en muchas situaciones solo existe una muestra), muchas más veces que las diez de este ejemplo, la media de los parámetros estimados tiende a los verdaderos valores del modelo, pero para una única realización muestral los resultados del cuadro son claramente ilustrativos. Otro tipo de análisis de sensibilidad que complementa el anterior, consiste en aumentar el valor de la varianza de las perturbaciones para comprobar como la dispersión de los valores de los parámetros estimados aumenta. Repetimos el ejercicio de simulación manteniendo las mismas variables aleatorias simuladas, los mismos valores de los parámetros, 2 y 5, pero aumentando la varianza del término de error, a un valor  $\sigma_\varepsilon^2 = 10$ . Los resultados de las estimaciones se recogen en el Cuadro 2.

Cuadro 2. Resultados de la estimación del modelo

Muestra	N	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_\varepsilon^2$	$R^2$
1	100	2,680	3,948	9,588	0,136
2	100	1,815	5,714	11,868	0,210
3	100	2,644	3,663	9,050	0,125
4	100	1,376	6,456	12,738	0,240
5	100	2,356	4,538	13,305	0,130
6	100	1,574	6,486	9,443	0,301
7	100	2,236	4,285	9,575	0,156
8	100	2,345	3,766	9,766	0,123
9	100	3,489	2,803	12,210	0,059
10	100	1,824	5,154	6,934	0,270
Medias		2,234	4,681	10,448	

Al estimar un modelo verdadero podemos hallar sensibles diferencias entre el valor del parámetro estimado y el verdadero valor, además de que el estadístico  $R^2$  puede tomar valores muy pequeños y eso no indica que la relación entre las variables X e Y no está adecuadamente especificada, sino que la relación entre la varianza del ruido y la varianza de la Y es muy grande.

El problema es que cuando estimamos el modelo conociendo únicamente los valores de las variables X e Y, suponiendo que han sido generados por el modelo  $y_t = \alpha + \beta x_t + \varepsilon_t$ , con parámetros desconocidos verdaderos,  $\alpha, \beta$  y  $\sigma_\varepsilon^2$ , nunca sabremos si los valores obtenidos  $\hat{\alpha}, \hat{\beta}$  y  $\hat{\sigma}_\varepsilon^2$  se separan mucho o poco de los anteriores. Y además generalmente dispondremos de una única muestra.

### **Modelo incorrectamente especificado**

Un modelo está incorrectamente especificado si se vulneran de forma grave alguna de las hipótesis establecidas. Cuantificar la gravedad no es tarea sencilla y se sitúa en el campo de la práctica, aunque existen herramientas para ayudar a realizar un diagnóstico, o evaluación, del modelo propuesto, una vez estimado.

Una cuestión previa se refiere a la naturaleza de los datos. Habitualmente se dice que los pares de datos de las variables X e Y forman una muestra. Pero no se trata de una muestra aleatoria. Si X es la renta disponible agregada de un país e Y es el consumo agregado, correspondiente a un conjunto determinado de años, los valores observados no son el resultado de ninguna operación de muestreo aleatorio. No son datos obtenidos mediante el diseño de un experimento aleatorio, son simplemente las únicas observaciones que existen. No podemos, un año cualquiera, dar valores a la renta disponible agregada, dentro del rango que nos parezca conveniente, y observar los valores del consumo agregado. Cada par de datos de cada año se interpretará como si hubiera sido el resultado de un experimento aleatorio que ha producido ese par de observaciones.

Al margen de este primer, y muy importante, escollo, los casos más frecuentes de deficiente o errónea especificación son los siguientes:

a) Los residuos no son estacionarios en media.

El análisis de los residuos es un poderoso instrumento para validar un modelo<sup>2</sup>. Si los residuos presentan algún tipo de tendencia o deambulan (rachas de valores por encima de la media y rachas por debajo de la media) no se verifica una de las hipótesis del modelo.

b) Los residuos presentan autocorrelación. Dicha presencia es una señal importante de mala especificación. Pueden existir varias causas diferentes para explicar la presencia de autocorrelación.

c) Los residuos no son estacionarios en varianza. Presentan heterocedasticidad.

Algunos de estos problemas pueden resolverse mediante las técnicas que proporciona la econometría, pero en otros casos refleja la existencia de relaciones espurias.

d) La relación entre las variables X e Y no es una relación de causalidad sino de correlación.

Dos variables pueden “moverse” a la vez y no guardar entre si ninguna relación. Es un clásico el ejemplo de Box, Hunter y Hunter, en su magnífico libro “Estadística para investigadores” , que muestran en un gráfico, prácticamente siguiendo una línea recta, la evolución del número de cigüeñas y el número de personas, en la ciudad alemana de

---

<sup>2</sup> El análisis de los residuos se realiza mediante los tests de raíces unitarias o mediante la metodología Box-Jenkins.



Oldenburg, durante el período 1930-1936. Nadie (o casi nadie) aceptaría que el número de cigüeñas determina el número de habitantes, y sin embargo la estimación de la ecuación

$$H_t = \alpha + \beta C_t + \varepsilon_t$$

daría, sin duda, desde un punto de vista puramente estadístico, buenos resultados. Esto nos conduce a la cuestión de la correlación y las relaciones espurias.

### **Correlación y relaciones espurias**

Supongamos que una teoría afirma la existencia de una relación entre dos variables que llamamos X e Y. Por ejemplo Y puede ser el nivel de precios y X la cantidad de dinero<sup>3</sup>. Esta supuesta relación puede establecerse de varias formas diferentes:

- a) La cantidad de dinero y el nivel de precios se mueven a la vez
- b) La cantidad de dinero determina el nivel general de precios
- c) La variación<sup>4</sup> de la cantidad de dinero y la variación del nivel de precios se mueven a la vez
- d) La variación de la cantidad de dinero determina la variación del nivel general de precios.

Pueden plantarse otras combinaciones, pero las anteriores son suficientes para la reflexión que viene a continuación.

a) y c) se refiere a relaciones de correlación, a comovimientos. Un índice posible, ampliamente utilizado, es el coeficiente de correlación lineal. Pero no es el único estadístico que puede medir la relación entre variables. Su uso está ampliamente extendido debido a su sencillez de cálculo, facilidad de interpretación, y que además guarda una estrecha relación con el estimador MCO  $\hat{\beta}$  de la ecuación  $y_t = \alpha + \beta x_t + \varepsilon_t$ ,

---

<sup>3</sup> Dejamos de lado la necesidad de precisar que se entiende operativamente por nivel general de precios (precios de consumo, deflactor del PIB, precios de consumo sin energía y alimentos no elaborados, ...) y por cantidad de dinero (alguno de los agregados denominados M1, M2, M3,....)

<sup>4</sup> La variación también necesita definirse. Primera diferencia, tasa de variación relativa, y la frecuencia: variación mensual, variación anual,...

ya que se demuestra que el coeficiente de correlación lineal entre las variables X e Y,  $\rho_{XY}$ , y el estadístico  $\hat{\beta}$ , están vinculados mediante la ecuación,  $\hat{\beta} = \rho_{XY} \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$

La presencia de una correlación lineal alta no dice nada de la determinación de una variable por otra. Las variables X e Y pueden estar vinculadas a otra u otras variables que actúen como causas o factores explicativos ocultos, sin que exista mediación directa de X a Y. En el caso de la población de una ciudad, o mejor el número de nacimientos, y el número de cigüeñas, la primera variable se explicará por factores causales en el ámbito de la demografía, y la segunda variable puede estar vinculada desde a factores de crecimiento económico, como al aumento de la sensibilidad medioambiental.

Pero además, de forma general sabemos, que la correlación alta entre variables que presentan tendencias en su comportamiento, esta asegurada. De forma intuitiva es posible entender este resultado si recordamos la definición del coeficiente de correlación lineal.

$$\rho = \frac{\sum_{t=1}^N (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^N (y_t - \bar{y})^2}}$$

Las dos raíces del denominador son cantidades positivas que actúan de factores de escala para acotar el rango de variación del valor del coeficiente. Si nos fijamos en el numerador, dos series con tendencia, por ejemplo creciente con el tiempo, van a tener sistemáticamente valores simultáneos por debajo de la media y valores por encima de la media, de tal modo que el signo de los factores del producto será básicamente el mismo, y el producto positivo, y como consecuencia correlación alta positiva.

Debido a este hecho, para evitar la posible estimación de relaciones espurias, es necesario eliminar la tendencia de las series Uno de los métodos más empleados es la diferenciación de la serie, o utilizar tasas logarítmicas, que además de ayudar a eliminar la tendencia, puede favorecer la corrección de la falta de varianza constante.

El diagnóstico de este problema, la falta de estacionariedad en media, y su solución está dentro del ámbito de la teoría de raíces unitarias, que ha generado procedimientos estándar dentro del trabajo econométrico aplicado. En cuanto a la falta de estacionariedad en varianza, se ha producido el desarrollo de los modelos de varianza condicional (ARCH, GARCH,...)

Suponiendo que se ha eliminado de las variables los factores que causan correlaciones espurias, las tendencias, la obtención de una correlación alta, indica la presencia de comovimientos, pero no de causalidad.

Para esta cuestión, que corresponde a los apartados b) y d) de la exposición anterior, la estimación de un modelo como el que estamos comentando no resolvería el problema, ya que establece una relación contemporánea entre la X y la Y. Aquí cabe un matiz, que tiene que ver con la frecuencia con la que se observan los datos, y con la naturaleza y la forma de medición de las variables del modelo. Si se trata de datos anuales, y fuera cierto que la variable X determina la variable Y, por ejemplo cantidad de dinero y precios, a lo largo de un año se han podido producir efectos desde la cantidad de dinero a los precios que pueden estar presentes en la cifra final anual de las dos variables. Si por ejemplo los datos son mensuales es más difícil aceptar que la relación de causalidad, en el caso de existir, pueda ser contemporánea. Por lo dicho, la frecuencia de observación de las variables de un determinado fenómeno no es indiferente, ni a los resultados que se pueden obtener al estimar una posible relación entre las variables, ni al diseño de los modelos adecuados a la detección de una relación de causalidad.

El contraste de causalidad más utilizado es el Granger en los siguientes términos:

La variable X causa (en el sentido de Granger) la variable Y si al estimar la ecuación

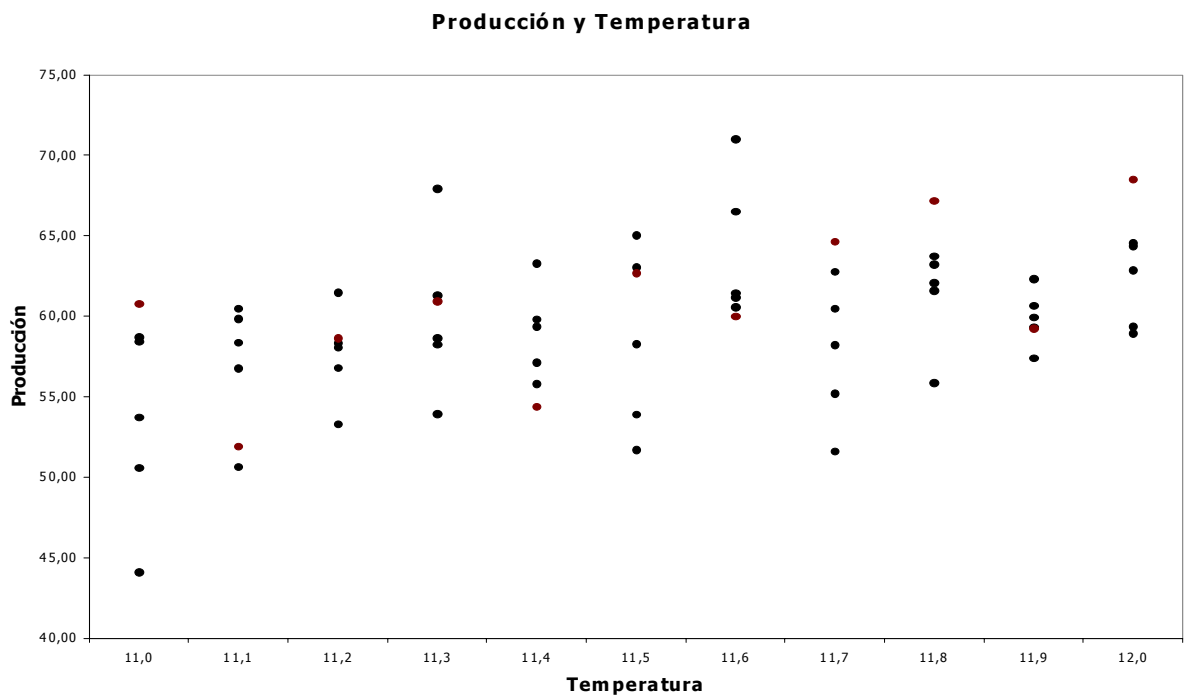
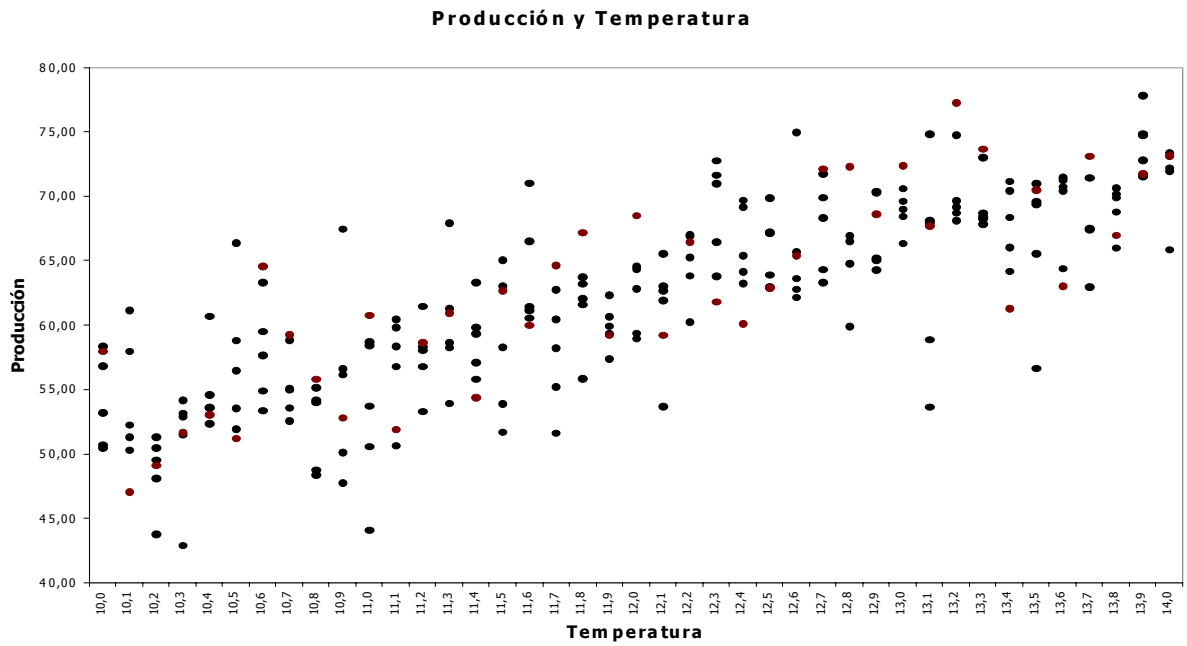
$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_m y_{t-m} + \beta_1 x_{t-1} + \dots + \beta_m x_{t-m} + \varepsilon_t$$

se rechaza la hipótesis de que los coeficientes de la X,  $\beta_1, \dots, \beta_m$ , son nulos.

Más formalmente: X no causa Y si la hipótesis nula  $H_0: \beta_1 = \dots = \beta_m = 0$  no se rechaza. El contraste se realiza con una F. De una forma más sencilla podemos plantear que la existencia de correlación entre dos variables estacionarias, poniendo en correspondencia valores no contemporáneos, es lo que se llama causalidad en el sentido de Granger.

### **Existencia de restricciones ocultas**

Una diferencia importante de la economía con otras ciencias está situada en la posibilidad de realizar experimentos controlados. En economía no es imposible pero para muchas de las cuestiones que se quieren investigar resulta imposible. Una rama de la estadística aborda el diseño de experimentos como cuestión de enorme importancia si se quiere hacer un uso adecuado de la estadística matemática para la interpretación de la realidad. La forma en la que se han generado los datos, no es neutral respecto a los procedimientos de estimación de relaciones entre las variables. Un ejemplo de Box, Hunter y Hunter es muy útil para clarificar la importancia de las condiciones de observación de las variables. Se refieren a un proceso de producción que tiene una fuerte dependencia con la temperatura, siendo muy sensible a sus variaciones. Si la temperatura está controlada dentro de un rango muy estrecho, puede parecer que no existe ninguna relación entre la temperatura y la producción.



En el gráfico superior está representada la relación entre la temperatura y la producción en un rango de temperaturas de 10°C a 14°C, mientras que en el gráfico

inferior el rango de variación de la temperatura está restringido entre 11°C y 12°C. La relación subyacente es bastante evidente en el primer gráfico pero aparece difuminada en el segundo.

El ejemplo anterior nos puede resultar útil para la econometría, ya que la imposibilidad de determinar el rango de variación de la variable de control (causa), se puede interpretar como la acción de restricciones ocultas que impiden estimar con precisión las hipotéticas relaciones subyacentes.

### **Problemas de agregación**

Existen relaciones teóricas que se justifican a un nivel micro (un agente individual) y posteriormente se estima una relación macro, utilizando datos agregados. Las preguntas que surgen son varias: ¿la aceptación o rechazo, de un modelo, en el nivel macro, implica lo mismo en el nivel micro?, ¿Qué relaciones existen entre los parámetros estimados en el nivel macro, con los parámetros del nivel micro?, ¿Qué restricciones adicionales hay que imponer a los modelos para poder realizar la agregación de los modelos de comportamiento individual?. La referencia obligada de estos temas es Theil (1959), pero veamos un ejemplo reciente de Granger (1999), que ilustra estos temas.

Granger toma como ejemplo la teoría (el modelo) de consumo de Hall, que básicamente dice que la utilidad marginal del agente representativo sigue un proceso de Markov de primer orden. Es decir partiendo de la ecuación de Euler que se obtiene al maximizar la suma de los valores esperados de consumo, aplicándolos un descuento debido a la preferencia temporal por el consumo más próximo en el tiempo, y teniendo en cuenta una restricción presupuestaria elemental, se tiene que

$$E_t U'(c_{t+1}) = \frac{1+\delta}{1+r} U'(c_t) \quad \text{o} \quad E_t U'(c_{t+1}) = \gamma U'(c_t) \quad \text{con} \quad \gamma = \frac{1+\delta}{1+r}$$

$\delta$  es el tipo de descuento de la preferencia por el consumo,  $r$  es el tipo de interés real,  $U$  es la función de utilidad, y  $E$  es el operador esperanza condicional a la información detentada en  $t$ .

La expresión anterior equivale a

$$U'(c_{t+1}) = \gamma U'(c_t) + \varepsilon_{t+1} \quad \text{suponiendo que} \quad E_t(\varepsilon_{t+1}) = 0$$

Para que la ecuación anterior pueda hacerse operativa, y permita algún tipo de contraste, es necesario particularizar la forma de la función de utilidad. El caso más sencillo se obtienen suponiendo una forma funcional cuadrática, y por lo tanto con derivada primera lineal.

Sea, por hipótesis, 
$$U(c_t) = -\frac{1}{2}(c_t - c_0)^2$$

Calculando las derivadas y sustituyendo obtenemos

$$c_{t+1} = \beta + \gamma c_t + u_{t+1} \quad \text{con} \quad \beta = c_0(1 - \gamma) \quad \text{y} \quad u_{t+1} = -\varepsilon_{t+1}$$

Una simplificación adicional es suponer que  $\gamma = 1$ , lo que transforma la ecuación anterior en un paseo aleatorio

$$c_{t+1} - c_t = u_{t+1}$$

La ecuación anterior también se formula aceptando la existencia de un término constante (paseo aleatorio con deriva)

$$\Delta c_{t+1} = \varepsilon_{t+1} \quad \text{o} \quad \Delta c_{t+1} = \beta + \varepsilon_{t+1}$$

Una vez obtenida esta ecuación se produce una explosión de trabajos que intentar refutar o no la hipótesis de paseo aleatorio del consumo.

Granger realiza unos penetrantes comentarios:

- a) Cuando obtenemos datos de consumo trimestrales se observa una fuerte componente estacional. De esta propiedad de los datos no dice nada la teoría anterior, por lo que el hecho empírico se intenta neutralizar. Aparece la desestacionalización, en esencia la obtención de unos datos de consumo a partir de los originales cuyo comportamiento estacional está alisado; y que no es una operación neutral, ya que supone pérdida de información, en sí misma, y en la medida que se altera la relación que puede tener el consumo con otras variables, a través de sus componentes estacionales. También se introducen otras pautas derivadas del modelo utilizado para la desestacionalización, y puede que se altere la volatilidad de la serie original.
- b) Si se estima una ecuación del tipo

$$\Delta c_{t+1} = \beta + \alpha_1 \Delta c_t + \alpha_2 \Delta c_{t-1} + \varepsilon_{t+1}$$

y se obtiene que los parámetros  $\alpha_1$  y  $\alpha_2$  no son significativamente distintos de cero, se está aceptando la hipótesis de paseo aleatorio. Pero ¿qué ocurriría si los parámetros fueran significativamente distintos de cero? ¿Significaría el rechazo de la teoría? Granger plantea el tema importante de la unidad de medida del tiempo, cuestión que la teoría ni se plantea, ni por lo tanto aclara, y recuerda que Halbrook Working demostró (Econometrica, 1960), que si una variable sigue un paseo aleatorio, medida con una unidad temporal determinada, seguirá una media móvil de orden uno (MA(1)) si se observa con una unidad temporal  $m$  veces la unidad anterior. Es decir el ruido en el segundo caso es de la forma

$$\varepsilon_t + b\varepsilon_{t-1}$$

verificándose la relación 
$$\frac{b}{1+b^2} = \frac{m^2-1}{2(2m^2+1)}$$



Esto plantea una nueva dificultad ya que un fenómeno económico puede tener diferentes representaciones estocásticas según el nivel de agregación temporal elegido.

- c) Otro aspecto relevante es el la agregación macro. La teoría se postula a nivel micro, pero se estima con datos agregados de consumo.

Supongamos que se intenta saber si se verifica a nivel micro que

$$\Delta c_{it+1} = \beta_i + \varepsilon_{it+1}$$

pero se estima la ecuación  $\Delta c_{t+1} = \beta + \varepsilon_{t+1}$  siendo  $\Delta c_{t+1} = \sum_{i=1}^N \Delta c_{it+1}$

Granger plantea que puede ser aceptada la ecuación agregada sin que eso demuestre la aceptación de las ecuaciones individuales. Estrictamente no puede contrastarse la teoría de Hall utilizando datos agregados.

- d) Por último Granger considera la posibilidad de otra forma funcional de función de utilidad, por ejemplo,

$$U(c_t) = c_t^{\frac{\sigma-1}{\sigma}}$$

función de elasticidad de sustitución constante, con lo que se llegaría a la ecuación

$$c_{it+1}^{\lambda} = \gamma c_{it}^{\lambda} + \varepsilon_{it+1} \quad \lambda = -\frac{1}{\sigma}$$

Ahora en este caso no existe la magnitud agregada que permita realizar el contraste ya que no se conoce el valor de  $\sum c_{it}^{\lambda}$

## Otras dificultades

- No linealidad
- Las interacciones no lineales entre variables explicativas
- Los efectos de feedback dinámicos (lineales y no lineales)

- d) Los fenómenos de umbral (saltos cualitativos)
- e) Los cambios del entorno subyacente: aparición de nuevas variables relevantes, influencia activa de variables dormidas (umbral), cambios de las relaciones de comportamiento,...

**Bibliografía:**

**Box, G.E.P. y Jenkins, G. M.** (1976): *Time Series Analysis forecasting and control*, Holden-Day, Oakland, California.

**Box, G.E.P., Hunter, W.G. y Hunter J.S.** (1989): *Estadística para investigadores*, Reverté, Barcelona.

**Desai, M.** (1981): *El monetarismo a prueba*, FCE, México.

**Granger, C.W.J.** (1999): *Empirical Modeling in Economics. Specification and Evaluation*, Cambridge University Press, Cambridge.

**Kennedy, P.** (1997): *Introducción a la Econometría*, FCE, México.

**Morgenstern, O.** (1970): *Sobre la exactitud de las observaciones económicas*, Tecnos, Madrid.

**Ruud, P.A.** (2000): *An Introduction to Classical Econometric Theory*, Oxford University Press, New York.

**Theil, H.** (1959): *Agregación lineal de relaciones económicas*, Aguilar, Madrid.