

INTRODUCCIÓN A LA ECONOMETRÍA E INFORMÁTICA

MODELOS ECONÓMICOS E INFORMACIÓN ESTADÍSTICA Eva Medina Moral

EXPRESIÓN DEL MODELO BÁSICO DE REGRESIÓN LINEAL

La expresión formal del modelo básico de regresión lineal, que es el modelo básico en econometría queda formulada como se expresa a continuación:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$$

Donde:

Y: es la variable endógena o explicada cuyo comportamiento se quiere analizar

X: cada una de las variables exógenas o explicativas y que son consideradas como las causas que crean transformaciones en la variable endógena.

B: son los parámetros cuyo valor desconozco y voy a estimar. A través de la estimación de los parámetros obtengo una cuantificación de las relaciones existentes entre la Y y cada una de las X.

U: perturbación aleatoria que recoge el efecto conjunto de otras variables no directamente explicitadas en el modelo, cuyo efecto individual sobre la endógena no resulta relevante.

i: es el subíndice que hace referencia a las diversas observaciones para las cuales se establece su validez. Según el tipo de valores con los que esté trabajando, el subíndice hará referencia a distintos momentos del tiempo (series temporales: las cotizaciones en bolsa diarias, los índices de predio al consumo mensuales, los datos anuales del PIB de un país, etc.) o a distintas unidades económicas (series de corte transversal: consumo de diferentes familias, inversión de distintas empresas, paro en diferentes provincias, etc.).

IMPORTANCIA DE LOS PARÁMETROS EN EL MODELO BÁSICO DE REGRESIÓN LINEAL

La principal utilidad que tienen los parámetros es la de cuantificar las relaciones que existen entre las variables explicativas y la variable endógena. Así:

- El parámetro que corresponde al término constante debe ser interpretado como el valor que toma la variable endógena cuando el resto de variables explicativas valen cero. Por ejemplo, en una función de consumo, aunque éste depende de la renta y de otras variables, cuando todas ellas valen cero el individuo realiza un consumo para sobrevivir, lo que es conocido como “autoconsumo”. Ese valor queda recogido en el modelo básico de regresión lineal a través del parámetro que corresponde al término constante.
- El resto de parámetros que acompañan a las variables explicativas miden la relación entre éstas y la variable endógena a través de su signo y su cuantía. El signo mide si la relación entre las variables es directa o inversa (si a medida que la explicativa incrementa también lo hace la endógena o viceversa). La cuantía sirve para medir la importancia de la relación entre la variable explicativa y la variable endógena.

Por tanto, el análisis de los parámetros estimados me permite conocer la estructura económica del fenómeno que estamos analizando, entendiendo por estructura el patrón de comportamiento de acuerdo con el cual se desarrolla una acción. De este modo, en un modelo en el que trato de explicar la evolución del consumo en función de la renta y de los tipos de interés, la estructura económica quedará definida como incrementos de consumo a medida que incrementa la renta y reducciones de consumo a medida que incrementan los tipos de interés.

Ahora bien, una vez estimado el modelo, admitimos que la estructura permanece constante para todo el periodo de estimación. Esto es, que los parámetros son los mismos para toda la muestra y que las relaciones permanecen constantes para todo el periodo analizado. Es por ello, que los parámetros no van acompañados de un subíndice en la expresión matemática del modelo básico de regresión lineal.

Sin embargo, la estructura o relaciones entre las variables pueden variar en el periodo analizado, lo que implicaría cambios en los valores de los parámetros. Los valores de los parámetros cambian cuando:

- Se incorpora una nueva variable al modelo. Ya que como en economía todo está relacionado entre sí, la inclusión de una nueva variable explicativa modifica las relaciones existentes entre las variables explicativas, y por tanto las relaciones existentes entre éstas y la variable endógena. Así, si dos variables explicativas están muy relacionadas entre sí, estarán explicando lo mismo del comportamiento de la endógena y al incluirlas juntas en el modelo su aportación a la evolución de la endógena se repartirá, mientras que si sólo incorporásemos una de ellas toda ella acumularía el peso importante en el

análisis de la endógena por lo que el valor de su parámetro sería superior que en el caso anterior.

- Se modifica el periodo muestral. Ya que la inclusión de nuevos años en el análisis implica incluir también nuevos factores explicativos de la variable endógena o una modificación en los pesos en que las variables explicativas participan en el comportamiento de la endógena.

El supuesto de constancia en el parámetro estimado conduce a errores en determinadas ocasiones. Esto ocurre cuando se produce un cambio estructural en el sistema económico a analizar. Así, en un mismo periodo se puede haber producido un cambio económico importante que implique una modificación radical de las relaciones existentes entre las variables. Por ejemplo: la entrada de España en la Unión Europea implica que el peso de las exportaciones en el crecimiento del PIB es mayor que durante los años anteriores a la inclusión en la UE, lo que modificaría las cuantías de los parámetros; países que han cambiado de un sistema económico socialista a uno capitalista implica un cambio radical en las relaciones económicas (antes tenía más peso el sector público y ahora las relaciones de libre mercado); el cambio que se produce en los hábitos de consumo con el trascurso de los años; la aparición de las tarjetas como medio de pago supone un cambio también en los hábitos de consumo.

UTILIDADES DE LOS MODELOS ECONÓMICOS

El modelo econométrico tiene tres utilidades principales:

- Análisis estructural: cuantificación de las relaciones que entre el periodo analizado ha existido entre las variables implicadas, a través del conocimiento del signo y valor de los parámetros estimados. Es decir, sirve para conocer como inciden en la endógena variaciones de las variables explicativas.
- Predicción: Dados unos valores a futuro para las variables explicativas, y conociendo la expresión matemática que relaciona las variables explicativas y la variable endógena, es posible predecir los valores que tomará a futuro la variable objeto de estudio.
- Simulación o evaluación de políticas: Efectos que tienen sobre la endógena diferentes estrategias que se planteen de las variables explicativas. Por ejemplo si analizamos las ventas de una empresa en función de los precios del producto y del nivel de gasto realizado en publicidad, podríamos estar interesados en analizar cuanto incrementarían las unidades vendidas si se mantienen los precios fijos y se incrementa el gasto en publicidad en un porcentaje determinado.

En general, el modelo econométrico es una herramienta de análisis que ayuda en la toma de decisiones tanto a nivel económico en general (macro) como en el ámbito de la dirección de empresas (micro).

CLASIFICACIÓN DE LOS MODELOS ECONOMETRÍCOS

Existe una tipología amplia de modelos econométricos en función de distintas clasificaciones:

Según el tipo de datos de las variables utilizadas en el modelo:

- Series temporales: los datos pueden corresponder a los valores de una variable en el tiempo. Estos pueden tener frecuencia, diaria, semanal, mensual o anual. Así podemos analizar las cotizaciones en bolsa diarias, los índices de predio al consumo mensuales, los datos anuales del PIB de un país, etc.
- Series de corte transversal: los valores corresponden a distintos sujetos para un mismo momento del tiempo. En este caso se trataría de series del tipo de consumo de diferentes familias, inversión de distintas empresas, paro en diferentes provincias, etc.

Según el momento del tiempo al que hacen referencia se distingue entre:

- Modelos estáticos: cuando el subíndice i hace referencia al mismo momento del tiempo o al mismo individuo económico tanto para la endógena como para todas las explicativas.
- Modelos dinámicos: cuando están involucradas las variables en diferentes puntos del tiempo. Así si estoy analizando la variable endógena consumo, utilizaré como variable explicativa la renta de ese mismo periodo, pero también podría utilizar la renta del año pasado, ya que mis decisiones de compra las tomaré en función de lo que pude ahorrar el año pasado. Al incluir variables en distintos momentos del tiempo podemos hablar de modelos dinámicos.

Según el número de variables endógenas que se desee explicar:

- Modelos uniecuaciones: únicamente existe una variable endógena.
- Modelos multiecuacionales: existen varias variables endógenas que deseamos explicar, algunas de las cuales pueden ser a su vez variables explicativas de otras ecuaciones.

Según la transformación de los datos que se realice:

- Modelo en niveles: las variables aparecen expresadas en unidades de medida.
- Modelo en tasas de variación: las variables aparecen expresadas como incrementos. Cuando una variable la expreso en vez de en niveles en incrementos estoy eliminando la tendencia. Al introducir las variables en niveles puedo encontrar un mayor número de variables explicativas aparentemente correctas, ya que es más fácil encontrar variables explicativas que tengan la misma tendencia que la endógena. Pero eso no significa que esas variables sean las que realmente son causas explicativas de los cambios de la

endógena. Por ello, al eliminar la tendencia de las variables exijo más al modelo, es decir, tengo en cuenta las variables que son realmente “causa”.

- Modelo en logaritmos. El modelo básico de regresión lineal permite únicamente trabajar con relaciones lineales. Pero no todas las variables tienen porque estar expresadas a través de una relación lineal. Cuando estimo un modelo únicamente con una variable endógena y una explicativa lo que trato es de encontrar la línea que mejor me recoja la información suministrada por ambas variables.

FASES PARA LA ELABORACIÓN DE UN MODELO ECONÓMETRICO

Las principales etapas que hay que cubrir en un modelo econométrico se pueden resumir en las cuatro siguientes:

Especificación: Esta etapa comprende tanto la determinación del tema objeto de análisis como la definición de las variables explicativas que se incluirán en el modelo.

- Selección del tema objeto de análisis. Este puede ser del campo de la economía, la gestión de empresas e incluso temas sociales no estrictamente económicos.
- Selección de las variables explicativas más importantes, la cual se realiza a través del:
 - Análisis de los antecedentes económicos: a través de las teorías económicas encuentro aquellas variables que a nivel general influyen de una manera importantes sobre la variable endógena
 - Análisis de los antecedentes econométricos: búsqueda de modelos similares a la materia objeto de análisis en libros y revistas sobre econometría.
 - Propio conocimiento del investigador.

Estimación: consiste en el cálculo del valor de los parámetros a través de la ayuda de un programa informático (Eviews). Para realizar esta fase es necesario previamente haber realizado una búsqueda y depuración de datos. Es necesaria la obtención de datos suficientes, homogéneos y actualizados.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2$$

En la expresión estimada no existe el componente de la perturbación aleatoria, ya que una vez que estimo, el valor de la endógena estimado se convierte en una combinación lineal exacta de las variables explicativas que he utilizado al realizar la estimación. La estimación de la perturbación aleatoria será el error que cometo con mi modelo al estimar, que incluirá precisamente las variables que dejo fuera de la

explicación (aquellas que tienen poca importancia sobre la variable que trato de analizar).

Validación: a través de la interpretación de los resultados analizaremos la bondad del modelo. De tal manera que si el modelo no es bueno para explicar a la variable endógena deberé perfeccionarlo a través de:

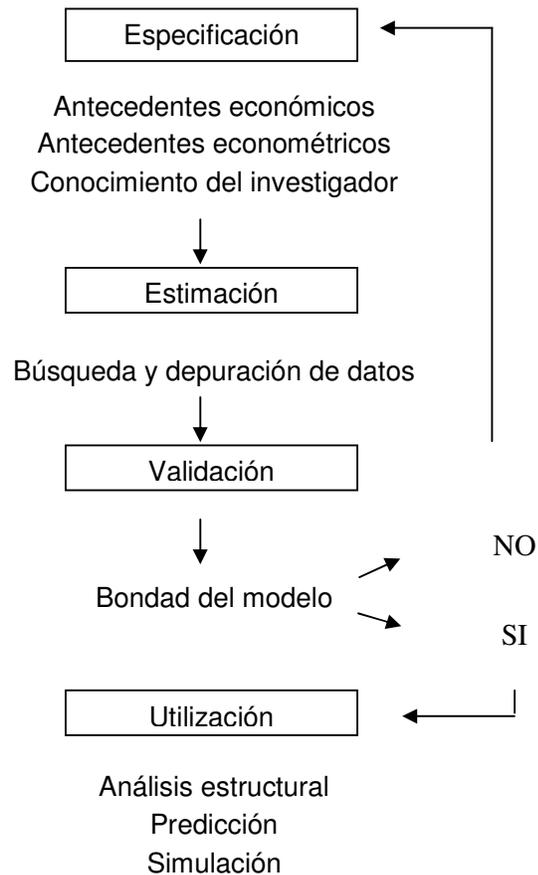
- Una reespecificación de las variables explicativas, es decir, es posible que haya olvidado incluir alguna variable importante por lo cual el modelo me esté dando un grado de error elevado.
- Una nueva búsqueda de los datos utilizados, ya que si no son los correctos pueden estar añadiendo un componente errático a la estimación.
- Etc.

Esta tarea puede repetirse en la práctica un número elevado de veces.

Utilización: para realizar:

- Análisis estructural: cuantificar las relaciones entre las X y la Y.
- Predicción: anticipar los valores a futuro de la Y
- Simulación: efectos sobre Y de distintas estrategias de las X.

Esquema a seguir en la elaboración de un modelo econométrico



TIPOLOGÍA DE VARIABLES Y DATOS EN LA MODELIZACIÓN ECONÓMETRICA

Se entiende por **variable** al concepto económico que queremos analizar.

Normalmente utilizaremos variables cuantitativas, es decir, aquellas cuyos valores vienen expresados de forma numérica.

Sin embargo, también existe la posibilidad de incluir en el modelo econométrico información cuantitativa, siempre que la información cualitativa pueda expresarse de forma cuantitativa. Dentro de este tipo de variables conviene conocer las siguientes posibilidades que se le pueden presentar al econométra:

- Variables proxies: son variables aproximadas a la variables objeto de análisis. Por ejemplo, si quiero utilizar una variable que mida el nivel cultural de un país (variable cualitativa) puedo utilizar como variable proxy el número de bibliotecas existentes en un país, que si bien no recoge el concepto exacto que yo quiero medir, si se aproxima al mismo.
- Variables ficticias o dummy: estas variables representan la incidencia que tiene sobre la variable endógena objeto de análisis un fenómeno cualitativo. Habitualmente, a la variable ficticia se le asignan dos valores arbitrarios según se de o no cierta cualidad en un fenómeno. Así, se le puede asignar el valor 1 si ocurre un determinado fenómeno y 0 en caso contrario. Estas variables pueden ser de dos tipos:
 - Ficticia de intervalo: Por ejemplo si estoy analizando la variable exportaciones en España desde 1970 hasta el año 2000, hay un hecho importante que es la entrada de España en la Unión Económica que debo recoger a través de la utilización de la variable ficticia.
 - Ficticia de escalón: Por ejemplo si está analizando el crecimiento económico de un país en el que en un año determinado hubo un acontecimiento meteorológico que tuvo una repercusión negativa sobre la economía, al tratarse éste de un dato casual (y no equilibrado con el resto de valores que toma la serie) debo introducir en el modelo este tipo de información para que la tenga en cuenta en la estimación y cometa un menor error.

Las variables ficticias se pueden incluir tanto en modelos temporales como en modelos de corte transversal. Por ejemplo, si se analiza la venta de coches a nivel provincial, se puede incluir una variable ficticia que valga 0 cuando la provincia no tiene metro y 1 cuando si que lo tiene.

Se entiende por **datos**, los diferentes valores que toma una variable.

Los datos pueden corresponder a los valores de una variable en el tiempo (serie temporal), o a valores para diferentes sujetos en un momento dado (datos de corte transversal).

Por otro lado, las formas alternativas en que un modelo puede venir expresado, obliga frecuentemente a transformaciones de los datos originales (tasas de crecimiento, diferencias, logaritmos) o incluso a un cierto tratamiento previo de los mismos (eliminación de tendencia, eliminación de estacionalidad, etc...)

Con la información disponible, el econométra deberá elaborar una base de datos de tal manera que:

Años	Individuos	Variable 1	Variable 2	Variable 3
1990	1	Datos	Datos	Datos
1991	2	Datos	Datos	Datos
...	...	Datos	Datos	Datos
2000	N	Datos	Datos	Datos

Los datos que formen la base de datos han de ser:

- Suficientes: como mínimo para poder realizar la estimación, el número de observaciones debe ser igual al número de parámetros que queremos estimar. Sin embargo si no se cumplen unos requisitos mínimos, aunque teóricamente se puede realizar la estimación, ésta no será fiable.
- Homogéneos: los datos que hacen referencia a una misma variable deben estar expresados de una forma homogénea, esto quiere decir que todos deben ir en niveles o en tasas de variación o en logaritmos. La homogeneidad de los datos también hace referencia al hecho de que todos deben o no ir corregidas de determinados efectos que se dan en la economía como la tendencia o la estacionalidad. Nunca se puede incluir en el modelo una variable cuyos datos estén expresados en distintas bases.
- Actuales: La falta de actualidad en los datos impide realizar un análisis completo del fenómeno económico, ya que éste se referirá únicamente al periodo muestral utilizado en la estimación.