PUBLICACIONES DE 4[er] CURSO

Grado: Economía

Asignatura: ECONOMETRÍA III

Tema 1: El Proyecto Econométrico
Apartado 1.2: Tipos de Variables

Grupos: 241, 242

Profesores: Antonio Aznar y M. Teresa Aparicio

Departamento de ANÁLISIS ECONÓMICO

Curso Académico 2014/15

Facultad de
Economía y Empresa
Universidad Zaragoza

Apartado 1.2

Tipos de Variables Económicas
Hemos visto que los modelos se componen de variables y relaciones que las unen. Las variables representan conceptos de forma cuantitativa. Una primera distinción relevante es entre variables continuas y variables discretas.
Una **variable continua es** aquella que puede tomar cualquier valor dentro del rango que tiene asignado. Por ejemplo, el PIB o la Inversión son variables continuas.
Una **variable discreta** es aquella que solo puede tomar un número finito de valores, generalmente números enteros. Tipos importantes de variables discretas son: **variables binarias** y variables de **opción múltiple.** Las primeras, también llamadas variables ficticias, toman sólo dos valores, en general 0 y 1. Por ejemplo, se define una variable para dar cuenta del sexo de una persona. Esta variable toma el valor 0 si la persona es hombre y 1 si es mujer. Las variables de opción múltiple toman más de dos valores dependiendo de las opciones que representan. Por ejemplo, se define una variable para dar cuenta del nivel de estudios de las personas consideradas en un determinado trabajo aplicado; esta variable toma el valor 0 si la persona no tiene estudios; toma el valor 1 si solo tiene estudios primarios; toma el valor 2 si tiene estudios secundarios y, por último, el valor 3 si tiene estudios universitarios.
Otra distinción relevante es entre variables para datos temporales, variables para datos de sección cruzada (o cross-section) y variables para datos panel.
Las **variables para datos temporales** son variables que toman valores a lo largo de diferentes periodos temporales, meses, trimestres o años. Describen,

cuantitativamente, el comportamiento de un individuo, una institución o un país a lo largo del tiempo. Ejemplos son el PIB trimestral de un país entre 1990-I y 2010-IV, o el volumen de ventas mensual de una empresa entre 2000/01 y 2012/12.

Las variables para datos de sección cruzada son variables que toman, valores en un mismo periodo, para un grupo de individuos, instituciones o países. Por ejemplo, la variable empleo en todos los países europeos en el año 2011 o el valor añadido bruto de las comunidades autónomas españolas en el año 2012.

Las variables para datos panel son variables que dan cuenta del comportamiento cuantitativo de un grupo de individuos, instituciones o países a lo largo de una serie de periodos temporales. Por ejemplo, el empleo anual en todos los países europeos observados entre 2000 y 2012, o el valor añadido bruto de las comunidades autónomas españolas entre 2002 y 2011.

Esta distinción es relevante porque cada tipo de variable reclama un tipo de modelización diferente. Para los datos de series temporales el principal problema es el tratamiento de la dependencia temporal mientras que, para los datos de sección cruzada, el principal problema es el de la heterogeneidad. Los datos panel combinan ambas dimensiones.

Dentro de las variables de series temporales vamos a distinguir entre variables estacionarias y variables no estacionarias. Esta distinción también se puede extender para los datos panel. Para caracterizar la estacionariedad vamos a considerar que cada observación de una variable corresponde a la observación de un elemento de un proceso estocástico con duración infinita ….1,2,…..T,….,. Se trata de caracterizar el comportamiento del elemento tésimo que podemos considerar como el elemento estandar del proceso. Sea y_t este elemento.

Variable Estacionaria

Decimos que una variable es estacionaria cuando la pauta de comportamiento temporal que sigue no está influida por el paso del tiempo. Hay dos sentidos de estacionariedad: en sentido fuerte y en sentido débil. Decimos que una variable es estacionaria en **sentido fuerte** cuando la distribución de probabilidad de cualquier conjunto del mismo número de elementos sucesivos del proceso no depende del periodo al que corresponde el primer elemento del conjunto. Si f(.) denota la función de probabilidad entonces la estacionariedad en sentido fuerte implica que,

$$f(y_1) = f(y_2) = ... = f(y_t)$$
$$f(y_1 y_2) = f(y_4, y_5) = ... = f(y_t, y_{t+1})$$
$$f(y_1, y_2, ...y_T) = f(y_{1+s}, y_{2+s}, ...y_{T+s})$$

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

$$Ey_t = \mu \quad \forall t$$
$$Var(y_t) = \sigma^2 \quad \forall t$$
$$Cov(y_t, y_s) = \gamma_{t-s} \quad \forall t, s$$

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 variación en torno a ese valor medio que no cambia con el tiempo y con una dependencia temporal que tampoco cambia.

La implicación de las propiedades asociadas con la estacionariedad débil es que cualquier elemento del proceso tiene momentos de primer y de segundo orden que no dependen del tiempo. Podemos decir que y_t es

$O_p(1)$ para indicar que cualquier elemento del proceso tiene una distribución de probabilidad bien definida y no precisa de ninguna normalización. Otras implicaciones se derivan para combinaciones de los elementos del proceso. Por ejemplo, el sumatorio de T elementos del proceso no tiene un limite bien definido pero si se divide por $\sqrt{T}$ entonces, por el teorema central del límite, sabemos que la expresión normalizada converge a una distribución de probabilidad bien definida. Podemos decir que $\sum_1^T y_t$ es $O_p(T^{\frac{1}{2}})$ o, alternativamente,

$$\frac{1}{T^{1/2}}\sum_1^T y_t \xrightarrow[d]{T\to\infty} N(0,\sigma_y^2) \qquad (1.2.1)$$

El mensaje importante es que algunas expresiones necesitan ser normalizadas para que cuando la evidencia empírica sea grande se obtenga una distribución con momentos finitos.

Otra expresión que tiene interés es la suma de los cuadrados de los elementos del proceso, $\sum_1^T y_t^2$.

Claramente se ve que este sumatorio se va a infinito cuando el tamaño muestral crece por lo que es necesaria la normalización. Es fácil demostrar que si esta suma se divide por T entonces si que converge en probabilidad, en este caso no en distribución, a una cantidad finita.

Podemos decir que $\sum_1^T y_t^2$ es $O(T)$, o bien que

$$\frac{1}{T}\sum_1^T y_t^2 \xrightarrow[p]{T\to\infty} q_{yy} \qquad (1.2.2)$$

Siendo q_{yy} una constante.

Una característica importante de las variables estacionarias es que los momentos muestrales convergen en probabilidad a los correspondientes momentos poblacionales y esta propiedad es muy importante para derivar las propiedades de estimadores y contrastes. Así, podemos escribir,

$$\frac{1}{T}\sum_{1}^{T} y_t \xrightarrow[p]{T\to\infty} \mu$$

$$\frac{1}{T}\sum_{1}^{T} (y_t - \bar{y})^2 \xrightarrow[p]{T\to\infty} \sigma_y^2$$

Variable No Estacionaria

Decimos que una variable es no estacionaria cuando la pauta de comportamiento temporal que sigue depende de la variable tiempo. Esta dependencia puede afectar solo a la media, solo a la varianza o a las dos. Podemos distinguir tres tipos de variables no estacionarias.

Variable solo con tendencia determinista

El modelo sería

$$y_t = \gamma_0 + \gamma_1 t + u_t \qquad (1.2.3)$$

Se supone que u_t es una variable estacionaria con media cero. Los momentos de esta variable son los siguientes,

$$Ey_t = \gamma_0 + \gamma_1 t$$

$$Var(y_t) = Var(u_t) = \sigma^2$$

$$Cov(y_t, y_s) = Cov(u_t, u_s) = \gamma_{t-s}$$

Se ve como, para esta variable, la dependencia temporal se manifiesta solo en la media.

Variable solo con tendencia estocástica

En este caso el modelo es

$$y_t = y_{t-1} + u_t = y_0 + \sum_{1}^{t} u_i \qquad (1.2.4)$$

Por lo que los momentos serán

$$Ey_t = y_0$$
$$Var(y_t) = t\sigma^2$$
$$Cov(y_t, y_s) = s\sigma^2$$

En este caso, la dependencia temporal se manifiesta en la varianza y en la covarianza.

Variable con las dos tendencias

El modelo es,

$$y_t = \delta + y_{t-1} + u_t = \delta t + y_0 + \sum_1^t u_i \qquad (1.2.5)$$

Los momentos serán

$$Ey_t = \gamma_0 + \gamma_1 t$$
$$Var(y_t) = t\sigma^2$$
$$Cov(y_t, y_s) = s\sigma^2$$

La dependencia temporal afecta tanto a la media como a la varianza y a las covarianzas.

Para analizar las implicaciones de la no estacionariedad consideremos la variable que solo tiene tendencia estocástica.

Lo primero que observamos es que, conforme el tamaño muestral se hace grande, $T \to \infty$, entonces y_∞ no tiene una distribución bien definida con momentos finitos porque, como ya hemos visto,

$$Var(y_t) = t\sigma^2 \xrightarrow{T\to\infty} \infty$$

Tenemos que normalizar para neutralizar la influencia del tiempo. Primero, tener en cuenta que t=rT con $0 \le r \le 1$, de forma que la varianza puede escribirse como

$$Var(y_t) = rT\sigma^2$$

A partir de esta expresión se puede pensar en una normalización dividiendo la variable por la raíz cuadrada del tamaño muestral de tal forma que,

$$Var(\frac{y_t}{T^{1/2}}) = r\sigma^2$$

Hemos neutralizado el efecto del tiempo pero no hemos llegado exactamente a un marco estacionario porque cada elemento del proceso normalizado tiene una varianza diferente, aunque finita. Recordar que, en el marco estacionario, todos los elementos tenían la misma varianza. Por eso, las convergencias que se van a derivar para el proceso normalizado van a ser diferentes de las obtenidas en el caso estacionario. En concreto, para la variable no estacionaria en varianza normalizada la convergencia es

$$\frac{y_t}{T^{1/2}} \xrightarrow[d]{T\to\infty} N(0, r\sigma^2) \cong \sigma W(r)$$

W(r) es una N(0,r) que se llama variable Browniana. En lo que respecta al sumatorio, hay que tener en cuenta que

$$\sum y_t = T^{1/2} T \frac{1}{T} \sum \frac{y_t}{T^{1/2}}$$

Por lo tanto, podemos decir que el sumatorio de los elementos del proceso es $O_p(T^{3/2})$. ¿Tendrá este sumatorio el mismo límite que el obtenido en el caso estacionario? No, porque estamos sumando variables con diferente varianza. El límite al que se llega es una suma de variables en el espacio continuo. El resultado es,

$$\frac{1}{T^{3/2}} \sum y_t \xrightarrow[d]{T\to\infty} \int_0^1 W(r)dr \qquad (1.2.6)$$

De forma similar puede derivarse la distribución a la que tiende la suma de cuadrados de los elementos del proceso. Se tiene que:

$$\sum y_t^2 = TT \frac{1}{T} \sum \frac{y_t^2}{T}$$

Por lo tanto, podemos decir que $\sum y_t^2$ es $O_P(T^2)$, pudiendose escribir,

$$\frac{1}{T^2}\sum y_t^2 \xrightarrow[d]{T\to\infty} \int_0^1 W(r)^2 dr \qquad (1.2.7)$$

Notar la diferencia con lo obtenido en el marco estacionario.

La ampliación de algunos de estos resultados puede verse en el Apéndice.

Contraste de Raíz Unitaria

Decimos que un proceso es integrado de orden cero y lo denotamos como I(0), si, después de quitar los elementos deterministas, es un proceso estacionario e invertible.

Decimos que un proceso es integrado de orden 1 y lo denotamos como I(1) si, después de quitar los elementos deterministas, la primera diferencia del mismo es un proceso I(0). También decimos que el proceso tiene una raíz unitaria o, equivalentemente, que tiene una tendencia estocástica.

Decimos que un proceso es integrado de orden d y lo denotamos como I(d) si, después de quitar los elementos deterministas, la diferencia d-ésima del mismo es un proceso I(0). También decimos que el proceso tiene d raíces unitarias.

Se verá posteriormente la importancia que tiene el conocer el orden de integración de las variables con las que se pretende especificar un modelo econométrico. Ahora vamos a describir un proceso de contraste para determinar el orden de integración de una variable. El punto de partida son las T observaciones y el gráfico con la variable tiempo en el eje de abscisas y, en ordenadas, los valores de la variable que se analiza. Con carácter general podemos encontrarnos con tres gráficos:

Gráfico 1. Los valores de la variable giran en torno a cero, que podemos considerar es la media de la variable. El grado de reversión hacia la media puede ser variable.
Gráfico 2. Los valores de la variable giran en torno a una constante diferente de cero. Tampoco en este caso el grado de reversión está definido.
Gráfico 3. Los valores de la serie van en torno a una pauta creciente o decreciente que, suponemos, es una función líneal del tiempo. Tampoco en este caso la reversión está definida.
Veamos ahora como diseñar el contraste del orden de integración de una serie según sea el gráfico observado.
Si observamos el Gráfico 1 entonces el modelo de contraste que utilizamos es

$$y_t = \rho y_{t-1} + u_t \qquad y_0 = 0 \qquad (1.2.8)$$

En principio, suponemos que u_t es un ruido blanco. La hipótesis nula del contraste es que la variable es integrada de orden 1, es decir que es I(1). En el marco del modelo de contraste la hipótesis nula se puede escribir como,

$$H_0 : \rho = 1$$

La hipótesis alternativa es que la serie es integrada de orden cero, es decir, es I(0) que se puede formular como

$$H_1 : \rho < 1$$

Es una hipótesis compuesta unilateral.
El estadístico de contraste es

$$t = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}} \qquad (1.2.9)$$

Con $\hat{\rho} = \dfrac{\sum y_{t-1} y_t}{\sum y_{t-1}^2}$ y $\hat{\sigma}_{\hat{\rho}} = \sqrt{\dfrac{\hat{\sigma}^2}{\sum y_{t-1}^2}}$

Siendo $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_t^2}{T} = \dfrac{\sum \left(y_t - \hat{\rho} y_{t-1} \right)^2}{T}$

La región crítica del contraste es decir, los valores de t para los que se rechaza la hipótesis nula, viene dada por

$$t < t_{\varepsilon} \qquad (1.2.10)$$

En donde t_{ε} es el punto crítico que cumple

$$\Pr ob.\{t < t_{\varepsilon} / H_0\} = \varepsilon$$

Siendo ε el nivel de significación elegido a priori. Los puntos críticos correspondientes a este caso pueden verse en la Tabla 1 en la columna encabezada por Gráfico 1. En esta columna aparecen varios puntos críticos dependiendo del tamaño muestral.

Normalmente el modelo de contraste que se utiliza no es el escrito en (1.2.8) sino otro equivalente que se obtiene restando a ambos lados de (1.2.8) el valor retardado del proceso, obteniéndose

$$\Delta y_t = \rho^* y_{t-1} + u_t \qquad (1.2.11)$$

Con $\rho^* = \rho - 1$. En este caso, la hipótesis nula se formula como $H_0 : \rho^* = 0$, siendo el estadístico de contraste

$$t^* == \frac{\hat{\rho}^*}{\hat{\sigma}_{\hat{\rho}^*}} \qquad (1.2.12)$$

Con $\hat{\rho}^* = \dfrac{\sum y_{t-1}\Delta y_t}{\sum y_{t-1}^2}$ y $\hat{\sigma}_{\hat{\rho}^*} = \sqrt{\dfrac{\hat{\sigma}^{*2}}{\sum y_{t-1}^2}}$

Los puntos críticos son los mismos comentados previamente. Por ejemplo, si T=100, entonces la hipótesis nula re rechaza si

$$t^* < -1,95$$

Si observamos el gráfico 2 entonces el modelo de contraste es

$$\Delta y_t = \delta_0 + \rho^* y_{t-1} + u_t \quad con \quad y_0 \neq 0 \qquad (1.2.13)$$

La hipótesis nula es la misma, aunque en este caso se puede demostrar que esa hipótesis nula implica también que $\delta_0 = 0$. El estadístico de contraste es el mismo escrito en (1.2.12) pero definiendo los estimadores a partir del modelo (1.2.13),

$$\begin{bmatrix} \hat{\delta}_0 \\ \hat{\rho}^* \end{bmatrix} = \begin{bmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \Delta y_t \\ \sum y_{t-1} \Delta y_t \end{bmatrix}$$

Los puntos críticos para definir la región crítica están recogidos en la Tabla 1 en la columna encabezada por Gráfico 2. Por ejemplo, si T=100, con un nivel de significación del 5% la hipótesis nula es rechazada si

$$t^* < -2,89$$

Tabla 1. Puntos Críticos. Nivel de Significación: 5%.

Tamaño Muestral	Gráfico 1	Gráfico 2	Gráfico 3
50	-1,95	-2,93	-3,50
100	-1,95	-2,89	-3,45
250	-1,95	-2,88	-3,43

Si observamos el gráfico 3 entonces el modelo de contraste es

$$\Delta y_t = \delta_0 + \delta_1 t + \rho^* y_{t-1} + u_t \qquad (1.2.14)$$

La hipótesis nula es la misma, aunque en este caso se puede demostrar que esa hipótesis nula implica también que $\delta_1 = 0$. El estadístico de contraste es el mismo escrito en (1.2.12) pero definiendo los estimadores a partir del modelo (1.2.14)

$$\begin{bmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \\ \hat{\rho}^* \end{bmatrix} = \begin{bmatrix} T & \sum t & \sum y_{t-1} \\ \sum t & \sum t^2 & \sum t y_{t-1} \\ \sum y_{t-1} & \sum t y_{t-1} & \sum y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum \Delta y_t \\ \sum t \Delta y_t \\ \sum y_{t-1} \Delta y_t \end{bmatrix}$$

Los puntos críticos para definir la región crítica están recogidos en la Tabla 1 en la columna encabezada por Gráfico 3. Por ejemplo, si T=100, con un nivel de significación del 5% la hipótesis nula es rechazada si

$$t^* < -3,45$$

Hasta ahora hemos supuesto que la perturbación de los tres modelos es un ruido blanco. Pero si la perturbación u_t esté afectada por **autocorrelación** serial entonces la estrategia tiene que cambiar porque la distribución de probabilidad del estadístico de contraste, t^*, ya no es la misma. Para evitar este problema lo que se hace es estimar los modelos tal como los hemos escrito y contrastar la presencia de autocorrelación utilizando los residuos. Si el resultado del contraste es que hay autocorrelación entonces se añade un retardo de la variable endógena como nuevo regresor estimándose el siguiente modelo, en el caso de observar el Gráfico 3

$$\Delta y_t = \delta_0 + \delta_1 t + \rho^* y_{t-1} + \gamma_1 \Delta y_{t-1} + u_t$$

Se vuelve a contrastar la autocorrelación y si se rechaza la hipótesis nula se añade un segundo retardo y el proceso sigue hasta que se llega a un modelo en el que la hipótesis nula de no autocorrelación no es rechazada.

El proceso de contraste que acabamos de describir se suele denominar **contraste de Dickey-Fuller** por ser estos dos autores los que lo desarrollaron.

Apendice

Resultado 2.6: Considerar un proceso generado como

$$y_t = y_{t-1} + u_t$$

En donde u_t es i.i.d $\left(0,\sigma^2\right)$ e $y_0 = 0$.

Entonces se tiene que:

a) $T^{-1/2}\sum u_t \xrightarrow{d} W(1) = N(0,1)$

b) $T^{-3/2}\sum y_{t-1} \xrightarrow{d} \int_0^1 W(r)dr$

c) $T^{-1/2}\bar{y} \xrightarrow{d} \int_0^1 W(r)dr$

d) $T^{-1}\sum y_{t-1}u_t \xrightarrow{d} \frac{1}{2}\left[W(1)^2 - 1\right]$

e) $T^{-3/2}\sum t\, u_t \xrightarrow{d} W(1) - \int_0^1 W(r)dr$

f) $T^{-2}\sum y_{t-1}^2 \xrightarrow{d} \int_0^1 W(r)^2 dr$

g) $T^{-5/2}\sum t\, y_{t-1} \xrightarrow{d} \int_0^1 rW(r)dr$

h) $T^{-3}\sum t\, y_{t-1}^2 \xrightarrow{d} \int_0^1 r\, W(r)^2 dr$

i) $T^{-(\nu+1)}\sum t^{\nu} \xrightarrow{p} \dfrac{1}{\nu+1}$

Prueba:

a) Es una consecuencia de la aplicación del Teorema Central del Límite

b) Ya demostrado anteriormente en (1.2.6).

c) $T^{-1/2}\bar{y} = T^{-3/2}\sum y_{t-1}$ que es b).

d) Tener en cuenta que:

$$y_t^2 = \left(y_{t-1} + u_t\right)^2 = y_{t-1}^2 + u_t^2 + 2y_{t-1}u_t$$

de donde:

$$y_{t-1}.u_t = \frac{1}{2}\left(y_t^2 - y_{t-1}^2 - u_t^2\right)$$

y sumando:

$$\sum y_{t-1}u_t = \frac{1}{2}\left(y_T^2 - y_0\right) - \frac{1}{2}\sum u_t^2$$

y como: $y_T \sim \mathrm{N}(0,T)$ y suponiendo que $y_0 = 0$:

$$\frac{1}{T}\sum y_{t-1}u_t = \frac{1}{2}\left(\frac{y_T}{\sqrt{T}}\right)^2 - \frac{1}{2}\frac{\sum u_t^2}{T} \xrightarrow{d} \frac{1}{2}\chi^2(1) - \frac{1}{2} = \frac{1}{2}\left[W(1)^2 - 1\right]$$

e) Primeramente vamos a demostrar que:

$$T^{-3/2}\sum t\,u_t = T^{-1/2}\sum u_t - T^{-3/2}\sum y_{t-1} \quad \text{(A.1)}$$

Escribimos:

$$T^{-3/2}\sum^{T} y_{t-1} = T^{-3/2}\big[u_1 + (u_1 + u_2) + (u_1 + u_2 + u_3) + \ldots..$$

$$+(u_1 + u_2 + \ldots + u_{T-1})\ \big] = T^{-3/2}\big[(T-1)u_1 + (T-2)u_2 + \ldots.$$

$$+(T-(T-1))u_{T-1}\ \big] = T^{-3/2}\sum_{1}^{T}(T-t)u_t$$

llegándose así a (A.1).

El resultado se obtiene aplicando a) y b) en (A.1).

f) Tener en cuenta que

$$T^{-2}\sum y_{t-1}^2 = T^{-1}\sum\left(\frac{y_{t-1}}{\sqrt{T}}\right)^2 \xrightarrow{d} \int W(r)^2 dr$$

g) $T^{-5/2}\sum t\, y_{t-1} = T^{-3/2}\sum\left(\frac{t}{T}\right)y_{t-1} \xrightarrow{d} \int_0^1 rW(r)\, dr$ para $r = \frac{t}{T}$.

h) $T^{-3}\sum t\, y_{t-1}^2 = T^{-2}\sum\left(\frac{t}{T}\right)y_{t-1}^2 \xrightarrow{d} \int_0^1 rW(r)^2 dr$.

i) Inductivamente:

$\sum_{t=1}^{T} t = \frac{T(T+1)}{2}$ que requiere una normalización por el factor T^2

$$\sum_{t=1}^{T} t^2 = \frac{T(T+1)(2T+1)}{6}$$ que requiere una normalización por el factor T^3

Referencias

Aznar, A. (2012): **“**Curso de Econometría” Copy Center Digital. Zaragoza.