CAPÍTULO 1. EL MODELO LINEAL GENERAL

1. Introducción

2. Hipótesis del modelo

Las hipótesis en las que vamos a basar el funcionamiento del Modelo Lineal General son las siguientes

1. Suponemos que tenemos una muestra de T valores de la variable aleatoria y. Al mismo tiempo, también conocemos los T valores de las k variables explicativas que vamos a utilizar como regresores del modelo. Es decir, conocemos el vector de observaciones de la variables endógena y' = (y_1, y_2, …, y_T), así como la matriz de regresores:

$$X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1T} & x_{2T} & \cdots & x_{kT} \end{bmatrix}$$

2. El proceso generador de los datos es y = Xβ + u, donde u es un vector de perturbaciones aleatorias que recoge la influencia de todas aquellas variables no incluidas en la matriz X.

3. El vector de perturbaciones aleatorias se distribuye de forma independiente según una variable aleatoria Normal de vector de medias el vector nulo y con matriz de varianzas y covarianzas σ2 I_T, donde I_T es la identidad de orden T. habitualmente esto se representa de la siguiente manera: u ~ niid(0, σ^2 I_T). Esta hipótesis contiene cuatro supuestos sobre el comportamiento de las perturbaciones aleatorias del modelo. En

primer lugar, asumimos que este vector tiene por esperanza matemática un vector de ceros

(E(u) = 0), lo que implica que la influencia media de las variables no incluidas en la especificación es nula. Dicho de otra manera, que la matriz X incluye todas aquellas variables que son relevantes a la hora de explicar el fenómeno determinado por la variable y.

El segundo de los supuestos formulados sobre el vector u es que tiene matriz de varianzas y covarianzas escalar. Para que esto se cumpla es necesario que la varianza de todos los componente del vector de perturbaciones sea la misma: $E\left(u_i^2\right) = \sigma^2$ i = 1, 2, …, T. A este supuesto se le conoce con el nombre de homoscedasticidad y bajo su cumplimiento todos los elementos de la diagonal principal de la matriz de varianzas y covarianzas del vector u son iguales. Su incumplimiento recibe el nombre de heteroscedasticidad: $E\left(u_i^2\right) = \sigma_i^2 \quad \forall i = 1, 2, ..., T$.

Al mismo tiempo, estamos suponiendo que los elementos fuera de la diagonal principal deben ser todos iguales a 0. Esto conlleva que $E(u_i u_j) = 0 \;\forall i \neq j$. A este supuesto se le conoce como de no autocorrelación. Obviamente, a su incumplimiento se le denomina autocorrelación $E(u_i u_j) \neq 0$ para al menos una pareja de valores (i, j), siendo i. $\neq$j.

Por último, también estamos suponiendo que este vector sigue una distribución normal. Dado que los elementos del vector de perturbaciones están incorrelacionados entre sí, al asumir la hipótesis de normalidad, es inmediato comprobar la independencia en la distribución de los mismos.

4. Los supuestos que formulamos sobre la matriz de varianzas y covarianzas son los siguientes. En primer lugar, asumimos que la matriz X es una matriz de tipo determinística. Esto supone que los elementos que la componen no son estocásticos. En ocasiones, sin embargo, este supuesto se puede relajar imponiendo una segunda condición plim $X'X = \Sigma_{XX}$, siendo Σ_{XX} una matriz de momentos finita y no singular. Este supuesto nos viene a decir que la distribución de las variables incluidas en X tiene momentos finitos de primer y segundo orden. Cuando los regresores tienen una naturaleza determinista esta condición no supone ninguna restricción ya que su cumplimiento es inmediato. Por el contrario, si los regresores son de naturaleza estocástica, este segundo supuesto es una restricción muy importante como podremos ver.

5. Suponemos que el vector de parámetros β es constante a lo largo de la muestra. Este supuesto excluye aquellas situaciones en las que los individuos responden de forma diferente ante variaciones de las variables explicativas. Tampoco considera posible aquellos periodos de tiempo en los que por acción de guerras, firma de acuerdos comerciales o efectos similares la respuesta agregada de los agentes cambia ante variaciones de las variables incluidas en la especificación del modelo.

6. El rango de la matriz X'X es k. Dado que la matriz X'X tiene dimensión kxk, este supuesto indica que esta matriz es de rango completo. Por tanto, no existe relaciones lineales perfectas entre los elementos de la matriz X. Este supuesto se denomina habitualmente como ausencia de multicolinealidad perfecta.

En presencia de estos supuestos, se puede demostrar que el vector de observaciones de la variable endógena se distribuye de acuerdo al siguiente vector de medias y matriz de varianzas y covarianzas:

$$E(y) = E(X\beta + u) = E(X\beta) + E(u) = X\beta$$

$$Var(y) = E[(y - E(y)) (y - E(y))'] = E[(X\beta + u - X\beta) (X\beta + u - X\beta)] = E(uu') = \sigma^2 I_T.$$

Si además, consideramos el supuesto de normalidad de las perturbaciones y, dado que el jacobiano de la transformación es igual a la unidad, resulta directo considerar que la variable y sigue una distribución $N(X\beta, \sigma^2 I_T)$. Por tanto, para poder representar el comportamiento de la variable y necesitamos, en primer lugar, las observaciones tanto de esta variable como de los componentes de la matriz X. Una vez conseguidos, es preciso conocer el valor de los elementos incluidos en el vector de parámetros de posición

$\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ y el del parámetro de dispersión σ^2. Para ello debemos estimarlos ya que no son desconocidos. Existen diversos medios para estimar los parámetros de los que depende la distribución de la variable y. El más utilizado es el método mínimo cuadrático ordinario. Este procedimiento implica la minimización de la suma de los cuadrados de los residuos de la estimación. La combinación de valores de los parámetros que minimice esta suma es la estimación mínimo cuadrático ordinaria.

Denotemos como u al vector de residuos, siendo $u = y - X\beta$. A partir de él podemos construir la suma residual como suma de los cuadrados de los elementos del vector u. Así, la suma residual se expresa como $SR = u'u =(y - X\beta)' (y - X\beta)$. Como vemos, esta suma residual depende del vector de observaciones de la variable endógena, de la matriz X y del vector de parámetros de posición. Dado que los dos primeros elementos vienen dados al estudio, el único elemento que puede variar es β. La minimización de la suma residual implica encontrar la combinación de elementos del vector β que minimice esta suma.

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

Para obtener el vector que minimiza esta suma, tan solo es necesario derivar la expresión anterior con respecto al vector b e igualar el resultado a 0:

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

Si resolvemos la igualdad anterior, es inmediato probar que el vector de soluciones es el siguiente:

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

que es la expresión con la que se identifica habitualmente al vector de estimadores mínimo cuadrático ordinarios. En la obtención de la expresión anterior hemos considerado implícitamente que la matriz X’X es no singular, por lo que existe $(X'X)^{-1}$.

A partir de este resultado podemos obtener la estimación MCO del parámetro de dispersión. En concreto, este estimador se define de la siguiente manera:

$$\hat{\sigma}^2 = \frac{SR}{T - k}$$

Las propiedades de este estimador, así como las del vector de estimadores las vamos a analizar a continuación.

Propiedades de los estimadores MCO

La primera propiedad del vector de estimadores MCO es la de insesgadez. Esto implica que el vector de estimadores se distribuye de forma centrada con respecto al vector de medias poblacionales. Para probar esta propiedad, resulta conveniente expresar el vector de estimadores de la siguiente manera:

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

A partir de este resultado, resulta directo probar la insesgadez del vector de estimadores MCO, asumiendo los supuestos formulados en el apartado 2,x, concretamente la determineidad de la matriz X y E(u) = 0.

$$E(\hat{\beta})=E(\beta)+E\left[(X'X)^{-1}X'u\right]=\beta+(X'X)^{-1}X'E[u]=\beta$$

Asimismo, la matriz de varianzas y covarianzas del estimador es la siguiente:

$$\mathrm{Var}(\hat{\beta})=E\left\{\left[\hat{\beta}-E(\hat{\beta})\right]\left[\hat{\beta}-E(\hat{\beta})\right]'\right\}=E\left\{\left[\beta+(X'X)^{-1}X'u-\beta\right]\left[\beta+(X'X)^{-1}X'u-\beta\right]'\right\}=$$

$$=E\left\{\left[(X'X)^{-1}X'u\right]\left[\beta+(X'X)^{-1}X'u\right]'\right\}=E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]=$$

$$=(X'X)^{-1}X'E[uu']X(X'X)^{-1}=(X'X)^{-1}X'\sigma^2 I_T X(X'X)^{-1}=\sigma^2(X'X)^{-1}X'X(X'X)^{-1}=$$

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

donde de nuevo juegan un papel esencial la no estocasticidad de la matriz X y la distribución del vector de perturbaciones, concretamente el hecho de que E(uu') = $\sigma^2 I_T$.

A partir del conocimiento de esta matriz de varianzas y covarianzas se puede demostrar que el vector de estimadores MCO es ELIO. Esa palabra quiere decir que el vector es Lineal, Insesgado y Óptimo, en el sentido de tener menor varianza entre todos los lineales e insesgados. La prueba de esta propiedad, conocida como Teorema de Gauss-Markov, no la incluimos aquí ya que vamos a demostrar que este vector es eficiente, de mínima varianza entre todos los insesgados. Para su prueba debemos comenzar por introducir la propiedad de normalidad del vector de perturbaciones aleatorias, por lo que posponemos su prueba hasta la presentación de la estimación máximo verosímil.

En cuanto a las propiedades del estimador del parámetro de dispersión, podemos empezar demostrando su insesgadez. Conviene para ello, expresar el vector de residuos como una combinación lineal del vector de perturbaciones aleatorias:

$$\hat{u} = y - X\hat{\beta} = X\beta + u - X\hat{\beta} = u - X\left(\hat{\beta} - \beta\right) = u - X\left(X'X\right)^{-1} X'u = M\,u$$

donde M es una matriz cuadrada de orden T, definida como $M = I_T - X\,(X'X)^{-1}\,X$. Esta matriz es simétrica:

$$M' = [I_T - X\,(X'X)^{-1}\,X']' = [I_T - X\,(X'X)^{-1}\,X'] = M$$

idempotente:

$$M\,M = [I_T - X\,(X'X)^{-1}\,X']\,[I_T - X\,(X'X)^{-1}\,X'] = I_T - X\,(X'X)^{-1}\,X' - X\,(X'X)^{-1}\,X' +$$

$$+ X\,(X'X)^{-1}\,X'\,X\,(X'X)^{-1}\,X' = I_T - X\,(X'X)^{-1}\,X' - X\,(X'X)^{-1}\,X' + X\,(X'X)^{-1}\,X' =$$

$$= I_T - X\,(X'X)^{-1}\,X' = M$$

Por último, es sencillo probar que la matriz M tiene rango T-k. Dado que M es simétrica e idempotente, es cierto que el rango coincide con la traza de la matriz. Esta traza es igual a:

$$\text{tr}\,M = \text{tr}[I_T - X\,(X'X)^{-1}\,X'] = \text{tr}\,I_T - \text{tr}\,[X\,(X'X)^{-1}\,X'] = \text{tr}\,I_T - \text{tr}\,[\,(X'X)^{-1}\,X'X] = \text{tr}\,I_T -$$

tr I_k = T-k = rg M.

Una vez que conocemos estas propiedades, es directo probar la insesgadez del estimador MCO de la varianza:

$$E\,\hat{\sigma}^2 = E\left(\frac{\hat{u}'\hat{u}}{T-k}\right) = \frac{1}{T-k} E(u'M'Mu) = \frac{1}{T-k} E(u'Mu) = \frac{1}{T-k} tr[E(u'Mu)] =$$

$$= \frac{1}{T-k} E[tr(u'Mu)] = \frac{1}{T-k} E[tr(Muu')] = \frac{1}{T-k} tr[E(Muu')] = \frac{1}{T-k} tr[M\,E(uu')] =$$

$$= \frac{1}{T-k} tr\left[M\sigma^2 I_T\right] = \frac{\sigma^2}{T-k} tr[M] = \frac{\sigma^2}{T-k}(T-k) = \sigma^2$$

En la demostración anterior hemos hecho uso de diversas propiedades del operador traza matemática: la traza de un escalar es igual al propio escalar, el operador traza y el operador esperanza son conmutables entre sí y tr (ABC) = tr (BCA).

A pesar de que el estimador MCO de la varianza es insesgado, no podemos decir que presente buenas propiedades, al menos en muestras finitas. A esta afirmación nos conduce el hecho de que este estimador no sea ni ELIO (no puede ser, por cuanto es evidente su no linealidad), ni tampoco es eficiente en muestras finitas. Por tanto, la distribución de este estimador está centrada sobre el verdadero valor del parámetro, pero sin poder afirmar muchas cosas sobre la dispersión de la distribución.

Estimación máximo-verosímil

Todos los resultados que hemos obtenido en el apartado anterior son válidos independientemente de cuál sea la distribución de las perturbaciones. Por tanto, no hemos hecho uso del supuesto de normalidad del vector de perturbaciones. SI lo tenemos en cuenta, podemos estimar los parámetro del modelo por un método distinto: estimación máximo verosímil. El fundamento de esta técnica de estimación es encontrar aquella combinación de parámetros que maximice la probabilidad de obtener la muestra observada, es decir, aquel conjunto de parámetros que sea más verosímil o probable de haber generado la muestra observada del vector y.

Como vimos, la introducción del supuesto de normalidad de las perturbaciones implica que el vector y siga una distribución N(XB, s2). Esto suponen que la función de densidad de cada uno de los componentes del vector y se puede expresar de la siguiente manera:

$$f\left(y_i / x_i, \beta, \sigma^2\right) = \left(2\pi\sigma^2\right)^{-T/2} \exp\left[-\frac{\left(y_i - x_i^{'}\beta\right)\left(y_i - x_i^{'}\beta\right)}{2\sigma^2}\right]$$

donde $x_i' = (x_{1i}, x_{2i}, \ldots, x_{ki})$. Dado que todas las observaciones son seleccionadas de forma independiente mediante un muestreo aleatorio simple, la función densidad conjunta del vector y no es sino el producto de las T funciones de densidad de cada unos de los componentes del vector. Por tanto la función de densidad conjunta es igual a:

$$f\left(y / X, \beta, \sigma^2\right) = f(y_1, y_2, \ldots, y_T) = f(y_1) f(y_2) \ldots f(y_T) =$$

$$= \left(2\pi\sigma^2\right)^{-T/2} \exp\left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right]$$

Una vez conocidas la función de densidad conjunta, podemos obtener la función de máxima verosimilitud, que denotaremos como L.

$$L\left(\beta, \sigma^2 / y, X\right) = \left(2\pi\sigma^2\right)^{-T/2} \exp\left[-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right]$$

Como se puede apreciar, desde el punto de vista matemático la formula anterior coincide con. Interpretación distinta.

La obtención de los estimadores MV implica la maximización de la función de verosimilitud. Esto conlleva que debemos derivar la expresión () con respecto al vector b y al parámetro s2. Transformación a la función log-verosímil por que sí.

$$\ln L\left(\beta, \sigma^2 / y, X\right) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\left(\sigma^2\right) - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}$$

Las derivadas primeras con respecto a b y s2 son

$$\frac{\partial L}{\partial \beta} = -\frac{1}{2\tilde{\sigma}^2}\left[-2X'y + 2X'X\tilde{\beta}\right] = \frac{X'y - X'X\tilde{\beta}}{\tilde{\sigma}^2}$$

$$\frac{\partial L}{\partial \sigma^2} = -\frac{T}{2}\,\frac{1}{\tilde{\sigma}^2} + \frac{2\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right)}{4\tilde{\sigma}^4} = -\frac{T}{2\tilde{\sigma}^2} + \frac{\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right)}{2\tilde{\sigma}^4}$$

Si igualamos las expresiones anteriores a 0 y con un poco de simplificación, obtenemos los estimadores máximo-verosímiles:

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

$$\tilde{\sigma}^2 = \frac{\left(y - X\tilde{\beta}\right)'\left(y - X\tilde{\beta}\right)}{T}$$

Observamos que, en primer lugar, el vector de estimadores MV coincide con el vector de estimadores MCO. No ocurre lo mismo con el estimador del parámetro de dispersión. Mientras que ahora en el denominador aparece el tamaño muestral, en el caso de la estimación MCO aparecía los grados de libertad del modelo.

Una vez definidos los estimadores, debemos pasar a analizar sus propiedades. EN el caso de $\tilde{\beta}$ no hace falta por cuanto sus propiedades son idénticas a las estudiadas para el vector $\hat{\beta}$. En cambio, los cambios si que son más sustanciales para el estimador MV del parámetro de dispersión, por cuanto es sencillo probar que es sesgado:

$$E\left(\tilde{\sigma}^2\right)=E\left(\frac{SR}{T}\right)=\frac{T-k}{T}E\left(\frac{SR}{T-k}\right)=\frac{T-k}{T}E\left(\hat{\sigma}^2\right)=\frac{T-k}{T}\sigma^2$$

Es asimismo evidente que el sesgo va a desaparecer asintóticamente. Por tanto, en virtud de [] debemos indicar que el estimador MV del parámetro de dispersión no va a ser ni ELIO, ni eficiente.

Precisamente, lo que vamos a abordar ahora es la cuestión de la eficiencia de los estimadores tanto MCO como MV. Recordemos que un estimador $\hat{\theta}$ del parámetro θ es eficiente siempre que sea insesgado y cov($\tilde{\theta}$)- cov($\hat{\theta}$)≥ 0. En el caso de que $\hat{\theta}$ sea un vector de estimadores insesgado, la versión multivariante de eficiencia necesita que cov($\tilde{\theta}$)- cov($\hat{\theta}$) sea una matriz semidefinida positiva. La cota de Cramer-Rao no proporciona una condición suficiente, pero no necesaria, para que un estimador sea eficiente. Estos dos autores probaron que la varianza de todo estimador insesgado tiene una cota inferior. Entonces, basta con comparar la matriz de varianzas y covarianzas del vector de estimadores con dicha cota para probar la eficiencia del mismo. Esta se obtiene de la siguiente manera:

$$I(\theta)=-E\left(\frac{\partial^2 L}{\partial\theta\,\partial\theta'}\right)$$

donde L es el logaritmo neperiano de la función de verosimilitud. Si la varianza del estimador coincide con este valor cota podemos asegurar que no existirá ningún otro estimador insesgado que presente una varianza inferior, por lo que será eficiente.

Para formar la matriz de información necesitamos las derivadas segundas de la función log-verosímil con respecto al conjunto de parámetros:

$$\frac{\partial^2 L}{\partial \beta\, \partial \beta'} = -\frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial \beta\, \partial \sigma^2} = -\frac{X'y - X'X\beta}{\sigma^4}$$

$$\frac{\partial L}{\left(\partial \sigma^2\right)^2} = \frac{T}{2\sigma^4} - \frac{(y - X\beta)'(y - X\beta)}{2\sigma^6}$$

Por tanto, la matriz de información será:

$$I\left(\beta, \sigma^2\right) = -E\begin{bmatrix} -\dfrac{X'X}{\sigma^2} & -\dfrac{X'y - X'X\beta}{\sigma^4} \\ -\dfrac{y'X - \beta'X'X}{\sigma^4} & \dfrac{T}{2\sigma^4} - \dfrac{(y - X\beta)'(y - X\beta)}{2\sigma^6} \end{bmatrix} =$$

$$= \begin{bmatrix} \dfrac{X'X}{\sigma^2} & E\left(\dfrac{X'y - X'X\beta}{\sigma^4}\right) \\ E\left(\dfrac{y'X - \beta'X'X}{\sigma^4}\right) & -\dfrac{T}{2\sigma^4} + \dfrac{E\left[(y - X\beta)'(y - X\beta)\right]}{2\sigma^6} \end{bmatrix}$$

Si tenemos en cuenta que:

$$E(X'y - X'X\beta) = E\left[X'(y - X\beta)\right] = X'E\left[(X\beta + u - X\beta)\right] = X'E(u) = 0$$

$$E\left[(y - X\beta)'(y - X\beta)\right] = E(u'u) = E\left(\sum_{i=1}^{T} u_i\right) = \sum_{i=1}^{T} E(u_i) = \sum_{i=1}^{T} \sigma^2 = T\sigma^2$$

La matriz de información se convierte con unas simples operaciones en:

$$I\left(\beta, \sigma^2\right) = \begin{bmatrix} \dfrac{X'X}{\sigma^2} & 0 \\ 0 & \dfrac{T}{2\sigma^4} \end{bmatrix}$$

La inversa de esta matriz es la cota de Cramer-Rao para el conjunto de estimadores que nos ocupa:

$$\begin{bmatrix} \sigma^2 (X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{T} \end{bmatrix}$$

EL siguiente paso es comparar las varianzas de los estimadores con cada uno de los bloques que componen esta matriz. Comenzando por el vector de estimadores MCO y MV vemos como su matriz de varianzas y covarianzas coincide con la cota de Cramer-Rao. Por tanto estos son eficientes.

Para estudiar la eficiencia del estimador MCO del parámetro de dispersión debemos comenzar por obtener su varianza. Bajo el supuesto de normalidad de las perturbaciones se puede demostrar que:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{u'Mu}{\sigma^2} \sim \chi^2_{T\text{-}k}$$

Asimismo, conocemos que, en general, $E\left(\chi^2_m\right) = m$ y $Var\left(\chi^2_m\right) = 2\,m$. Entonces, es cierto que:

$$\mathrm{Var}\left(\hat{\sigma}^2\right)=\mathrm{Var}\left(\frac{\hat{u}'\,\hat{u}}{T-k}\right)=\frac{\sigma^4}{(T-k)^2}\mathrm{Var}\left(\frac{\hat{u}'\,\hat{u}}{\sigma^2}\right)=\frac{\sigma^4}{(T-k)^2}\mathrm{Var}\left(\chi^2_{T-k}\right)=\frac{\sigma^4}{(T-k)^2}2(T-k)=$$

$$=\frac{2\,\sigma^4}{T-k}$$

Comparando el valor anterior con el segundo elemento de la matriz de valores cota vemos que es superior, por lo que no el estimador MCO de la varianza no es eficiente. Otro tanto podemos decir del estimador MV de la varianza, dado que es sesgado. A pesar de que los estimadores del parámetro de dispersión no son eficientes en muestras finitas, debemos indicar que sí que lo son asintóticamente.

XX. Inferencia

Una vez que hemos estimado los parámetros de los que depende la distribución de la variable y, podemos pasar a analizar diversas hipótesis. Habitualmente, los primeros contrastes que debemos efectuar son aquellos que analizan la significatividad, tanto individual como conjunta, de los parámetros de posición del modelo. Esto nos dará una idea de la posible introducción de variables irrelevantes en la especificación anterior. Además, la teoría económica nos ofrece muchas veces cuestiones que son interesantes estudiar. Un ejemplo de ello es la denominada paridad del poder adquisitivo. Según esta teoría, muy extendida dentro de la macroeconomía abierta, las variaciones entre los diferenciales de precios de dos países son directamente absorbidas por el tipo de cambio nominal de las monedas de estos países. Para contrastar esta teoría, por ejemplo, podríamos realizar una regresión entre el diferencial de precios y el tipo de cambio nominal y contrastar si el parámetro asociado a esta variable es la unidad. Del mismo modo, también podríamos estudiar si los incrementos en la inflación de una economía esperada son recogidos directamente por el tipo de interés nominal de la misma. Esto se conoce como efecto Fisher. Otros ejemplos también muy utilizados son la existencia de rendimientos constantes a escala en una empresa.

Para realizar este tipo de contrastes, debemos comenzar por la definición de un estadístico general que englobe un gran número de opciones, definir cuál es la hipótesis nula y calcular la distribución del estadístico bajo esta hipótesis nula. El caso más general es aquel en el que queremos estudiar algunas hipótesis sobre los elementos del vector de parámetros de posición. Supongamos que queremos contrastar la hipótesis nula

H_o: RB = c, donde β es un vector columna de orden k y R es una matriz de orden r x k. Como podemos comprobar, en la hipótesis anterior están recogidas todo el conjunto de hipótesis lineales. Todo depende de la correcta selección de la matriz R. Así, por ejemplo, sí queremos contrastar la hipótesis nula individual H_o: $\beta_i = 0$, donde i es un valor entre 1 y k, la matriz R deberá adoptar la siguiente forma:

$$R = (0, 0, \ldots, 0, \underset{i}{1}, 0, \ldots 0)$$

por lo que al multiplicar la matriz R por el vector β el resultado es β_i. Si hacemos r=0, tenemos los requisitos para contrastar la hipótesis nula de interés. Si, por el contrario, queremos contrastar una hipótesis como la de rendimientos constantes a escala, la matriz R debe adoptar la forma:

$$R = (0, 1, 1)$$

de esta forma, $R\beta = \beta_2 + \beta_3$. Haciendo r =1, podemos ahora contrastar esta hipótesis.

Por último, si estamos interesados en contrastar una hipótesis conjunta, por ejemplo la significatividad global de las variables explicativas, la matriz R debería ser igual a:

$$R = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

y posteriormente hacer r=0.

Una vez que hemos visto como se pueden contrastar las diversas hipótesis, vamos a calcular un estadístico que sea válido para el caso general. Para ello debemos comenzar por recordar que el vector de estimadores de parámetros de posición sigue una distribución $N(\beta,\ \sigma^2\ (X'X)^{-1})$. Por tanto, una combinación lineal del vector de estimadores $R\hat{\beta}$ sigue una distribución $N(R\beta,\ \sigma^2\ R(X'X)^{-1}R')$ ya que:

$$E(R\hat{\beta}) = R\ E(\hat{\beta}) = R\beta$$

$$Var(R\hat{\beta}) = E\{[(R\hat{\beta} - E(R\hat{\beta})]\ \ [(R\hat{\beta} - E(R\hat{\beta})]'\} = E\{[R(\hat{\beta} - \beta)]\ \ [R(\hat{\beta} - \beta)]'\} =$$

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

Con una simple operación es sencillo probar que, bajo la hipótesis nula H_o: $R\beta = r$, la combinación $R\hat{\beta} - r$ sigue una distribución $N(0,\ \sigma^2\ R\ (X'X)^{-1}\ \ R')$. De aquí podemos pasar a:

$$\frac{\left(R\hat{\beta} - r\right)\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta} - r\right)'}{\sigma^2} \propto \chi^2_r$$

La demostración de este último resultado es tediosa, pero no difícil. Debemos comenzar expresando el numerador del cociente anterior como una forma cuadrática del vector de perturbaciones aleatorias:

$$\begin{aligned}&\left(R\hat{\beta}-r\right)' \left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)=\left(R\hat{\beta}-R\beta\right)' \left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-R\beta\right)=\\ &=\left(R\beta+R(X'X)^{-1}X'u-R\beta\right)' \left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\beta+R(X'X)^{-1}X'u-R\beta\right)=\\ &=u'X(X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R\right]^{-1}R(X'X)^{-1}X'u=u'Bu\end{aligned}$$

donde la matriz B es cuadrada de orden T y tiene las siguientes propiedades:

- Simetría: $B'=X(X'X)^{-1}R'\left[R\left(X'X\right)^{-1}R\right]^{-1}R(X'X)^{-1}X'=B$

- Idempotencia:

$$B\,B = X(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1}R(X'X)^{-1}\underline{X'\ \ X(X'X)^{-1}}R'[R\,(X'X)^{-1} R']^{-1} R(X'X)^{-1} X'$$

$$= X(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1}\underline{R(X'X)^{-1}R'[R\,(X'X)^{-1} R']^{-1}} R\,(X'X)^{-1} X' =$$

$$= X(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} R\,(X'X)^{-1} X' = B$$

- tr B = rg B = r

$$\mathrm{tr}\,B = \mathrm{tr}\{X(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} R\,(X'X)^{-1} X'\} = \mathrm{tr}\{[R(X'X)^{-1} R']^{-1} R\,(X'X)^{-1} \underline{X'\,X(X'X)^{-1}} R'\}= \mathrm{tr}\{[R(X'X)^{-1} R']^{-1} R\,(X'X)^{-1} R'\}= \mathrm{tr}\,(I_r) = r$$

Si asumimos que el vector u se distribuye según $N(0, \sigma^2 I_T)$, dado que la matriz B es simétrica, idempotente y de rango r, se puede demostrar que:

$$\frac{u'Bu}{\sigma^2}\propto\chi^2_r$$

Desgraciadamente, el cociente anterior depende de σ^2 que es un parámetro que desconocemos. Para eliminar esa dependencia, podemos dividir [] por el cociente entre la suma residual y la varianza de la perturbación. Como vimos anteriormente, este estadístico sigue una distribución del tipo x2, por lo que el estadístico resultante podría seguir una distribución F, siempre y cuando las dos formas cuadráticas del vector u sean independientes. Pero esta condición se cumple ya que:

$$B\,M = X(X'X)^{-1} R'[R(X'X)^{-1} R']^{-1} R\,(X'X)^{-1} X'\,M = 0$$

por cuanto la matrices X y M son ortogonales. En consecuencia, el estadístico general que nos va a permitir contrastar las diversas hipótesis es el siguiente:

$$F\frac{\dfrac{\left(R\hat{\beta}-r\right)\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)'}{\sigma^2}}{\dfrac{\hat{u}'\,\hat{u}}{\sigma^2}}\,\frac{T-k}{r}=\frac{\left(R\hat{\beta}-r\right)\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)'}{r\dfrac{\hat{u}'\,\hat{u}}{T-k}}=$$

$$=\frac{\left(R\hat{\beta}-r\right)\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)'}{r\,\hat{\sigma}^2}\propto F_{r,\,T-k}$$

Dado un nivel de significación (que habitualmente será del 5%), podemos encontrar los valores críticos de una distribución F de r grados de libertad en el numerador y T-k grados de libertad en el denominador. Entonces, si $F \le F^{0'05}_{r,\,T-k}$, podemos decir que no hay suficiente evidencia empírica para rechazar la hipótesis nula. Por el contrario, si $F > F^{0'05}_{r,\,T-k}$, la muestra disponible no sustenta la hipótesis nula, por lo que la rechazaremos.

Ahora podemos particularizar el estadístico F a una serie de casos particulares de gran interés en el análisis econométrico. Una primera hipótesis que suele ser analizada con asiduidad es si las variables explicativas son conjuntamente significativas. A este contraste se le denomina análisis de la varianza. Suponiendo que tenemos un modelo cuyo primer regresor es un término independiente, por lo que la primera columna de la matriz X no es sino un vector de unos. La matriz X la podemos particionar de la siguiente manera $X = [X_1, X_2] = [\mathit{1}, X_2]$, donde *1* es un vector de unos de orden T y X_2 es una matriz de orden Tx (k-1) que recoge las observaciones de las k-1 variables explicativas del modelo. De acuerdo con esta partición, el vector de parámetros de posición lo podemos expresar como $\beta = (b_1, b_2')'$, donde $b_2 = (\beta_2, \beta_3, \ldots, \beta_{k-1})'$. Entonces, el contraste de la significación global de las k-1 variables explicativas se puede realizar sin más que estudiar la veracidad de la hipótesis nula H_o: $R\,\beta = 0$, donde $R = [0_{k-1}, I_{k-1}]$, siendo 0_{k-1} un vector columna de ceros de orden k-1. De acuerdo a estas restricciones, el estadístico F adopta la siguiente forma:

$$F=\frac{\left(R\hat{\beta}-r\right)\left[R\left(X'X\right)^{-1}R\right]^{-1}\left(R\hat{\beta}-r\right)'}{r\hat{\sigma}^2}=\frac{\hat{b}_2'\left[[0_{k-1},I_{k-1}]\begin{bmatrix}1'1 & 1'X_2\\ X_2' & X_2'X_2\end{bmatrix}^{-1}\begin{bmatrix}0_{k-1}\\ I_{k-1}\end{bmatrix}\right]^{-1}\hat{b}_2}{r\hat{\sigma}^2}=$$

Dado que el segundo elemento de la diagonal principal de la matriz $(X'X)^{-1}$ puede demostrarse que es igual X2' Q X2, donde $Q = I_T - \frac{1}{T}1'1$, el estadístico anterior se puede expresar así:

$$F=\frac{\hat{b}_2'\;X_2'\,Q\,X_2\;\hat{b}_2}{r\hat{\sigma}^2}$$

Bajo la hipótesis nula, este estadístico sigue una distribución $F_{k-1,\,T-k}$. La interpretación de este estadístico no es directa a simple vista. No obstante, si tenemos en cuenta que el numerador no es sino la suma explicada del modelo y que en el denominador aparece la suma residual, podemos decir que este estadístico mide la bondad del ajuste. Cuanto mayor sea la suma explicada en comparación con la suma residual, con mayor facilidad rechazaremos la hipótesis nula de no significatividad global de los parámetros de posición asociados a las variables explicativas del modelo.

Contrastes de significación individual

Otro de los contrastes que habitualmente se realizan es el análisis de la significatividad individual de los parámetros de posición. En realidad, lo que hacemos es estudiar si todos los regresores en la especificación empírica del modelo son individual significativos. Si llegamos a la conclusión de que todos las variables son individualmente significativas, el modelo está correctamente especificado. En otro caso, aquellas que no son significativas debería ser excluidas de la especificación.

El estadístico que debemos utilizar ahora es algo diferente. Para empezar, la hipótesis nula es individual y no conjunta. En general, podemos estar interesados en estudiar la significatividad del parámetro i-ésimo, lo que se puede explicitar así: $H_o : \beta_i = 0$. Por tanto, R ya no será una matriz, sino un vector fila de dimensión k que presenta k-1 ceros más un valor unitario en el elemento i-ésimo. Por tanto, $R\hat{\beta}$ será igual ahora a $\hat{\beta}_i$, lo mismo que $R\beta$ será β_i, que toma el valor 0 bajo la hipótesis nula. Bajo estas condiciones, el estadístico de contraste toma la siguiente forma:

$$F = \frac{\hat{\beta}_i \left[(0 \cdots 1 \cdots 0)\,(X'X)^{-1} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \right]^{-1} \hat{\beta}_i}{\hat{\sigma}^2} = \frac{\hat{\beta}_i^2 \left[(X'X)_{ii}^{-1} \right]^{-1}}{\hat{\sigma}^2} = \frac{\hat{\beta}_i^2}{\hat{\sigma}^2\,(X'X)_{ii}^{-1}} = \frac{\hat{\beta}_i^2}{V\hat{a}r\left(\hat{\beta}_i\right)} \sim F_{1,T\text{-}k}$$

ya que $(X'X)_{ii}^{-1}$ representa el elemento i-ésimo de la diagonal principal de la matriz $(X'X)^{-1}$. Este estadístico se puede simplificar si tenemos en cuenta que una distribución F de un grado de libertad en el numerador y T-k grados de libertad en el denominador, es equivalente al cuadrado de una distribución t de Student de T-k grados de libertad. Por tanto, podemos realizar el contraste de significación individual de los parámetros de posición del modelos a partir del siguiente estadístico:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2\,(X'X)_{ii}^{-1}}} = \frac{\hat{\beta}_i}{D.\hat{T}.\left(\hat{\beta}_i\right)} \sim t_{T\text{-}k}$$

donde $D.\hat{T}.\left(\hat{\beta}_i\right)$ es la desviación típica estimada del estimador del parámetro de posición i-ésimo. Entonces, dado un nivel de significación α, podemos tomar el valor crítico teórico de esta distribución y compararlo con el valor empírico del estadístico. Si resulta cierto que $t_{\hat{\beta}_i} \leq t_{T-k}^{\alpha/2}$, entonces concluiremos que la hipótesis nula H_o: $\beta_i = 0$ es cierta. Por tanto, deberemos interpretar que la variable i-ésima no contribuye individualmente a la explicación del fenómeno que queremos analizar, aconsejándose su exclusión de la especificación del modelo. Por contra, si $t_{\hat{\beta}_i} > t_{T-k}^{\alpha/2}$, rechazaremos la hipótesis nula y consideraremos que la variable está actuando de forma correcta en la especificación del modelo.

Contraste de hipótesis a partir de las sumas residuales

En las secciones anteriores hemos visto como podemos utilizar los diversos estadísticos para contrastar algunas hipótesis nulas de interés. Lo que vamos a realizar a continuación es una simple reparametrización del estadístico general de contraste, de forma que este quede en función de las sumas residuales procedentes de la estimación de dos modelos. El primero es el modelo original y el segundo el resultante de introducir en este la hipótesis nula.

Supongamos que tenemos el modelo original $y = X\,\beta + u$, donde X es una matriz de orden Txk e y es el vector fila de observaciones de la variable que queremos estudiar. Si lo estimamos por mínimos cuadrados ordinarios, la suma residual resultante la representamos como $\hat{u}'\hat{u}$. Supongamos que queremos estudiar si la hipótesis nula $R\beta = b_2 = 0$ es cierta o no, siendo b_2 un conjunto de s parámetros del vector β. Si introducimos esta restricción en el modelo anterior nos queda $y = X_1\, b_1 + u_1$. Si lo estimamos, los residuos que resultan son $\hat{u}_1 = y - X_1\,\hat{b}_1 = y - X_1(X_1'X_1)^{-1}\,X_1\,y = M_1\,y$, donde $M_1 = I_T - X_1(X_1'X_1)^{-1}\,X_1$.

Multipliquemos ahora el modelo original por la matriz M_1, con lo que queda:

$$M_1\,y = M_1\,X\,\hat{\beta} + u = M_1\,X_1\,\hat{b}_1 + M_1\,X_2\,\hat{b}_2 + M_1\,\hat{u} = M_1\,X_2\,\hat{b}_2 + M_1\,\hat{u}$$

dada la ortogonalidad entre X_1 y M_1. Si tenemos en cuenta, además, que $M_1\,\hat{u} = \hat{u}$, por cuanto:

$$M_1\,\hat{u} = [I_T - X_1(X_1'X_1)^{-1}\,X_1]\;\hat{u} = \hat{u} - X_1(X_1'X_1)^{-1}\,X_1\,\hat{u}$$

siendo $X_1\,\hat{u} = 0$. Entonces, la expresión anterior queda como:

$M_1 y = M_1 X_2 \hat{b}_2 + \hat{u}$

Entonces, la suma residual del modelo resultante de introducir la restricción que queremos estudiar se puede expresar como:

$$\hat{u}_1' \hat{u}_1 = y' M_1' M_1 y = y' M_1 y = \left(M_1 X_2 \hat{b}_2 + \hat{u}\right)' \left(M_1 X_2 \hat{b}_2 + \hat{u}\right) =$$

$$= \hat{b}_2' X_2' M_1' M_1 X_2 \hat{b}_2 + \hat{b}_2' X_2' M_1' \hat{u} + \hat{u}' M_1 X_2 \hat{b}_2 + \hat{u}'\hat{u} = \hat{b}_2' X_2' M_1 X_2 \hat{b}_2 + \hat{u}'\hat{u}$$

Por tanto, resulta cierto que $\hat{b}_2' X_2' M_1 X_2 \hat{b}_2 = \hat{u}_1' \hat{u}_1 - \hat{u}'\hat{u}$, por lo que podemos contrastar la hipótesis nula H_o: $b_2 = 0$, a partir del siguiente estadístico:

$$F = \frac{\hat{u}_1' \hat{u}_1' - \hat{u}'\hat{u}}{s\hat{\sigma}^2}$$

Luego, tan solo necesitamos las sumas residuales del modelo original y del modelo restringido. La interpretación de este estadístico se hace en los siguientes términos. Supongamos que las s variables sobre las que realizamos el contraste no contribuyen nada a la explicación de la variable y. En ese caso, el incremento en la suma residual producida por su exclusión de la especificación inicial del modelo no será muy elevada. Por lo que el numerador de [] se aproximará hacia 0. La conclusión final a la que llegaríamos es que la hipótesis nula es cierta, lo cuál es correcto dado nuestro planteamiento. Analicemos el caso contrario. Imaginemos que las variables sobre las que hacemos el contraste sí que explican parte del comportamiento de la variable y. Cuanto mayor sea este porcentaje de explicación del fenómeno, mayor será la suma residual del modelo restringido en función de la del modelo original. Por tanto, el numerador del estadístico tenderá a tomar valores elevados y más sencillo será rechazar la hipótesis nula. De nuevo, el estadístico nos conduce al resultado correcto.

Una cuestión: esto es valido en general, aunque no lo demostremos. Ejemplo PPP

Errores de Especificación

Supongamos que el proceso generador de los datos viene definido por la siguiente relación: y = Xb + u, donde la matriz X la particionamos de la siguiente manera $X=[X_1, X_2]$. De acuerdo a esta partición, el vector de parámetros de posición queda $\beta = (b_1, b_2)$. Imaginemos ahora que el investigador no es capaz de determinar correctamente el conjunto de variables explicativas que debe incluirse en la matriz y, erróneamente, se plantea el modelo y = X1 b1 + u1. El vector de estimadores MCO será ahora $\hat{b}_1 = (X_1'X_1)^{-1} X_1$ y. Las propiedades de este vector son las siguientes:

$E(\hat{b}_1) = E[(X_1'X_1)^{-1} X_1' \, y] = E[(X_1'X_1)^{-1} X_1'(X_1 b_1 + X_2 b_2 + u)] = E[(X_1'X_1)^{-1} X_1'X_1 b_1 + (X_1'X_1)^{-1} X_1' X_2 b_2 + (X_1'X_1)^{-1} X_1' u] = b_1 + (X_1'X_1)^{-1} X_1' X_2 b_2 + (X_1'X_1)^{-1} X_1' E(u) = b_1 + (X_1'X_1)^{-1} X_1' X_2 b_2$

Por tanto, en general es cierto que $E(\hat{b}_1) \neq b_1$, por lo que este vector de estimadores será sesgado, excepto en el caso en el que $X_1'X_2 = 0$. Por otro lado, la matriz de varianzas y covarianzas del vector de estimadores $\hat{b}_1$ será la siguiente:

$$Var(\hat{b}_1) = E\{[\hat{b}_1 - E(\hat{b}_1)][\hat{b}_1 - E(\hat{b}_1)]'\} = E\{[(X_1'X_1)^{-1} X_1' u] [(X_1'X_1)^{-1} X_1' u]'\} =$$

$$= E[(X_1'X_1)^{-1} X_1' u u' X_1(X_1'X_1)^{-1}] = (X_1'X_1)^{-1} X_1' E[u u'] X_1 (X_1'X_1)^{-1} =$$

$$= (X_1'X_1)^{-1} X_1' \sigma^2 I_T X_1 (X_1'X_1)^{-1} = \sigma^2 (X_1'X_1)^{-1} \underline{X_1'X_1 (X_1'X_1)^{-1}} = \sigma^2 (X_1'X_1)^{-1}$$

Como podemos comprobar, aunque el vector de estimadores $\hat{b}_1$ es sesgado, su varianza es la misma que en el caso en el que no se produce una omisión de un subconjunto de variables relevantes.

La segunda de las posibilidades que vamos a contemplar es aquella en la que el investigador introduce un conjunto de variables que son irrelevantes a la hora de abordar el estudio de un fenómeno económico. En este caso, el proceso generador de los datos viene dado por $y = X_1 b_1 + u$, mientras que el investigador plantea la siguiente especificación y = X1 b1 + X2 b2 + u1 = XB + u1. En este caso, el investigador debe estimar el vector b. Usando la técnica de los mínimos cuadráticos ordinarios, el vector de estimadores es igual a:

$$\hat{\beta} = (X'X)^{-1} X' y = (X'X)^{-1} X' (X_1 b_1 + u) = (X'X)^{-1} X' X_1 b_1 + (X'X)^{-1} X'u$$

Si tenemos presente que $(X'X)^{-1} X' X = I_k$, si desarrollamos esta igualdad observamos que

$$(X'X)^{-1} X' X = (X'X)^{-1} X' [X_1, X_2] = [(X'X)^{-1} X'X_1, (X'X)^{-1} X'X_2] =$$

$$= \begin{bmatrix} I_s & 0_{s\,x(k-s)} \\ 0_{(k-s)\,x\,s} & I_{k-s} \end{bmatrix} = I_k$$

En consecuencia, el vector de estimadores queda de esta manera:

$$\hat{\beta} = (X'X)^{-1} X' X_1 b_1 + (X'X)^{-1} X'u = \begin{bmatrix} I_s \\ 0_{(k-s)\,x\,s} \end{bmatrix} b_1 + (X'X)^{-1} X'u = \begin{bmatrix} b_1 \\ 0_{(k-s)\,x\,s} \end{bmatrix} + (X'X)^{-1} X'u$$

Por tanto, se puede demostrar su insesgadez:

$$E(\hat{\beta}) = E\left\{ \begin{bmatrix} b_1 \\ 0_{(k-s)xs} \end{bmatrix} + (X'X)^{-1} X'u \right\} = \begin{bmatrix} b_1 \\ 0_{(k-s)xs} \end{bmatrix} + (X'X)^{-1} X'E(u) = \begin{bmatrix} b_1 \\ 0_{(k-s)xs} \end{bmatrix}$$

Como vemos, mientras que la estimación de los s primeros elementos tiene por media el vector b1, por lo que son insesgados, la estimación de los (k-s) restantes se distribuyen alrededor del valor 0, valor que es el correcto por cuanto no aparecen en el proceso generador de los datos. Por tanto, el vector $\hat{\beta}$ es insesgado.

Por otro lado, la matriz de varianzas y covarianzas del vector $\hat{\beta}$ se puede calcular de la siguiente manera:

Var($\hat{\beta}$)= E[]

Medidas de bondad

En esta sección vamos a definir una serie de estadísticos que nos permitan determinar la bondad del ajuste. De alguna forma, el objetivo de la estimación es intentar explicar la variabilidad de la variable. Por tanto, parece adecuado comenzar por aproximarse a la varianza muestral de la variable y. A partir de esta información, definimos como suma total, y lo representaremos como ST, al siguiente estadístico:

$$ST = \sum_{i=1}^{T} (y_i - \overline{y})^2 = y'y - T\,\overline{y}^2$$

donde $\overline{y}$ es la media muestral de la variable y.

$$y'y = (X\hat{\beta} + \hat{u})'\,(X\hat{\beta} + \hat{u}) = \hat{\beta}'X'X\hat{\beta} + \hat{\beta}'X'\hat{u} + \hat{u}'X\hat{\beta} + \hat{u}'\hat{u} = \hat{\beta}'X'X\hat{\beta} + \hat{u}'\hat{u}$$

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

En la demostración anterior hemos hecho uso de la ortogonalidad entre la matriz de regresores y el vector de residuos MCO. Si restamos a ambos lados de la igualdad T $\bar{y}^2$, queda una relación la suma total del modelo, la suma explicada y la suma residual.

$$ST = y'y - T\,\bar{y}^2 = \hat{\beta}'X'y - T\,\bar{y}^2 + \hat{u}'\hat{u} = SE + SR$$

$$SE = \sum_{i=1}^{T}\left(\hat{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{T}\hat{y}_i^2 - 2\,\bar{y}\sum_{i=1}^{T}\hat{y}_i + T\,\bar{y}^2$$

Si el modelo tiene término independiente, entonces el sumatorio de los residuos es igual a 0. Esto implica que, en este caso, $\sum_{i=1}^{T} y_i = \sum_{i=1}^{T}\hat{y}_i$. Introduciendo esta información en [], la suma explicada queda como:

$$SE = \sum_{i=1}^{T}\hat{y}_i^2 - 2\,\bar{y}\sum_{i=1}^{T} y_i + T\,\bar{y}^2 = \sum_{i=1}^{T}\hat{y}_i^2 - 2\,T\,\bar{y}^2 + T\,\bar{y}^2 = \sum_{i=1}^{T}\hat{y}_i^2 - T\,\bar{y}^2$$

Como vemos, hemos descompuesto la suma total en la suma de la parte explicada por el modelo y la suma que no el modelo no es capaz de explicar. Por tanto, a medida que la SE sea mayor en comparación con SR, mejor será la bondad del ajuste. Desgraciadamente, esta medida no estaría acotada, lo que suele ser útil en muchas ocasiones. Por ello, si dividimos por ST a ambos lados de la igualdad de [], resulta que:

$$\frac{ST}{ST} = \frac{SE + SR}{ST} \Rightarrow 1 = \frac{SE}{ST} + \frac{SR}{ST}$$

$$R^2 = 1 - \frac{SR}{ST}$$

$$\bar{R}^2 = 1 - \frac{SR}{ST}\,\frac{T-1}{T-k} = 1 - \left(1 - R^2\right)\frac{T-1}{T-k}$$

Descomposición normal y matricial

Una aproximación distinta al método de selección son aquellos estadísticos basados en el denominado criterio de información. Este criterio busca aunar ña precisión de la estimación con la regla de parsimonia en la especificación del modelo empírico. Un primer criterio perteneciente a esta familia es el propuesto en Akaike

$$AIC = \ln\left(\hat{\sigma}^2\right) + \frac{2\ k_1}{T}$$

Otra posibilidad es la sugerida en Schwarz (1978).

$$SBIC = \ln\left(\hat{\sigma}^2\right) + \frac{k_1\ \ln(T)}{T}$$

La ventaja de este último criterio frente al de Akaike es que es consistente. Así, en muestras de gran tamaño, el criterio de Akaike tendería a tomar aquellos modelos con un mayor número de regresores ya que la penalización por incluir una nueva variable explicativa sería 0. Esto no sucede para el criterio SBIC que, en términos generales, nos conduce a modelos de dimensión más pequeña que la proporcionada por el AIC.

Predicción

Uno de los objetivos de la estimación de un modelo es predecir los valores futuros de la variable que estamos analizando. No es el único fin, ya que también podemos estar interesados en un simple análisis de las relaciones estructurales que mantiene la variable endógena con el resto de las variables explicativas.

Comenzamos por el predictor óptimo. Dado que conocemos las buenas propiedades del vector de estimadores MCO, parece lógico pensar en un predictor basado en la proyección extramuestral de las estimaciones de la variable endógena. Así, si la especificación planteada es y = Xb + u, un posible predictor de los valores futuros de y vendría dado por:

$$y_{T+1} = x'_{T+1}\,\hat{\beta}$$

donde x'T+1 recoge las observaciones del periodo T+1 de las diferentes variables explicativas. Si consideramos conocidos los valores de este vector, es sencillo probar la insesgadez de este predictor:

$$E(\hat{y}_{T+1}) = E(x'_{T+1}\,\hat{\beta}) = x'_{T+1}\,\beta$$

Al mismo tiempo, tiene la siguiente varianza:

$$Var(\hat{y}_{T+1}) = E\{[\hat{y}_{T+1}-E(\hat{y}_{T+1})]^2\}= E\{[x'_{T+1}(\hat{\beta}-\beta) + u_{T+1}][x'_{T+1}(\hat{\beta}-\beta) + u_{T+1}]'\}= E\{[x'_{T+1}(X'X)^{-1}X'u + u_{T+1}][x'_{T+1}(X'X)^{-1}X'u + u_{T+1}]'\} = E[x'_{T+1}(X'X)^{-1}X'u\,u'X(X'X)^{-1}x_{T+1} + x'_{T+1}(X'X)^{-1}X'u\,u_{T+1} + u_{T+1}u'X(X'X)^{-1}x_{T+1} + u^2_{T+1}] = \sigma^2 x'_{T+1}(X'X)^{-1}x_{T+1} + \sigma^2 = \sigma^2[1 + x'_{T+1}(X'X)^{-1}x_{T+1}]$$

Se puede demostrar que esta varianza es la menor entre todos los predictores lineales e insesgados. Por tanto, el predictor es ELIO.

Por otro lado, dado que continuamos asumiendo la hipótesis de normalidad de las perturbaciones, el predictor sigue una distribución $N(x'_{T+1}\beta,\ \sigma^2[1 + x'_{T+1}(X'X)^{-1}x_{T+1}])$. Por tanto, es inmediato que podamos construir el estadístico:

$$\frac{x'_{T+1}\hat{\beta}-x'_{T+1}\beta}{\sigma^2\left[1+x'_{T+1}(X'X)^{-1}x_{T+1}\right]}\sim N(0,1)$$

EL problema que presenta el anterior estadístico es la presencia del parámetro σ^2 en el denominador. Para solucionarlo, podemos construir el siguiente estadístico:

$$t_{PE}=\frac{x'_{T+1}\hat{\beta}-x'_{T+1}\beta}{\hat{\sigma}\sqrt{\left[1+x'_{T+1}(X'X)^{-1}x_{T+1}\right]}}=\frac{e_T(1)}{D.\hat{T}.(\hat{y}_{T+1})}\sim t_{T-k}$$

La hipótesis nula es que no existe cambio estructural con posterioridad al periodo T. Caso de ser eso cierto el error de predicción de primer orden estará próxima a cero, lo que hará que el numerador del estadístico tienda hacia 0. Por el contrario, cuanto mayor sea la ruptura, mayor será la diferencia entre el vector de estimadores y el de parámetros, lo que implicará que el numerador tome valores elevados y más fácil será el rechazo de la hipótesis nula.

A partir de [] también podemos construir un intervalo de confianza para la predicción, para un nivel de confianza 100(1-α) dado:

$$y_{T+1}\in\left(\hat{y}_{T+1}-t^{\alpha/2}_{T-k}D.\hat{T}.(\hat{y}_{T+1}),\ \hat{y}_{T+1}-t^{\alpha/2}_{T-k}D.\hat{T}.(\hat{y}_{T+1})\right)$$

También es posible predecir para un periodo temporal más amplio. En este caso, el proceso generador de los datos será: $y_h = X_h\ \beta + u_h$, donde y_h es un vector fila de tamaño h y X_h es una matriz de orden hxk. El predictor del vector de valores futuros vendrá dado por $\hat{y}_h = X_h\ \hat{\beta}$. Sus propiedades son las mismas que las observadas para el caso de que el horizonte de predicción sea de un periodo. En primer lugar, es un vector de predictores insesgados y, en segundo, tiene matriz de varianzas y covarianzas:

Salkever.

Intervalos de confianza.