

Capítulo 1

El modelo lineal general.

Especificación para datos de sección cruzada y para datos en forma de serie temporal. Estimadores mínimo cuadrático ordinarios (MCO). Propiedades para muestras finitas y para grandes muestras. Estimador del método generalizado de los momentos y del método de máximo verosimilitud.

1.1. Introducción.

El objeto de la econometría consiste en:

1. Especificar un modelo de relación entre variables económicas.
2. Utilizar información muestral acerca de los valores de las variables al objeto de cuantificar la magnitud de la dependencia entre ellas.
3. Evaluar la validez de las hipótesis formuladas por la teoría económica acerca de la relación entre las variables objeto de estudio y, en algunos casos,
4. Efectuar un ejercicio de seguimiento coyuntural y de predicción de las variables estudiadas.

El objeto del estudio será un modelo de relación entre variables económicas, que denotaremos por:

$y = f(x_1, x_2, \dots, x_k, u/\beta)$, que trata de explicar el comportamiento de la variable y utilizando la información suministrada por el conjunto de k variables x_i , variables explicativas con significado económico, así como de una variable aleatoria no observable y sin significado económico, que denotamos por u y llamamos término de error. La relación de dependencia se define a través de un vector de parámetros, β , que es el que queremos averiguar.

La información muestral consiste en una lista ordenada de valores de las variables y, x_1, x_2, \dots, x_k . La muestra es de sección cruzada si los conjuntos de valores corresponden a información proporcionada por diversos agentes económicos en el mismo instante de tiempo, y de series temporales si los datos corresponden a una misma unidad económica en diversos instantes de tiempo. Por tanto, disponemos de una lista de relaciones $y_i = f(x_{1i}, x_{2i}, \dots, x_{ki}, u_i/\beta)$, $i = 1, 2, \dots, N$.

En general, las relaciones que trataremos serán siempre lineales de la forma:

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + u_i, \quad i = 1, 2, \dots, N$$

Que denominamos modelo de regresión lineal simple, o **modelo lineal general**.

La variable y se denomina variable endógena o explicada, las variables x se denominan variables exógenas o explicativas, al término u se le denomina término de error. A los β_k se les denomina coeficientes del modelo y reflejan la influencia de cada variable explicativa en la variable endógena. Si hacemos $x_{1i} = 1$ para todas las i tenemos un modelo con un término independiente.

1.2. Especificación para datos de sección cruzada y para datos en forma de serie temporal.

El modelo lineal general se puede expresar de forma matricial como sigue:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} & x_{2N} & \cdots & x_{kN} \end{pmatrix} \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}$$

$$y = X\beta + u$$

Para que el análisis econométrico que vamos a realizar sea consistente, el modelo debe cumplir una serie de hipótesis. Hay un conjunto que son comunes, y otro conjunto que son específicas para modelos de series temporales/datos transversales.

1.2.1. El modelo es lineal, estocástico y constante.

Es decir, el proceso generador de los datos es del tipo:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i$$

Esto implica que el proceso estocástico del que provienen los datos es de naturaleza lineal, no se trata de una aproximación o proyección.

Además la relación entre las variables explicativas y la variable endógena no es determinista, existe un término de error distinto de cero. Esto se justifica por las siguientes razones:

- El modelo es sólo una aproximación a la verdadera relación entre las variables.
- Las variables económicas del modelo están sujetas a errores de medida.
- Se reconoce la posible existencia de otros factores determinantes del comportamiento de y que no se han incluido en el modelo.

También suponemos que los coeficientes son los mismos para toda la muestra de que disponemos y para los valores que queremos estimar. Si no fuese así, el problema de estimación sería más complejo.

1.2.2. El modelo no tiene multicolinealidad.

Esto quiere decir que no hay una relación lineal entre las variables X , es decir, que la matriz $E(x_i x_i')$ > 0. Este supuesto tiene dos razones: una eminentemente práctica, ya que como veremos en caso de multicolinealidad perfecta no se pueden obtener los estimadores de los parámetros, y otra de índole teórica, ya que en este caso no podríamos estimar el efecto de un cambio en un regresor manteniendo el resto constantes. Es importante fijarse en que la hipótesis descarta la multicolinealidad perfecta, pero sí que permite la correlación entre las variables explicativas.

1.2.3. Exogeneidad del modelo.

Para datos de sección cruzada esta hipótesis exige que $E(\varepsilon_i | \mathbf{x}_i) = 0$. Para datos de series temporales, que $E(\varepsilon_t | \mathbf{x}_t) = 0$. Es decir, el valor esperado de la perturbación condicionado a los valores de las variables explicativas es nulo. Por tanto, en media el error no depende de los valores que tomen las variables explicativas. Este supuesto implica que \mathbf{x}_i y ε_i están incorrelacionadas, pero no se produce la implicación al revés, ya que la correlación solo mide la relación lineal. Sin embargo, si ambas variables son independientes sí que se cumple el supuesto.

Si se cumple este supuesto diremos que tenemos variables explicativas exógenas. Si alguna de las variables explicativas está correlacionada con la perturbación, entonces diremos que esa variable es endógena.

En caso de que se cumpla que $E(\varepsilon_i | \mathbf{X}) = 0$, es decir, que la esperanza condicionada de la perturbación es cero para todas las observaciones, diremos que tenemos **exogeneidad estricta**.

Esta hipótesis implica que $E(\varepsilon_i) = 0$, ya que por la ley de las esperanzas totales, $E(\varepsilon_i) = E[E(\varepsilon_i | \mathbf{x}_i)]$ y por tanto, $E[E(\varepsilon_i | \mathbf{x}_i)] = E[0] = 0$. Esto podría parecer muy restrictivo, sin embargo, si el modelo tiene término independiente y $E(\varepsilon_i) = \mu \neq 0$, podemos expresar el término independiente como $\beta_1 + \mu$, y el error como $\varepsilon - \mu$, con lo que se cumpliría esta hipótesis.

1.2.4. Muestra aleatoria.

Las variables aleatorias multidimensionales $(X_{1i}, X_{2i}, \dots, X_{ki}, Y_i)$ son independientes e idénticamente distribuidas. Es decir, los conjuntos de datos provienen de individuos seleccionados de forma aleatoria de una población.

Esta hipótesis normalmente no se cumple en el caso de series temporales. En efecto, para muchas variables económicas el valor de la variable en el futuro está influido por el valor de la misma en el presente, es decir, hay correlación entre observaciones próximas. Es por eso que para datos de series temporales se adopta la siguiente hipótesis.

Esta hipótesis implica que no haya correlación entre las perturbaciones (autocorrelación).

1.2.5. Correlación decreciente.

Las variables aleatorias multidimensionales $(X_{1t}, X_{2t}, \dots, X_{kt}, Y_t)$ tienen la misma distribución a lo largo del tiempo, y la dependencia entre $(X_{1t}, X_{2t}, \dots, X_{kt}, Y_t)$ y $(X_{1t-j}, X_{2t-j}, \dots, X_{kt-j}, Y_{t-j})$ disminuye rápidamente al aumentar j .

Si exigimos simultáneamente las hipótesis de exogeneidad y muestra aleatoria, automáticamente se cumple el supuesto de exogeneidad estricta. Sin embargo, en el caso de la hipótesis de correlación decreciente esto no es cierto.

1.2.6. Existe una relación causal entre las variables explicativas y la variable endógena:

Es decir, existe una justificación teórica del modelo.

1.2.7. Las variables explicativas son deterministas:

Es decir, si volviésemos a obtener la misma muestra, los valores de las x se mantendrían constantes.

1.3. Estimadores mínimo cuadrático ordinarios.

El primer objetivo del análisis econométrico es obtener estimadores de los parámetros del modelo. Es decir, si expresamos el modelo en forma matricial, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, queremos obtener un estimador del vector $\boldsymbol{\beta}$. Lo denotaremos por $\hat{\boldsymbol{\beta}}$.

Una vez obtenido $\hat{\boldsymbol{\beta}}$, se puede calcular para cada elemento de la muestra: $\hat{y}_i = \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \dots + \hat{\beta}_k x_{ki}$, estimadores de las variables y_i . Definimos el **residuo** como la diferencia entre el valor real de la variable endógena y su estimación, $\hat{u}_i = y_i - \hat{y}_i$. La serie de residuos representados en forma matricial será:

$$\hat{\mathbf{u}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Llamamos **estimador de mínimos cuadrados** a aquel estimador de $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{MCO}$, que minimiza la suma de los cuadrados de los residuos. Así:

$$\begin{aligned}\hat{\mathbf{u}}'\hat{\mathbf{u}} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{y}'\mathbf{y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ \frac{\partial \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\boldsymbol{\beta}}} &= -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \\ \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y} \\ \frac{\partial^2 \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}'} &= \mathbf{X}'\mathbf{X}\end{aligned}$$

Como $\mathbf{X}'\mathbf{X}$ es siempre semidefinida positiva, tenemos un mínimo. La ecuación $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$ es un sistema de k ecuaciones lineales (sistema de ecuaciones normales) con una incógnita por cada uno de los k parámetros del vector $\hat{\boldsymbol{\beta}}$. Éste sistema tiene generalmente una única solución, que será nuestro estimador de mínimos cuadrados ordinarios:

$$\hat{\boldsymbol{\beta}}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

Si la matriz $\mathbf{X}'\mathbf{X}$ es singular, no se podrá invertir, y el sistema de ecuaciones tiene infinitas soluciones. Esto se produce cuando se vulnera el supuesto de que las variables explicativas no sean linealmente dependientes. A este fenómeno se le llama multicolinealidad.

El sistema tendrá solución única siempre que:

- Las variables explicativas no sean linealmente dependientes.

- El número de observaciones sea igual o mayor que el número de parámetros a estimar.

Para lograr precisión en la estimación MCO es necesario que el número de observaciones sea mucho mayor que el número de parámetros a estimar. Al valor $N - k$ se le conoce como número de grados de libertad de la estimación.

1.4. Propiedades para muestras finitas y para grandes muestras.

El estimador MCO es un vector aleatorio, ya que depende de las y_i que son variables aleatorias (ya que dependen del término de error, u). A partir de las hipótesis básicas del modelo podemos definir una serie de propiedades del estimador, que caracterizan su distribución de probabilidad.

1.4.1. Insesgadez.

El estimador es insesgado siempre que se cumpla el supuesto de esperanza condicionada nula. Por la ley de las esperanzas totales esto implica que $E(\mathbf{u}) = \mathbf{0}_N$. Entonces:

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} \\ E(\hat{\beta}|\mathbf{X}) &= E\left[\beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}|\mathbf{X}\right] = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{u}|\mathbf{X}) = \beta\end{aligned}$$

Ya que β es un vector constante, aunque desconocido.

Como por la ley de las esperanzas totales, $E[E(\hat{\beta}|\mathbf{X})] = E(\hat{\beta})$, $E(\hat{\beta}) = \beta$ y el estimador es insesgado.

Como consecuencia, podemos expresar el error de estimación como:

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

Para que se de la propiedad de insesgadez, basta que se cumpla el supuesto de exogeneidad. Sin embargo, si se cumple que $E(\varepsilon) = 0$ también se da la insesgadez.

1.4.2. Varianza del estimador MCO.

Si además de las suposiciones de la especificación del modelo añadimos la hipótesis de que el término de error es homocedástico, es decir, $Var(\varepsilon = \sigma^2 \mathbf{I}_n)$. Esto quiere decir que la varianza del término de error es constante, y que las covarianzas cruzadas son nulas. Si este supuesto falla, diremos que el modelo presenta heteroscedasticidad.

$$\begin{aligned}Var(\hat{\beta}) &= E\left[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'\right] = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = \\ &= E\left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\sigma_u^2 \mathbf{I}_N \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

La diagonal principal de esta matriz nos da la varianza de cada uno de los estimadores de los parámetros del modelo. Veamos qué factores influyen en estas varianzas:

- Cuanto mayor sea la varianza de la perturbación, mayor será la varianza de los estimadores. Esto es lo esperado, cuanto más se aparte nuestro sistema del modelo que hemos especificado, menos precisos serán los estimadores.
- Cuanto mayor sea la dispersión de las variables explicativas, o mayor sea el tamaño de la muestra, menor será la varianza, ya que la matriz está dividiendo. Esto es lógico pues por n lado, cuanto más repartida esté la muestra por el rango de variación posible más información capturaremos, y a mayor tamaño de la muestra más eficiente será nuestro estimador.
- Cuanta menos multicolinealidad presenten las variables explicativas, menor será la varianza. Si las variables explicativas presentan un comportamiento muy cercano a la multicolinealidad, la matriz será muy próxima a ser singular, con un determinante próximo a cero. Por tanto, su inversa será muy grande y por tanto las varianzas también, dando origen a estimadores muy poco eficientes.

1.4.3. Teorema de Gauss-Markov:

Teorema 1 Teorema de Gauss-Markov: *Bajo las hipótesis del modelo lineal, y en presencia de heteroscedasticidad, el estimador MCO es el más eficiente dentro de la clase de estimadores lineales insesgados.*

Sea $\tilde{\beta} = \tilde{A}y$ estimador lineal insesgado de β . Sea $A = \tilde{A} - (X'X)^{-1}X'$. Por tanto,

$$\tilde{\beta} = [A + (X'X)^{-1}X']y = [A + (X'X)^{-1}X'](X\beta + u) = AX\beta + \beta + [A + (X'X)^{-1}X']u$$

$$E(\tilde{\beta}) = AX\beta + \beta$$

Y por tanto, $AX = 0_{k \times k}$ ya que el estimador es insesgado. Por tanto, $\tilde{\beta} = \beta + [A + (X'X)^{-1}X']u$

$$Cov(\tilde{\beta}) = E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = E\left[\left([A + (X'X)^{-1}X']u\right)\left([A + (X'X)^{-1}X']u\right)'\right] = \sigma_u^2 AA' + \sigma_u^2 (X'X)^{-1}$$

Y como AA' es una matriz semidefinida positiva, la matriz de varianzas y covarianzas de $\tilde{\beta}$ será mayor que la de $\hat{\beta}$.

1.4.4. Cada una de las variables explicativas es ortogonal al vector de residuos:

O lo que es lo mismo, $X'\hat{u} = 0_N$

$$X'\hat{u} = X'(y - X\hat{\beta}) = X'y - X'X(X'X)^{-1}X'y = 0_N$$

Como consecuencia, si el modelo tiene término independiente, $\sum_{i=1}^N \hat{u}_i = 0$, la suma de los residuos es cero.

1.4.5. Expresiones de la suma residual.

$$SR = \hat{u}'\hat{u} = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta} = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X(X'X)^{-1}X'y =$$

$$y'y - \hat{\beta}'X'y$$

$$SR = \hat{u}'\hat{u} = y'y - \hat{y}'\hat{y}, \text{ ya que } \hat{y}'\hat{y} = \hat{\beta}'X'X\hat{\beta} = \hat{\beta}'X'X(X'X)^{-1}X'y = \hat{\beta}'X'y$$

1.4.6. El vector de residuos es una transformación lineal del término de error.

$$\hat{u} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = [I_N - X(X'X)^{-1}X']y = My = Mu, \text{ con}$$

$$M = I_N - X(X'X)^{-1}X' \text{ y dado que } MX = 0_{N \times k}.$$

Por tanto, $\hat{u}'\hat{u} = u'M'Mu = u'Mu$ ya que M es singular, simétrica e idempotente.

Además, se deduce que $E(\hat{u}) = 0_N$, $Var(\hat{u}) = \sigma_u^2 M$.

1.4.7. Sumas de cuadrados.

Definimos las siguientes expresiones:

Suma Total: $ST = \sum_{i=1}^N (y_i - \bar{y})^2$. Es la varianza muestral de la variable endógena multiplicada por el número de observaciones en la muestra. Es una medida de las fluctuaciones de la variable.

Suma Explicada: $SE = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$. Es la fluctuación de los estimadores de la variable endógena generados por el modelo alrededor de la media muestral, es decir, es el grado de fluctuación que explica el modelo.

Suma Residual: $SR = \sum_{i=1}^N (y_i - \hat{y}_i)^2$. Es la fluctuación no explicada por el modelo, es decir, indica el nivel de error del modelo al explicar la relación entre las variables explicativas y la variable endógena.

Si entre las variable explicativas hay un término constante, entonces $ST = SE + SR$.

$$ST = \sum_{i=1}^N (y_i - \bar{y})^2 = ST = \sum_{i=1}^N y_i^2 - 2\bar{y} \sum_{i=1}^N y_i + \sum_{i=1}^N \bar{y}^2 = \sum_{i=1}^N y_i^2 - N\bar{y} = y'y - N\bar{y}$$

Hemos visto que $SR = \hat{u}'\hat{u} = y'y - \hat{y}'\hat{y}$, por tanto, $y'y - N\bar{y} = \hat{y}'\hat{y} - N\bar{y} + \hat{u}'\hat{u}$.

$$SE = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 = \hat{\mathbf{y}}' \hat{\mathbf{y}} - 2\bar{y} \sum_{i=1}^N \hat{y}_i + \sum_{i=1}^N \bar{y}^2 = \hat{\mathbf{y}}' \hat{\mathbf{y}} - 2\bar{y} \sum_{i=1}^N y_i - 2\bar{y} \sum_{i=1}^N \hat{u}_i + \sum_{i=1}^N \bar{y}^2 = \hat{\mathbf{y}}' \hat{\mathbf{y}} - N\bar{y} - 2\bar{y} \sum_{i=1}^N \hat{u}_i$$

Y como si el modelo tiene término independiente, $\sum_{i=1}^N \hat{u}_i = 0$, $ST = SE + SR$.

1.4.8. Estimación de σ_u^2 .

Para estimar la matriz de covarianzas de $\hat{\beta}$, debemos estimar σ_u^2 .

Sabemos que $\hat{\mathbf{u}} = [\mathbf{I}_N - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{u} = \mathbf{M}\mathbf{u}$. Por tanto, $E(\hat{\mathbf{u}}'\hat{\mathbf{u}}) = E(\mathbf{u}'\mathbf{M}\mathbf{u}) = E[\text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})]$, donde tr es el operador traza y utilizamos que $\mathbf{u}'\mathbf{M}\mathbf{u}$ es un escalar y un escalar es igual a su traza. Por las propiedades del operador traza,

$E[\text{tr}(\mathbf{u}'\mathbf{M}\mathbf{u})] = E[\text{tr}(\mathbf{M}\mathbf{u}'\mathbf{u})] = \text{tr}[E(\mathbf{M}\mathbf{u}'\mathbf{u})] = \text{tr}[\mathbf{M}\sigma_u^2\mathbf{I}_N] = \sigma_u^2 \text{tr}(\mathbf{M}) = \sigma_u^2(N - k)$. Y, por tanto, un estimador insesgado de σ_u^2 será: $\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N - k}$.

1.4.9. Bondad del ajuste.

Definimos el coeficiente de determinación como: $R^2 = 1 - \frac{SR}{ST}$. Mide la proporción de variación de la variable endógena explicada por el modelo. A su raíz cuadrada positiva, cuando existe, se le denomina coeficiente de correlación lineal, R .

Como consecuencia de las propiedades de la suma de residuos, si el modelo tiene término independiente,

$$R^2 = \frac{SE}{ST}.$$

El valor del coeficiente de determinación depende del tamaño de la muestra y del número de regresores.

Esto hace que no sea útil para comparar distintos modelos. Para ello se define el coeficiente de determinación corregido, que elimina estos efectos: $\bar{R}^2 = 1 - \frac{SR/(N-k)}{ST/(N-1)} = 1 - \frac{N-1}{N-k} (1 - R^2)$.

Otras medidas de la bondad del ajuste son el criterio de Schwarz, $SC = \ln \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N} + \frac{k}{N} \ln N$ y el criterio de información de Akaike, $CI_{AK} = \ln \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N} + \frac{2k}{N}$.

1.4.10. Distribución de los estimadores.

Si además de las hipótesis de especificación del modelo y la hipótesis de homocedasticidad, añadimos la hipótesis de normalidad del término de error, es decir, $\epsilon \sim N(\mathbf{0}, \mathbf{I}_n\sigma^2)$ completamos el modelo clásico de regresión lineal. Bajo estos supuestos, podemos afirmar:

Teorema 2 *Bajo los supuestos del modelo clásico de regresión lineal se cumple que:*

$$\begin{aligned}\hat{\beta} &\sim N(\beta, \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}) \\ \mathbf{X}\hat{\beta} &\sim N(\mathbf{X}\beta, \sigma^2 \mathbf{P}) \\ \hat{\mathbf{u}} &\sim N(\mathbf{0}, \sigma^2 \mathbf{M})\end{aligned}$$

Sin embargo, la hipótesis de normalidad del término de error es bastante fuerte, ya que para unas X fijas implica la normalidad de la variable dependiente, algo que no siempre podemos afirmar.

No conocer la distribución muestral de los estimadores MCO hace que estos pierdan gran parte de su utilidad, ya que nos impide hacer inferencia sobre los mismos, y por tanto no podremos contrastar hipótesis sobre el modelo. Es por esto que vamos a ver el siguiente teorema:

Teorema 3 *Bajo las hipótesis elementales del modelo, de linealidad, no multicolinealidad, exogeneidad y aleatoriedad, o su equivalente para ST, y suponiendo que grandes atípicos sean poco probables, la distribución del estimador MCO es asintóticamente normal a medida que crece el tamaño de la muestra, con media β y varianza $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$.*

Este teorema se basa en el Teorema Central del Límite, y por ello la distribución es aproximada, no exacta, y cuanto mayor sea el tamaño de la muestra mejor será la aproximación. Podemos considerar que si $n > 100$ la aproximación será lo suficientemente confiable, salvo que haya indicios que nos indiquen lo contrario.

Ya hemos visto que la varianza disminuye al aumentar n , con lo que además vemos que el estimador es consistente.

1.5. Estimador del método generalizado de los momentos y del método de máxima verosimilitud.

1.5.1. Estimador del método generalizado de los momentos.

El método generalizado de los momentos produce estimadores con propiedades muy deseables, sin embargo estas propiedades solo se cumplen para los casos de muestras muy grandes. Típicamente, son estimadores asintóticamente eficientes en muestras muy grandes, pero pierden esa eficiencia para muestras de menor tamaño.

El método se basa en el método de los momentos, que consiste en igualar los momentos respecto al origen poblacionales a los momentos muestrales, y resolver el sistema de ecuaciones resultante. Los estimadores obtenidos a partir de este método son consistentes.

Para aplicar este método a nuestro problema, consideremos nuestro modelo: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Hemos visto que, si está bien especificado, debe cumplirse que $E[\mathbf{X}'\mathbf{u}] = 0$. Teniendo en cuenta que $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, podemos escribir que $E[\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = 0$.

Aplicamos el principio del método generalizado de los momentos, y sustituimos el momento poblacional por el momento muestral. Como sabemos que $\boldsymbol{\beta}$ hace que el momento poblacional sea cero, asumiremos que una buena estimación hará que el momento muestral valga cero, y por tanto:

$$\frac{1}{n}\mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0$$

Resolviendo esta ecuación obtenemos la estimación por el método generalizado de los momentos, que será:

$$\hat{\boldsymbol{\beta}}_{MCO} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Y como vemos, coincide con el estimador MCO.

1.5.2. Estimador de máxima verosimilitud.

En lugar de aplicar el criterio de mínimos cuadrados, utilizamos el método de máxima verosimilitud para estimar los valores de $\boldsymbol{\beta}$ y σ_u^2 .

Si suponemos que el vector de términos de error sigue una distribución normal, $\mathbf{u} \sim N(\mathbf{0}_N, \sigma_u^2 \mathbf{I}_N)$, la función de densidad es:

$$f(\mathbf{u}) = \frac{1}{(2\pi)^{N/2}} \frac{1}{(\sigma_u^2)^{N/2}} e^{-\frac{1}{2\sigma_u^2}\mathbf{u}'\mathbf{u}}$$

Como $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, hacemos el cambio de variable. El jacobiano de la transformación es la matriz identidad, por tanto,

$$\begin{aligned} L(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma_u^2) &= \frac{1}{(2\pi)^{N/2}} \frac{1}{(\sigma_u^2)^{N/2}} e^{-\frac{1}{2\sigma_u^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \\ \ln L(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma_u^2) &= -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma_u^2 - \frac{1}{2\sigma_u^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ \frac{\partial \ln L(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma_u^2)}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma_u^2} [2\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = \mathbf{0}_k \\ \frac{\partial \ln L(\mathbf{y}, \mathbf{X}; \boldsymbol{\beta}, \sigma_u^2)}{\partial \sigma_u^2} &= -\frac{N}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0 \end{aligned}$$

Y por tanto,

$$\begin{aligned} \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} &= \mathbf{X}'\mathbf{y} \\ \sigma_u^2 &= \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{N} = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{N} \end{aligned}$$

Es decir, el estimador máximo verosímil de $\boldsymbol{\beta}$ coincide con el estimador de mínimos cuadrados. Por lo tanto, $E(\hat{\boldsymbol{\beta}}_{MV}) = \boldsymbol{\beta}$, $Var(\hat{\boldsymbol{\beta}}_{MV}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$. Sin embargo, el estimador de σ_u^2 es distinto, y además es sesgado:

$E(\hat{\sigma}_{MV}^2) = \frac{N-k}{N}\sigma_u^2$, aunque al aumentar el tamaño muestral el sesgo se hace cada vez más pequeño.

Capítulo 2

Inferencia en el modelo lineal.

Distribución muestral de los estimadores MCO. Contraste de hipótesis. Contraste acerca de un coeficiente del modelo. Contraste de un subconjunto paramétrico. Inferencia para grandes muestras: consistencia, eficiencia y normalidad asintótica. Contrastes para grandes muestras basados en el Multiplicador de Lagrange. Predicción en el modelo lineal.

2.1. Inferencia en el modelo lineal.

Una vez hemos estimado el modelo puede ser interesante o necesario el contraste de distintas hipótesis sobre los coeficientes del modelo, para comprobar si cumplen las hipótesis teóricas subyacentes al mismo. Esto contrastes pueden ser de significación (contrastar si una o varias variables explicativas influyen en el valor de la variable dependiente), de valor del mismo, o puede ser interesante la definición de intervalos o regiones de confianza para el valor de los parámetros.

2.2. Contraste de hipótesis.

A partir de ahora, mantendremos el supuesto de que el término de error del modelo sigue para cada observación una distribución normal, de media cero y varianza constante para todas las observaciones σ_u^2 , es decir el vector \mathbf{u} se distribuye según una normal multivariante $N(\mathbf{0}_N; \sigma_u^2 \mathbf{I}_N)$. Como el estimador MCO es una transformación lineal del vector \mathbf{u} , se tiene que $\hat{\beta}_{MCO} \sim N_k(\beta; \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1})$.

Además, se puede demostrar que siendo $\mathbf{x} \sim N_k(\mathbf{0}_k; \sigma^2 \mathbf{I}_k)$ y \mathbf{A} una matriz simétrica e idempotente de rango r , entonces $\frac{1}{\sigma^2} \mathbf{x}' \mathbf{A} \mathbf{x} \sim \chi^2(r)$. Por tanto, como $\hat{\mathbf{u}}' \hat{\mathbf{u}} = \mathbf{u}' \mathbf{M} \mathbf{u}$ se deduce que $\frac{1}{\sigma_u^2} \hat{\mathbf{u}}' \hat{\mathbf{u}} = (N - k) \frac{\hat{\sigma}_u^2}{\sigma_u^2} \sim \chi^2(N - k)$.

Como $\hat{\beta}_{MCO} \sim N_k(\beta; \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1})$, se deduce que $\mathbf{X}(\hat{\beta} - \beta) \sim N_k(\mathbf{0}_k; \sigma_u^2 \mathbf{I}_k)$, así que

$$\frac{1}{\sigma_u^2} (\hat{\beta} - \beta)' \mathbf{X}' \mathbf{I}_k \mathbf{X} (\hat{\beta} - \beta) = \frac{1}{\sigma_u^2} (\hat{\beta} - \beta)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta) \sim \chi^2(k)$$

Por otro lado, sabemos que $\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{u}$ y que $\hat{\mathbf{u}}' \hat{\mathbf{u}} = \mathbf{u}' \mathbf{M} \mathbf{u}$. Como una forma lineal $\mathbf{L} \mathbf{x}$ y una forma cuadrática $\mathbf{x}' \mathbf{A} \mathbf{x}$ se distribuyen de forma independiente si $\mathbf{L} \mathbf{A} = \mathbf{0}$, y

$(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{M} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' [\mathbf{I}_N - \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'] = \mathbf{0}_N$, podemos decir que $\hat{\beta} - \beta$ y $\hat{\mathbf{u}}' \hat{\mathbf{u}} = (N - k) \hat{\sigma}_u^2$ se distribuyen de forma independiente.

Combinando todos estos resultados, podemos llegar a la conclusión de que el estadístico:

$$\frac{\frac{1}{\sigma_u^2} (\hat{\beta} - \beta)' \mathbf{X}' \mathbf{X} (\hat{\beta} - \beta) / k}{\frac{1}{\sigma_u^2} \hat{\mathbf{u}}' \hat{\mathbf{u}} / (N - k)} = (\hat{\beta} - \beta)' [\hat{\sigma}_u^2 (\mathbf{X}'\mathbf{X})^{-1}]^{-1} (\hat{\beta} - \beta) / k \sim F_{k, N-k}$$

Con este estadístico podemos contrastar la hipótesis nula $H_0 : \beta = \beta^0$, substituyendo en el estadístico el valor de β por β^0 y los estimadores por los valores obtenidos en la estimación. SI el valor del estadístico es menor que el valor de $F_{k, N-k}$ para el nivel de significación elegido, no podremos rechazar la hipótesis nula. En otro caso, la rechazaremos.

2.3. Tratamiento general del contraste de hipótesis.

El apartado anterior nos permite contrastar una hipótesis sobre todos los coeficientes del modelo, pero nos puede interesar contrastar hipótesis sobre el valor de uno o varios coeficientes del modelo, o sobre la significación de uno o varios de los coeficientes (es decir, sobre si su valor es cero). Para ello vamos a desarrollar un método más general de contrastación de hipótesis.

Definimos una hipótesis general, $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, siendo \mathbf{R} una matriz $q \times k$, siendo q el número de restricciones, con los coeficientes de los parámetros $\boldsymbol{\beta}$ en cada una de las restricciones y \mathbf{r} un vector con q filas con los valores de las restricciones. De esta forma podemos definir cualquier conjunto de hipótesis lineales sobre los coeficientes del modelo. Por ejemplo, para la hipótesis $H_0 : \beta_1 = \beta_1^0$, \mathbf{R} sería una matriz $1 \times k$ con el primer término igual a 1 y el resto cero, y \mathbf{r} sería un escalar con valor β_1^0 .

Como \mathbf{R} es una matriz constante, $\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim N_k(\mathbf{0}; \sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')$. Si la hipótesis nula es cierta,

$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ y por tanto $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \sim N_q(\mathbf{0}_q; \sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}')$. Finalmente,

$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\sigma_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi^2(q)$, y por tanto,

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k}} \sim F_{q, N-k}$$

O, lo que es lo mismo, $(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' [\hat{\sigma}_u^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q \sim F_{q, N-k}$. Y por tanto, si el valor del estadístico es menor que el valor de $F_{q, N-k}$ para el nivel de significación elegido, no podremos rechazar la hipótesis nula. En otro caso, la rechazaremos.

2.4. Contraste acerca de un coeficiente del modelo.

En este caso, $H_0 : \beta_i = \beta_i^0$. Entonces, $\mathbf{R} = [0, 0, \dots, 1, 0, \dots, 0]$ ocupando el 1 la posición i -ésima, y $\mathbf{r} = \beta_i^0$, y por tanto $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \hat{\beta}_i - \beta_i^0$, escalar. El producto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' = a_{ii}$, siendo a_{ii} el elemento i -ésimo de la

diagonal de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$, y el estadístico se convierte en $\frac{(\hat{\beta}_i - \beta_i^0)^2}{\hat{\sigma}_u^2 a_{ii}} \sim F_{1, N-k}$. Como

$Var(\hat{\boldsymbol{\beta}}) = \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$, $\hat{Var}(\hat{\beta}_i) = \hat{\sigma}_u^2 a_{ii}$. Además, aplicando raíces cuadradas, $\frac{\hat{\beta}_i - \beta_i^0}{\hat{\sigma}_u \sqrt{a_{ii}}} \sim t_{N-k}$. Para

contrastar la significación de la variable explicativa x_i en el modelo, contrastamos que el valor de su coeficiente sea igual a cero, es decir, utilizamos el estadístico $\frac{\hat{\beta}_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \sim t_{N-k}$, que se suele conocer como el estadístico t del cociente estimado $\hat{\beta}_i$.

2.5. Contraste de un subconjunto paramétrico.

Ahora contrastamos la significación de un subconjunto de variables explicativas. Sin pérdida de generalidad supondremos que son las últimas del modelo, por tanto la matriz será: $\mathbf{R} = [\mathbf{0}_{s, k-s}; \mathbf{I}_s]$ y el vector $\mathbf{r} = \mathbf{0}_s$. Particionaremos la matriz $\mathbf{X} = [\mathbf{X}_1; \mathbf{X}_2]$, con \mathbf{X}_1 de dimensión $N \times k - s$ y \mathbf{X}_2 de dimensiones $N \times s$, y el vector de parámetros en $\boldsymbol{\beta} = (\boldsymbol{\beta}_1; \boldsymbol{\beta}_2)$. El modelo econométrico puede escribirse como

$\mathbf{y} = (\mathbf{X}_1; \mathbf{X}_2) \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} + \hat{\mathbf{u}} = \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\mathbf{u}}$. Por tanto, $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \hat{\boldsymbol{\beta}}_2$, y el producto $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}'$ tiene

como resultado la submatriz $s \times s$ inferior derecha de $(\mathbf{X}'\mathbf{X})^{-1}$. Esta submatriz es igual a $(\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2)^{-1}$, con $\mathbf{M}_1 = \mathbf{I}_N - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$, y por tanto,

$$F = \frac{\hat{\boldsymbol{\beta}}_2' (\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2) \hat{\boldsymbol{\beta}}_2 / s}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k}} \sim F_{s, N-k}$$

Si en lugar de contrastar la significación queremos contrastar los valores de los parámetro, el estadístico sería:

$$F = \frac{(\hat{\beta}_2 - \beta_2^0)' (\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2) (\hat{\beta}_2 - \beta_2^0) / s}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k}} \sim F_{s, N-k}$$

2.6. Contraste de significación global del modelo.

Si queremos contrastar la significación de todas las variables del modelo la matriz será: $\mathbf{R} = [\mathbf{0}_{k-1,1}; \mathbf{I}_{k-1}]$ y el vector $\mathbf{r} = \mathbf{0}_{k-1}$ (el término independiente no se contrasta). Particionaremos la matriz $\mathbf{X} = [\mathbf{1}_N; \mathbf{X}_2]$, con \mathbf{X}_1 de dimensión $N \times 1$ y \mathbf{X}_2 de dimensiones $N \times k - 1$, y el vector de parámetros en $\beta = (\beta_1; \beta_2)$.

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} N & \mathbf{1}_N' \mathbf{X}_2 \\ \mathbf{X}_2' \mathbf{1}_N & \mathbf{X}_2' \mathbf{X}_2 \end{pmatrix}$$

$$F = \frac{\hat{\beta}_2' (\mathbf{X}_2' \mathbf{Q} \mathbf{X}_2) \hat{\beta}_2 / (k - 1)}{\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k}} \sim F_{k-1, N-k}$$

Con $\mathbf{Q} = \mathbf{I}_N - \frac{1}{N} \mathbf{1}_N \mathbf{1}_N'$. Este estadístico admite una expresión alternativa:

$$F = \frac{\frac{SCE}{k-1}}{\frac{SCR}{N-k}} = \frac{R^2 / (k-1)}{(1 - R^2) / (N - k)}, \text{ sólo si el modelo contiene un término independiente. Aunque ninguna de}$$

las variables sea significativa, el término independiente sería aproximadamente igual a la media de la variable endógena, y por tanto debería ser significativo.

2.7. Intervalos y regiones de confianza.

Obtener una estimación puntual a partir de una muestra dada no es muy útil a no ser que se pueda proporcionar un intervalo de confianza de dicha estimación. Para obtener estos intervalos utilizaremos los resultados anteriores.

2.7.1. Intervalo de confianza para un solo coeficiente.

Hemos visto que el estadístico $\frac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_u \sqrt{a_{ii}}} \sim t_{N-k}$. Si definimos $t_{N-k; \alpha/2}$ como el valor para el que se cumple

$P(t \geq t_{N-k; \alpha/2}) = \alpha/2$ en una distribución t de Student, un intervalo de confianza del $(1 - \alpha) \%$ será $[\hat{\beta}_i - t_{N-k; \alpha/2} \hat{\sigma}_u \sqrt{a_{ii}}, \hat{\beta}_i + t_{N-k; \alpha/2} \hat{\sigma}_u \sqrt{a_{ii}}]$. El tamaño del intervalo depende directamente de la varianza del estimador de β_i . Una vez construido el intervalo, contrastar una hipótesis $H_0 : \beta_i = \beta_i^0$ equivale a rechazar la hipótesis nula si β_i^0 está fuera del intervalo de confianza y no rechazarla si β_i^0 está dentro del intervalo de confianza.

2.7.2. Regiones de confianza para varios coeficientes.

Cuando se busca un rango de valores para varios coeficientes, no se tiene un intervalo, se tiene una región de confianza. Análogamente a la anterior sección, utilizaremos el estadístico F . La matriz \mathbf{R} será la que seleccione los coeficientes para los que queremos hallar la región de confianza, y el estadístico resultante será:

$$\frac{(\hat{\beta}_2 - \beta_2)' [(\mathbf{X}'\mathbf{X})_2^{-1}]^{-1} (\hat{\beta}_2 - \beta_2) / s}{\hat{\sigma}_u^2} \sim F_{s, N-k}$$

donde β_2 denota el subvector cuya región de confianza se quiere construir y $(\mathbf{X}'\mathbf{X})_2^{-1}$ denota la submatriz correspondiente extraída de la matriz $(\mathbf{X}'\mathbf{X})^{-1}$. Si colocamos los coeficientes que nos interesan al final del modelo, ya hemos visto que $(\mathbf{X}'\mathbf{X})_2^{-1} = (\mathbf{X}_2' \mathbf{M}_1 \mathbf{X}_2)^{-1}$, con $\mathbf{M}_1 = \mathbf{I}_N - \mathbf{X}_1 (\mathbf{X}_1' \mathbf{X}_1)^{-1} \mathbf{X}_1'$, y la región de confianza se obtiene mediante:

$$1 - \alpha = P \left[\left(\hat{\beta}_2 - \beta_2 \right)' (X_2' M_1 X_2) \left(\hat{\beta}_2 - \beta_2 \right) \leq \lambda_\alpha s \hat{\sigma}_u^2 \right]$$

Con λ_α un valor tal que $P(F \leq \lambda_\alpha) = 1 - \alpha$ en una distribución $F_{s, N-k}$. Es conveniente recordar que $(X_2' M_1 X_2)$ es la submatriz de $(X' X)$ correspondiente a los coeficientes considerados.

La dificultad surge de que la expresión a la izquierda de la desigualdad es un polinomio de grado igual al número de coeficientes considerados, por lo que estas regiones sólo son útiles para el caso de dos coeficientes. La región de confianza consiste en una elipse, cuyas dimensiones dependen de las varianzas con las que se hayan estimado los coeficientes y su inclinación de la covarianza entre las estimaciones de los coeficientes. Se puede contrastar la hipótesis conjunta examinando si el punto compuesto por los valores de la hipótesis está dentro de la región de confianza.

2.8. Estimación bajo restricciones.

En la mayoría de las ocasiones en econometría se contrasta una hipótesis de cuya validez se está razonablemente seguro, por lo que si el contraste no la rechaza se consideran como válidas en el modelo que se esté considerando. Dado que el concepto de potencia de un contraste hace que una hipótesis nula no se descarte aun siendo errónea, conviene contrastar únicamente hipótesis nulas aceptables desde el punto de vista conceptual.

Parece lógico entonces que si contrastamos una hipótesis nula y ésta no se rechaza, el modelo ganaría en precisión incorporando esta hipótesis al proceso de estimación. Veamos cómo.

Buscaremos un estimador que minimice la suma de cuadrados, pero esta vez imponiendo un conjunto de restricciones lineales, $R\beta = r$. El lagrangiano del problema será:

$$L = (y - X\beta)'(y - X\beta) - 2\lambda'(R\beta - r)$$

donde λ es un vector de dimensiones $q \times 1$, siendo q el número de restricciones a considerar, de multiplicadores de Lagrange. Minimizamos el lagrangiano tomando derivadas:

$$\begin{aligned} \frac{\partial L}{\partial \beta} &= -2X'y + 2X'X\beta - 2R'\lambda = 0_k \quad (k \text{ derivadas}) \\ \frac{\partial L}{\partial \beta} &= -2(R\beta - r) = 0_q \quad (q \text{ derivadas}) \end{aligned}$$

las soluciones a este sistema de ecuaciones son el estimador de **mínimos cuadrados restringidos**, MCR , y el vector de precios sombra (multiplicadores de Lagrange) de las restricciones.

Premultiplicando la primera ecuación por $R(X'X)^{-1}$:

$$R\hat{\beta}_R - R(X'X)^{-1}X'y - R(X'X)^{-1}R'\lambda = 0_k$$

Y como $\hat{\beta}_{MCO} = (X'X)^{-1}X'y$, y según la segunda ecuación $R\hat{\beta}_R = r$, tenemos que:

$$\lambda = \left[R(X'X)^{-1}R' \right]^{-1} (r - R\hat{\beta}_{MCO})$$

Y, sustituyendo λ en la primera ecuación:

$$\begin{aligned} (X'X)\hat{\beta}_R - X'y - R' \left[R(X'X)^{-1}R' \right]^{-1} (r - R\hat{\beta}_{MCO}) &= 0_k \\ \hat{\beta}_R &= \hat{\beta}_{MCO} + (X'X)^{-1}R' \left[R(X'X)^{-1}R' \right]^{-1} (r - R\hat{\beta}_{MCO}) \end{aligned}$$

donde $\hat{\beta}_R$ es el estimador de mínimos cuadrados restringidos del modelo. La interpretación que podemos inferir es que el estimador restringido es una corrección del estimador sin restringir. Esta corrección será mayor cuanto más lejos esté el estimador MCO de satisfacer las restricciones. Hay una serie de características del estimador restringido:

1. El estimador MCR es insesgado sólo si las restricciones bajo las que se ha obtenido son ciertas.
2. El estimador MCR satisface las restricciones.
3. El estimador MCR difiere del estimador MCO sólo si éste no satisface las restricciones en H_0 , si el estimador MCO satisface las restricciones coincide con el estimador MCR .

4. La matriz de covarianzas del estimador MCR es siempre inferior a la matriz de covarianzas del estimador MCO , incluso si las restricciones no son ciertas. Esto ocurre porque al imponer las restricciones limitamos la región del espacio paramétrico en que buscamos el estimador.

La matriz de covarianzas del estimador será:

$$\begin{aligned} Var(\hat{\beta}_R) &= \sigma_u^2 \left[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left[\mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right]^{-1} \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \right] \\ Var(\hat{\beta}_{MCO}) - Var(\hat{\beta}_R) &= \sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \left[\mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}' \right]^{-1} \mathbf{R} (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

la diferencia es una matriz semidefinida positiva, es decir, los elementos de la diagonal principal de $Var(\hat{\beta}_R)$ serán menores que los de $Var(\hat{\beta}_{MCO})$.

2.9. Contraste de cambio estructural: Test de Chow.

Un contraste bastante importante es el que contrasta la hipótesis nula de que dos submuestras han sido generadas por la misma estructura, es decir, el mismo modelo. Se utiliza cuando se tiene información acerca de un cambio que pueda afectar al modelo en un momento dado y se quiere contrastar si esa variación afecta a los coeficientes del modelo. Se considera el modelo restringido, MR :

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i \quad i = 1, 2, \dots, N_1, N_1 + 1, \dots, N$$

y el modelo sin restringir, MSR :

$$\begin{aligned} y_i &= \mathbf{x}'_i \boldsymbol{\beta}_1 + u_i & i = 1, 2, \dots, N_1 \\ y_i &= \mathbf{x}'_i \boldsymbol{\beta}_2 + u_i & i = N + 1, \dots, N \end{aligned}$$

con una regresión diferente para cada submuestra. La hipótesis nula es: $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. A este contraste se le llama Test de Chow. Llamamos suma residual restringida, SRR , a la suma de los residuos del modelo restringido, y suma residual sin restringir, SRS , al agregado de las sumas residuales de los dos modelos sin restringir, que denotamos por SR_1 y SR_2 . El estadístico para el contraste de la hipótesis de ausencia de cambio estructural será:

$$F = \frac{\frac{SRR - (SR_1 + SR_2)}{k}}{\frac{SR_1 + SR_2}{N - 2k}} \sim F_{k, N-2k}$$

Si el estadístico F es mayor que el valor de $F_{k, N-2k}$ para el nivel de significación contemplado, rechazaremos la hipótesis nula de ausencia de cambio estructural.

Muchas veces se utiliza este contraste para comprobar si las últimas observaciones recibidas suponen un cambio respecto al resto de la muestra. En estos casos puede que no se pueda estimar el segundo modelo sin restricciones por falta de grados de libertad.

Si estamos en el caso límite en el que $N_2 = k$, los residuos son cero, y por tanto $SR_2 = 0$, por lo que el estadístico se reduce a:

$$F = \frac{(SRR - SR_1) / T_2}{SR_1 / (T_1 - k)} \sim F_{T_2, T_1 - k}$$

Si tenemos que $T_2 < k$ se puede demostrar que el estadístico anterior sigue siendo válido.

2.10. Predicción en el modelo lineal.

El objeto final de los modelos lineales es, una vez estimados los mismos, utilizarlos para hacer predicciones sobre la variable endógena conocidos los valores de las variables explicativas. Esto tiene sentido ya que el modelo representa la relación entre las variables, y es válido a menos que la relación sea muy inestable.

2.10.1. Cálculo de las predicciones.

El estimador MCO refleja la mejor relación lineal entre las variables explicativas y la variable endógena para la muestra de la que disponemos. Suponemos que esa relación estimada es también la mejor relación entre las

variables fuera de la muestra. Bajo este supuesto, denotamos por E_T el valor esperado en base a la información disponible hasta el momento T . La función a utilizar para predecir y_{T+1} será:

$$E_T y_{T+1} = E_T (\mathbf{x}'_{T+1} \boldsymbol{\beta} + u_{T+1}) = E_T (\mathbf{x}'_{T+1} \boldsymbol{\beta}) + E_T u_{T+1} = E_T (\mathbf{x}'_{T+1}) \hat{\boldsymbol{\beta}}_T + E_T u_{T+1}$$

Por tanto, para predecir $\mathbf{x}'_{T+1} \boldsymbol{\beta}$ multiplicamos los valores previstos para las \mathbf{x}'_{T+1} por el estimador *MCO* de los coeficientes. Este estimador lo representamos como $\hat{\boldsymbol{\beta}}_T$ para explicitar que se ha obtenido con datos hasta T . Hemos supuesto que es lo suficientemente estable como para poder usarlo para predecir y_{T+1} . Así pues, necesitaremos predecir el vector de variables explicativas y el término de error.

Las variables explicativas pueden ser conocidas de antemano (ventas en función de precios, demanda de inversión en función de saldos monetarios) o puede ser necesario estimarlas.

En cuanto al término de error, dado que es una sucesión de variables aleatorias independientes entre sí y la muestra no nos proporciona ninguna información acerca de él, lo predecimos con su esperanza matemática, que ya hemos visto que es cero.

Por tanto, para obtener buenas predicciones necesitaremos:

- Que la relación lineal entre las variables se mantenga fuera de la muestra.
- Que los coeficientes sean lo suficientemente estables como para que sus estimaciones obtenidas con la muestra sean una buena aproximación de los valores que se obtendrían incorporando las observaciones que queremos predecir.
- Que se conozcan los valores de las variables \mathbf{x} para los casos que queremos predecir, o que se puedan estimar de forma suficientemente fiable.
- Que el modelo esté bien especificado.
- Que el horizonte de predicción no esté muy lejano.

Por tanto, en caso de que se cumplan estos requisitos, la predicción mínimo-cuadrática de y_{T+1} sería:

$$E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T.$$

De todas formas, una predicción no es muy útil si no podemos dar un intervalo de confianza de la misma. Vamos a calcularlo.

2.10.2. Error de predicción.

El error de predicción se define como la diferencia entre el valor de la variable a predecir y la predicción obtenida:

$$e_T(1) = y_{T+1} - E_T y_{T+1} = \mathbf{x}'_{T+1} \boldsymbol{\beta} - \mathbf{x}'_{T+1} \hat{\boldsymbol{\beta}}_T + u_{T+1} = \mathbf{x}'_{T+1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) + u_{T+1}$$

Que es una variable aleatoria con un valor desconocido, puesto que su realización ocurrirá en el instante $T + 1$. Las fuentes de este error son:

- El error en la predicción de \mathbf{x}'_{T+1} .
- El error en la estimación del vector $\boldsymbol{\beta}$.
- El error estocástico inherente al modelo, u_{T+1} .

Como el estimador es insesgado, el error de predicción tiene esperanza cero. Así, cuando las variables exógenas son conocidas de antemano la predicción es insesgada.

La varianza del error de predicción será:

$$\begin{aligned} \sigma_e^2 &= E \left\{ \mathbf{x}'_{T+1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)' \mathbf{x}_{T+1} + 2 \mathbf{x}'_{T+1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) u_{T+1} + u_{T+1}^2 \right\} = \\ &= \mathbf{x}'_{T+1} E \left[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T) (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_T)' \right] \mathbf{x}_{T+1} + E(u_{T+1}^2) = \sigma_u^2 \mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{x}_{T+1} + \sigma_u^2 \end{aligned}$$

Donde se ha usado que $E(\hat{\boldsymbol{\beta}}_T u_{T+1}) = 0$, ya que u_{T+1} es independiente de los errores anteriores. De esta fórmula el único parámetro desconocido es σ_u^2 , que sustituiremos por su estimador.

2.10.3. Intervalos de confianza para la predicción.

Bajo el supuesto de normalidad del término de error, el error de predicción es combinación lineal de dos variables con distribución normal:

$$e_T(1) = -\mathbf{x}'_{T+1} (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) + u_{T+1} = -\mathbf{x}'_{T+1} (\mathbf{X}'_T \mathbf{X}_T)^{-1} \mathbf{X}'_T u_T + u_{T+1}$$

y por tanto, $e_t(1) \sim N(0, \sigma_e^2)$, donde σ_e^2 lo hemos calculado antes. Por un razonamiento análogo al realizado con las inferencias sobre coeficientes, tenemos que

$$\frac{e_T(1)}{\hat{\sigma}_e^2} = \frac{y_{T+1} - E_T y_{T+1}}{\hat{\sigma}_e^2} \sim t_{T-k}$$

y podemos utilizar esta expresión para calcular un intervalo de confianza para el valor futuro y_{T+1} . Estos resultados sólo son válidos si se cumplen los supuestos que hemos asumido para obtenerlos. En particular, si las variables \mathbf{x}_{T+1} no son conocidos con certeza, las expresiones son cotas inferiores para la varianza del error de predicción. En tales situaciones se podría emplear la desigualdad de Tchebichev, expresada como $P[|E_T y_{T+1} - y_{t+1}| \geq \lambda \sigma_e] \leq \frac{1}{\lambda^2}$, asignando a λ un valor apropiado (por ejemplo, tal que $\frac{1}{\lambda^2} = 0,05$, y suponiendo que sustituir σ_e por su estimador no supondrá un gran error. Por tanto

$$P[E_T y_{T+1} - \lambda \hat{\sigma}_e \leq y_{T+1} \leq E_T y_{T+1} + \lambda \hat{\sigma}_e] \geq 1 - \frac{1}{\lambda^2}$$

Capítulo 3

Modelo lineal con perturbaciones no esféricas. Propiedades del estimador mínimo cuadrático ordinario. El estimador de mínimos cuadrados generalizado. El estimador de máxima verosimilitud.

3.1. Modelo lineal con perturbaciones no esféricas.

Al analizar las propiedades de los estimadores *MCO*, siempre hemos supuesto que el término de error tiene una matriz de covarianzas escalar: todos sus elementos son cero, excepto los de la diagonal principal, y estos son todos iguales a σ_u^2 . Sin embargo, existen situaciones en las que la matriz de covarianzas tiene una estructura más compleja; en estas situaciones las propiedades analizadas bajo este supuesto podrían dejar de ser válidas.

Una de estas situaciones se produce cuando la matriz es diagonal, pero sus elementos diagonales son distintos unos de otros, es decir, $Var(u_i) = \sigma_i^2$, con $\sigma_i^2 \neq \sigma_j^2$ si $i \neq j$. A esta situación se la denomina **heteroscedasticidad**. Al caso en que la varianza de u_i es constante se le llama **homoscedasticidad**.

Una segunda situación ocurre cuando los términos de error de distintas observaciones no son independientes entre sí, es decir, la matriz de covarianzas no es diagonal. A esta situación se la denomina **autocorrelación**, para reflejar el hecho de que el término de error está correlado consigo mismo.

A la hora de estimar un modelo econométrico no se puede suponer la ausencia de heteroscedasticidad y autocorrelación sino que es necesario analizar en qué medida afectan a la estimación.

Veremos cómo afecta la presencia de estos fenómenos al estimador de mínimos cuadrados ordinarios, y que técnicas se pueden aplicar para minimizar sus efectos.

3.2. Propiedades del estimador mínimo cuadrático ordinario.

Dado el modelo $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, donde \mathbf{X} es la matriz de valores de las variables explicativas, \mathbf{y} el vector de valores de la variable dependiente y \mathbf{u} el vector de términos de error, que cumple $E(\mathbf{u}) = \mathbf{0}$ y $Var(\mathbf{u}) = \sigma_u^2 \boldsymbol{\Sigma}$, el estimador de mínimos cuadrados ordinarios del vector de parámetros $\boldsymbol{\beta}$ es una solución al sistema de ecuaciones normales $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$

Cuando la matriz $(\mathbf{X}'\mathbf{X})$ es invertible, el sistema tiene una solución única, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$, si no tiene infinitas soluciones. Se puede demostrar que $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}$.

3.2.1. El estimador MCO es insesgado.

Para que el estimador sea insesgado, se tiene que cumplir $E[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u}] = \mathbf{0}_k$. Si las variables explicativas son deterministas, y siempre que se cumpla la condición de que $E(\mathbf{u}) = \mathbf{0}$, vemos que

$E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{0}_N = \mathbf{0}_k$. Por tanto, el estimador MCO es insesgado independientemente de la matriz de covarianzas del término de error.

3.2.2. Matriz de covarianzas del estimador MCO.

Calculemos la varianza, teniendo en cuenta que la matriz \mathbf{X} es determinista y que $Var(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma_u^2\mathbf{\Sigma}$:

$$\begin{aligned} Var(\hat{\beta}_{MCO}) &= E\left[\left(\hat{\beta} - \beta\right)\left(\hat{\beta} - \beta\right)'\right] = E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] = \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}\mathbf{u}')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

Es fácil ver que si $\mathbf{\Sigma} = \mathbf{I}_N$ (matriz de covarianzas escalar), la varianza del estimador se reduce a $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, que es la matriz de covarianzas que se deduce aplicando el supuesto de covarianzas escalares del término de error. Es decir, la expresión que hemos deducido es la matriz de covarianzas del estimador MCO en cualquier caso. EL hecho de que la matriz de covarianzas presente esta forma hace que los estadísticos utilizados para realizar inferencias sobre el modelo no se distribuyan según las distribuciones F y t , a menos que la matriz de covarianzas sea escalar.

El estimador es una función lineal del término de error. Si el término de error presenta una distribución normal, el estimador MCO también se distribuye normalmente, con

$$\hat{\beta}_{MCO} \sim N_k\left(\beta; \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right).$$

3.3. El estimador de mínimos cuadrados generalizados.

Sabemos que el estimador MCO bajo el supuesto de $Var(\mathbf{u}) = \sigma_u^2\mathbf{I}_N$ es el estimador lineal insesgado de mínima varianza. Esta propiedad no se mantiene necesariamente si el supuesto sobre la matriz de covarianzas no se cumple, y puede haber un estimador lineal insesgado con menor varianza.

En estas circunstancias, intentamos transformar el modelo en otro con los mismos coeficientes, pero cuyo término de error tenga una matriz de covarianzas escalar. En este caso podríamos utilizar el estimador MCO, y sabríamos que es eficiente. Para ello premultiplicamos el modelo por una matriz \mathbf{P} de dimensiones $N \times N$: $\mathbf{Py} = \mathbf{PX}\beta + \mathbf{Pu}$, y denotamos $\mathbf{y}^* = \mathbf{Py}$, $\mathbf{X}^* = \mathbf{PX}$ y $\mathbf{u}^* = \mathbf{Pu}$. Realmente hemos hecho un cambio de variable en el modelo, las nuevas \mathbf{y}^* son combinaciones lineales de las \mathbf{y} antiguas y los coeficientes de esas combinaciones lineales son las filas de la matriz \mathbf{P} . Algo similar ocurre con las \mathbf{X}^* y las \mathbf{u}^* . Los coeficientes del modelo transformado son los mismos que los del modelo original. La matriz de covarianzas del término de error del nuevo modelo es: $Var(\mathbf{u}^*) = Var(\mathbf{Pu}) = \sigma_u^2\mathbf{P}\mathbf{\Sigma}\mathbf{P}'$.

Como la matriz $\mathbf{\Sigma}$ es simétrica y definida positiva, sabemos que siempre existe una matriz cuadrada no singular \mathbf{V} de modo que $\mathbf{\Sigma} = \mathbf{V}\mathbf{V}'$, o equivalentemente, $\mathbf{V}^{-1}\mathbf{\Sigma}(\mathbf{V}^{-1})' = \mathbf{I}_N$, $\mathbf{\Sigma}^{-1} = (\mathbf{V}^{-1})'\mathbf{V}^{-1}$. Así, si hacemos $\mathbf{P} = \mathbf{V}^{-1}$ el término de error del modelo transformado tiene matriz de covarianzas escalar. El estimador MCO de los parámetros del modelo transformado es:

$$\hat{\beta}_{MCG} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{y}^*$$

y se le llama estimador de mínimos cuadrados generalizados de los coeficientes del modelo original. En función de las variables originales tenemos que:

$$\hat{\beta}_{MCG} = \left(\mathbf{X}(\mathbf{V}^{-1})'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}(\mathbf{V}^{-1})'\mathbf{V}^{-1}\mathbf{y} = (\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}\mathbf{\Sigma}^{-1}\mathbf{y}$$

claramente vemos que es distinto del estimador MCO aplicado al modelo original.

3.3.1. Propiedades del estimador MCG.

Como obtenemos el estimador a partir de aplicar MCO a un modelo transformado, podemos asegurar algunas propiedades:

- El estimador MCG es insesgado: $\hat{\beta}_{MCG} = \beta + (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}\mathbf{X}^{*'}\mathbf{u}^*$ y como $E(\mathbf{u}^*) = \mathbf{0}_N$, entonces $E(\hat{\beta}_{MCG}) = \beta$.
- Matriz de covarianzas: $Var(\hat{\beta}_{MCG}) = \sigma_u^2(\mathbf{X}^{*'}\mathbf{X}^*)^{-1} = \sigma_u^2(\mathbf{X}'\mathbf{\Sigma}^{-1}\mathbf{X})^{-1}$.

- Ecuaciones normales: el estimador MCG satisface un sistema de ecuaciones normales. Si $(\mathbf{X}^{*'}\mathbf{X}^*)\hat{\beta} = \mathbf{X}^{*'}\mathbf{y}^*$ es el sistema de ecuaciones normales que satisface el estimador MCO del modelo transformado, para el estimador MCG tendremos:

$$(\mathbf{X}'\Sigma^{-1}\mathbf{X})\hat{\beta}_{MCG} = \mathbf{X}'\Sigma^{-1}\mathbf{y}$$

que es el sistema de ecuaciones normales que satisface el estimador MCG.

- Hay dos formas equivalentes de obtener el estimador MCG: descomponiendo la matriz Σ y transformando las matrices de datos del modelo original para aplicar MCO al modelo transformado, o usando la expresión matricial del estimador,
- Eficiencia: el estimador MCG es el estimador lineal insesgado de mínima varianza del vector de parámetros β , ya que el modelo transformado satisface todas las condiciones necesarias para que se le pueda aplicar el teorema de Gauss-Markov, así que su estimador MCO es el estimador lineal insesgado de mínima varianza para sus coeficientes. Como ese estimador coincide con el estimador MCG del modelo original y los coeficientes son los mismos, el estimador MCG será el de varianza mínima. Puede ocurrir que ambos estimadores coincidan aunque la matriz de varianzas del término de error no sea escalar.

3.3.2. Estimación del parámetro σ_u^2 .

Como el modelo transformado tiene matriz de covarianzas escalar, tenemos que

$$\hat{\sigma}_{MCG}^2 = \frac{\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*}{N-k} = \frac{\hat{\mathbf{u}}'_{MCG}\Sigma^{-1}\hat{\mathbf{u}}_{MCG}}{N-k}$$

donde $\hat{\mathbf{u}}_{MCG} = \mathbf{y} - \mathbf{X}\hat{\beta}_{MCG}$. Este estimador es insesgado.

3.3.3. El coeficiente de determinación.

En este contexto no podemos utilizar el estadístico R^2 como medida del ajuste del modelo. El modelo transformado puede no tener término constante, con lo que el R^2 calculado no estaría acotado entre cero y uno, además de que mide el ajuste de las variables transformadas, que no es lo que nos interesa. Calculándolo con las variables de interés tampoco se puede asegurar que esté acotado.

3.4. El estimador de máxima verosimilitud.

Vamos a obtener un estimador MV de un modelo cuyo término de error tiene una distribución normal, pero con matriz de covarianzas no escalar.

Tenemos el modelo $\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$, y sus términos de error, $\mathbf{u} \sim N(\mathbf{0}_N; \sigma_u^2\Sigma)$. La función de verosimilitud será:

$$L(\mathbf{u}/\sigma_u^2, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}}} \frac{1}{(\sigma_u^2)^{\frac{N}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_u^2}\mathbf{u}'\Sigma^{-1}\mathbf{u}}$$

Aplicamos la transformación $\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$, cuyo jacobiano será:

$$|J| = \left| \frac{\partial \mathbf{u}_i}{\partial \mathbf{y}_i} \right| = \begin{vmatrix} \frac{\partial u_1}{\partial y_1} & \cdots & \frac{\partial u_N}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial u_1}{\partial y_N} & \cdots & \frac{\partial u_N}{\partial y_N} \end{vmatrix} = |\mathbf{I}_N| = 1$$

Y por tanto, tenemos la función de verosimilitud:

$$\begin{aligned} L(\mathbf{y}, \mathbf{X}/\beta, \sigma_u^2, \Sigma) &= \frac{1}{(2\pi)^{\frac{N}{2}}} \frac{1}{(\sigma_u^2)^{\frac{N}{2}}} \frac{1}{|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma_u^2}(\mathbf{y}-\mathbf{X}\beta)'\Sigma^{-1}(\mathbf{y}-\mathbf{X}\beta)} \\ \ln L(\mathbf{y}, \mathbf{X}/\beta, \sigma_u^2, \Sigma) &= -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma_u^2 - \frac{1}{2} \ln |\Sigma| - \frac{1}{2\sigma_u^2}(\mathbf{y}-\mathbf{X}\beta)'\Sigma^{-1}(\mathbf{y}-\mathbf{X}\beta) \\ \frac{\partial L}{\partial \beta} &= -\frac{1}{\sigma_u^2} \mathbf{X}'\Sigma^{-1}(\mathbf{y}-\mathbf{X}\beta) = \mathbf{0} \\ \mathbf{X}'\Sigma^{-1}\mathbf{y} &= \mathbf{X}'\Sigma^{-1}\mathbf{X}\hat{\beta}_{MV} \end{aligned}$$

que como vemos, coincide con el sistema de ecuaciones normales del estimador MCG. Por tanto, el estimador de máxima verosimilitud del vector $\boldsymbol{\beta}$ coincide con su estimador de mínimos cuadrados generalizados. Calculamos ahora el estimador de la varianza:

$$\begin{aligned}\frac{\partial L}{\partial \sigma_u^2} &= -\frac{N}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \mathbf{u}' \boldsymbol{\Sigma}^{-1} \mathbf{u} = \mathbf{0} \\ \hat{\sigma}_{MV}^2 &= \frac{\hat{\mathbf{u}}_{MCG}' \boldsymbol{\Sigma}^{-1} \hat{\mathbf{u}}_{MCG}}{N}\end{aligned}$$

y el estimador MCG y el MV difieren sólo en el denominador por el que se divide la suma residual. El estimador MV es sesgado, puesto que $E(\hat{\sigma}_{MV}^2) = E\left(\frac{N}{N-k} \hat{\sigma}_{MCG}^2\right) = \frac{N}{N-k} \sigma_u^2$, pero podemos ver que el sesgo se hace más pequeño cuanto mayor es el tamaño muestral.

Capítulo 4

Heteroscedasticidad. Posibles causas de heteroscedasticidad. Estimación mínimo cuadrática en presencia de heteroscedasticidad. Contrastes de heteroscedasticidad. Transformación Box-Cox. Heteroscedasticidad condicional Autorregresiva.

4.1. Introducción.

Dentro de los casos en los que la matriz de covarianzas del término de error no es escalar, un caso particularmente interesante es cuando la matriz de covarianzas es diagonal, pero los elementos de la diagonal principal no son iguales entre sí. En este caso se dice que el modelo presenta **heteroscedasticidad**, siendo la varianza del término de error distinta para las distintas observaciones de la muestra. El estimados de mínimos cuadrados generalizados en estos casos pasa a llamarse de mínimos cuadrados ponderados, porque asigna diferentes ponderaciones a las distintas observaciones dependiendo de la varianza de sus términos de error.

4.2. Posibles causas de heteroscedasticidad.

- Puede ocurrir que la varianza del término de error dependa de los valores de una o varias de las variables explicativas. Así, por ejemplo, el gasto en consumo de las familias depende de su nivel de ingresos, pero una vez satisfechas las necesidades primordiales las familias con más ingresos tienen más renta disponible para repartir entre consumo y ahorro, por lo que es de suponer que la variabilidad de esta decisión sea mayor.
- Si los datos de que se dispone son agregados o promedios de valores individuales agrupados por grupos, cabe esperar que los grupos con menos miembros tengan una variabilidad mayor.
- Si se omite una variable explicativa relevante se puede producir, aparte de sesgo en los estimadores, una situación de heteroscedasticidad. Así, si en lugar de estimar el modelo $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ estimamos el modelo $y_i = \beta_1 + \beta_2 x_{2i} + v_i$, entonces $v_i = \beta_3 x_{3i} + u_i$, y por tanto, $Var(v_i) = \beta_3^2 x_{3i}^2 + \sigma_u^2$, que es creciente con el valor de x_3 . Es por esto que si se detecta que la varianza del término de error depende de una variable no incluida en el modelo es conveniente considerar su inclusión en el mismo.

4.3. Estimación mínimo cuadrática en presencia de heteroscedasticidad.

Supongamos el modelo lineal $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, y supongamos que la matriz de varianzas del término de error es:

$$\text{Var}(\mathbf{u}) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{pmatrix}$$

esta matriz de covarianzas se puede descomponer en:

$$\text{Var}(\mathbf{u}) = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{pmatrix} = \mathbf{V}\mathbf{V}'$$

y por tanto,

$$\mathbf{V}^{-1} = \mathbf{P} = \begin{pmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_N \end{pmatrix}$$

$$\mathbf{y}^* = \mathbf{V}^{-1}\mathbf{y} = \begin{pmatrix} y_1/\sigma_1 \\ y_2/\sigma_2 \\ \vdots \\ y_N/\sigma_N \end{pmatrix}$$

$$\mathbf{X}^* = \mathbf{V}^{-1}\mathbf{X} = \begin{pmatrix} x_{11}/\sigma_1 & x_{21}/\sigma_1 & \cdots & x_{k1}/\sigma_1 \\ x_{12}/\sigma_2 & x_{22}/\sigma_2 & \cdots & x_{k2}/\sigma_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N}/\sigma_N & x_{2N}/\sigma_N & \cdots & x_{kN}/\sigma_N \end{pmatrix}$$

Y por tanto,

$$\hat{\boldsymbol{\beta}}_{MCG} = \begin{pmatrix} \sum_{i=1}^N (x_{1i}^2/\sigma_i^2) & \sum_{i=1}^N (x_{1i}x_{2i}/\sigma_i^2) & \cdots & \sum_{i=1}^N (x_{1i}x_{ki}/\sigma_i^2) \\ \sum_{i=1}^N (x_{2i}x_{1i}/\sigma_i^2) & \sum_{i=1}^N (x_{2i}^2/\sigma_i^2) & \cdots & \sum_{i=1}^N (x_{2i}x_{ki}/\sigma_i^2) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^N (x_{ki}x_{1i}/\sigma_i^2) & \sum_{i=1}^N (x_{ki}x_{2i}/\sigma_i^2) & \cdots & \sum_{i=1}^N (x_{ki}^2/\sigma_i^2) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^N (x_{1i}y_i/\sigma_i^2) \\ \sum_{i=1}^N (x_{2i}y_i/\sigma_i^2) \\ \vdots \\ \sum_{i=1}^N (x_{ki}y_i/\sigma_i^2) \end{pmatrix}$$

Ahora vemos por qué se llama al estimador de mínimos cuadrados ponderado: cada observación se pondera por el inverso de la desviación típica de su término de error, y luego se aplican mínimos cuadrados ordinarios. El problema que surge con la heteroscedasticidad es que si la varianza del término de error es distinta para cada observación de la muestra, el número de parámetros a estimar crecerá con el número de observaciones, y siempre tendremos más parámetros que ecuaciones. Por tanto es preciso establecer algún supuesto acerca de cómo varía la varianza dentro de los elementos de la muestra.

Esa restricción es importante porque tanto la detección de la heteroscedasticidad como la estimación del modelo se ven condicionados a que la forma que presente la heteroscedasticidad se ajuste al supuesto específico que se haya establecido. Así, los contrastes de heteroscedasticidad sólo la detectan si muestra una determinada estructura. Además, los estimadores MCG sólo serán eficientes solamente si la heteroscedasticidad presenta la estructura que se ha supuesto al diseñar el estimador.

Por tanto puede ser buena idea utilizar el estimador MCO si se desconoce la forma de la heteroscedasticidad, ya que en este caso la ganancia en eficiencia del estimador MCG queda en entredicho. En cualquier caso, habrá que aproximar la matriz de varianzas, ya que ésta forma parte de la matriz de covarianzas del estimador MCO. Además, para estimar el parámetro σ_u^2 que acompaña a la matriz $\boldsymbol{\Sigma}$ no se puede utilizar la suma residual de la estimación MCO, ya que dicha suma residual sufre un problema de escala con respecto a la derivada de los residuos MCG, que es la que se debería utilizar.

White (1980) ha propuesto una aproximación a la matriz de covarianzas del estimador MCO que no precisa de una representación de la forma funcional de la heteroscedasticidad, por lo que no genera los sesgos que

podrían derivar de una representación incorrecta. La sugerencia de White es:

$$\text{Var}(\hat{\beta}_{MCO}) = N\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1} \frac{1}{N} (\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1}$$

Y estimamos $\frac{1}{N} (\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X})$ por $\frac{1}{N} \widehat{(\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X})} = \frac{1}{N} \sum_{i=1}^N \hat{u}_i^2 (\mathbf{x}_i \mathbf{x}_i')$. Se ha comprobado que este estimador tiene buenas propiedades y es recomendable su uso. Su comparación con la expresión $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}$ puede dar una idea de la heteroscedasticidad presente en el modelo.

Como el estimador MCG es el estimador lineal insesgado de mínima varianza, la diferencia entre su matriz de covarianzas y la del estimador MCO siempre será una matriz semidefinida negativa, y por tanto los elementos de la diagonal principal de la matriz de covarianzas del estimador MCG serán menores que los de la correspondiente al estimador MCO. Por otro lado, si el término de error sigue una distribución normal, el estimador MCG coincide con el estimador de máxima verosimilitud, y por tanto pasa a ser el estimador insesgado óptimo.

El procedimiento de estimación por MCG en presencia de heteroscedasticidad es como sigue:

1. Se estima el modelo por MCO, ignorando la heteroscedasticidad.
2. Se establece un supuesto acerca de la estructura de las σ_i^2 .
3. Se utilizan los residuos MCO para estimar la forma funcional supuesta para las σ_i^2 en el apartado anterior.
4. Se divide cada observación por la estimación $\hat{\sigma}_i$.
5. Se vuelve a estimar el modelo con las variables transformadas.

Esta forma de estimar el modelo da mayor ponderación a las observaciones cuyo término de error tiene menor varianza. Esto es lógico, ya que éstas contendrán más información acerca del modelo y un menor componente aleatorio.

4.4. Contrastes de heteroscedasticidad.

Estos contrastes se utilizan para detectar la presencia de heteroscedasticidad. Contrastan la hipótesis nula de ausencia de heteroscedasticidad. Algunos sugieren la forma funcional de la heteroscedasticidad si se rechaza la hipótesis nula, otros no proporcionan tal información.

4.4.1. Contraste de Goldfeld y Quandt.

Parte del supuesto de que la magnitud de las σ_i^2 depende de una variable z_i . Esta es generalmente una de las variables explicativas, aunque no es necesario para llevar a cabo el contraste de la hipótesis. En cualquier caso, es necesario disponer de información muestral de esa variable.

Supongamos que la dependencia es positiva (σ_i^2 crece con z_i). El contraste consiste en:

- Ordenamos las observaciones de menor a mayor valor de las z_i .
- Omitimos p observaciones en mitad de la muestra.
- Estimamos por MCO el modelo original con las $\frac{N-p}{2}$ primeras observaciones y luego con las $\frac{N-p}{2}$ últimas observaciones de la muestra. El número de observaciones omitidas debe ser lo suficientemente pequeño para que se puedan estimar los dos modelos.
- Sean SR_1 y SR_2 las sumas residuales de las dos estimaciones. Entonces, bajo el supuesto de homoscedasticidad y normalidad del término de error, se cumple que:

$$\lambda = \frac{SR_2}{SR_1} = \frac{\sigma_2^2}{\sigma_1^2} \sim F_{m,m}$$

donde $m = \left(\frac{N-p}{2}\right) - k$.

Si existe heteroscedasticidad con la estructura propuesta el valor del estadístico será muy grande, ya que hemos ordenado las observaciones según el valor de las z_i , y por tanto según el valor de las σ_i^2 , es decir, si el valor del estadístico excede al de $F_{m,m}$ para la significación elegida, rechazaremos la hipótesis nula de ausencia de heteroscedasticidad.

Si sospechamos que la varianza depende inversamente de los valores z_i , el contraste es igual, pero ordenando las observaciones de mayor a menor valor de las z_i .

En cuanto al número de observaciones a excluir, si excluimos demasiadas se pierden muchos grados de libertad para las dos regresiones, con lo que éstas pierden precisión y el contraste pierde potencia. Por otro lado, cuantas más excluyamos, más divergirán los valores de las sumas residuales y aumentará la potencia. Se recomienda no eliminar más de un tercio de las observaciones disponibles.

Si de resultados de este contraste admitimos la hipótesis nula, puede deberse a que no hemos especificado bien la dependencia de las σ_i^2 , que podría depender de una variable distinta a la que hemos supuesto. Por ello el contraste debería llevarse a cabo sucesivamente con variables de las que podamos sospechar a priori que puede depender la varianza del término de error.

4.4.2. Contraste de Breusch y Pagan.

En este caso suponemos que la varianza del término de error depende de un vector de variables \mathbf{z}_i de dimensión p : $\sigma_i^2 = h(\mathbf{z}_i' \boldsymbol{\alpha}) = h(\alpha_0 + \alpha_1 z_{1i} + \alpha_2 z_{2i} + \dots + \alpha_p z_{pi})$. Si todas las α_i menos α_0 son cero, estaremos en ausencia de heteroscedasticidad. Por tanto, si podemos estimar los coeficientes $\alpha_1, \alpha_2, \dots, \alpha_p$, un contraste de la hipótesis nula de homoscedasticidad vendría dado por un contraste conjunto de las p restricciones lineales $H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0$.

El contraste se realiza como sigue:

1. Se estima por MCO el modelo original y se obtienen los residuos.
2. Se obtiene la serie o sección cruzada de residuos normalizados al cuadrado: $\hat{e}_i = \frac{\hat{u}_i^2}{\hat{\sigma}_u^2}$, donde $\hat{\sigma}_u^2$ es la estimación MV de la varianza del término de error bajo la hipótesis nula (homoscedasticidad), es decir $\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N}$.
3. Se estima una regresión de los \hat{e}_i^2 sobre una constante y las variables \mathbf{z}_i y se obtiene la suma explicada de dicha regresión. Como el modelo tiene un término independiente la expresión $ST = SE + SR$ es válida.
4. Bajo la hipótesis nula de homoscedasticidad y supuesta una distribución normal para el término de error, el cociente $\frac{SE}{2}$ para la regresión de los \hat{e}_i^2 se distribuye según crece el tamaño muestral como una $\chi^2(p)$.

La lógica detrás del contraste es que si se cumple la hipótesis nula de homoscedasticidad las variables \mathbf{z}_i no tendrán poder explicativo sobre los residuos, y por tanto la suma explicada de su regresión deberá ser pequeña. Si $\frac{SE}{2}$ es mayor que el valor de la chi-cuadrado para el nivel de significación escogido, se rechaza la hipótesis nula de homoscedasticidad.

En ese caso se podría dividir cada observación por $\sqrt{\mathbf{z}_i' \boldsymbol{\alpha}}$ como una aproximación a la desviación típica para cada elemento de la muestra. La estimación MCO de este modelo transformado sería la estimación MCG del modelo original. En general no es muy recomendable, porque no está asegurado que la aproximación lineal de la función sea lo suficientemente buena para realizar con ella la estimación MCG.

Se utilizan los cuadrados de los residuos porque lo importante es la magnitud de los residuos, no su signo.

Para que este contraste funcione no es necesario conocer la forma funcional de la dependencia, se entiende la aproximación lineal como una aproximación en serie de Taylor lo suficientemente buena para los propósitos del contraste.

La lista de variables \mathbf{z}_i debería ser corta y no incluir muchas variables que no estén presentes en el modelo, ya que esto nos induciría a pensar que el modelo está mal especificado y convendría incluir las variables de las que depende la varianza.

En presencia de estacionalidad las \mathbf{z}_i deben incluir variables ficticias estacionales. Los cuadrados de las variables explicativas también son candidatos a formar parte del vector \mathbf{z}_i .

4.4.3. Contraste de Glesjer.

Intenta determinar la estructura de la heteroscedasticidad, no limitándose a estructuras lineales. Sin embargo, sólo resulta útil cuando se cree que la heteroscedasticidad puede explicarse con una sola variable, quizás junto con un término constante. El contraste se realiza así:

- Estimar el modelo MCO y obtener los correspondientes residuos.
- Estimar una regresión del valor absoluto de los residuos o su cuadrado sobre una potencia de la variable z_i , es decir $|\hat{u}_i| = \delta_0 + \delta_1 z_i^h + v_i$, para distintos valores del exponente h : $h = \left\{-1, 1, \frac{1}{2}, -\frac{1}{2}\right\}$. Escoger el valor de h que proporciona una mejor regresión. Para esto debe tenerse en cuenta la significación del coeficiente δ_1 y el valor de la suma residual.
- Una vez seleccionado el parámetro h , se estima el modelo por MCO, pero con las variables divididas por $\delta_0 + \delta_1 z_i^h$ si se estimó una regresión para $|\hat{u}_i|$, o por $\sqrt{\delta_0 + \delta_1 z_i^h}$ si se estimó la regresión de \hat{u}_i^2 . Con esto se obtiene el estimador MCG del modelo original.

Si ningún h produce una regresión aceptable, puede ser que el supuesto inicial de que la variable z_i explique la estructura de la heteroscedasticidad sea incorrecto. Se debería repetir el proceso con otra variable. De hecho, aunque la regresión sea aceptable, nada nos indica que no pueda existir otra variable con una regresión al menos tan buena como la utilizada. Por tanto, deberían utilizarse todas las alternativas que tengan sentido.

4.4.4. Contraste de Harvey.

Contrasta la heteroscedasticidad del tipo $\sigma_i^2 = e^{z_i' \alpha}$, que incluye como caso particular la expresión $\sigma_i^2 = \sigma_u^2 x_i^s$ ($z_i = (\ln \sigma_u, \ln x_i)$ y $\alpha = (2, s)$). El contraste consta de las siguientes etapas:

- Estimar el modelo por MCO ignorando la posible heteroscedasticidad y obtener los residuos \hat{u}_i .
- Estimar por MCO la regresión $\ln \hat{u}_i^2 = z_i' \alpha + \varepsilon_i = \alpha_1 + \alpha_2 z_{2i} + \dots + \alpha_p z_{pi} + \varepsilon_i$.
- El estadístico F para el contraste de significación global para esta regresión sigue una distribución chi-cuadrado con $p - 1$ grados de libertad, y puede interpretarse como un contraste de la hipótesis nula de homoscedasticidad, de forma similar al contraste de Breusch-Pagan.

Harvey además demostró que las estimaciones de los α_i son consistentes (disminuye su sesgo al aumentar el tamaño de la muestra) excepto para el término independiente, que es sesgado con independencia del tamaño muestral.

Este sesgo no tiene influencia en el contraste, ya que no se contrasta su significación, ni tampoco a la hora de estimar los parámetros del modelo por MCG, ya que al ser la varianza del término de error $\sigma_i^2 = e^{\alpha_1} e^{\alpha_2 z_{2i} + \dots + \alpha_p z_{pi}}$ el término e^{α_1} forma parte del multiplicador constante de la matriz de varianzas, que no interviene en la estimación de $\hat{\beta}$ en los MCG. Si se rechaza la hipótesis de homoscedasticidad, habría que dividir las observaciones de cada período por $\sqrt{e^{\alpha_2 z_{2i} + \dots + \alpha_p z_{pi}}}$. La matriz de covarianzas del estimador se obtiene multiplicando la matriz $(X^{*'} X^*)^{-1}$ por la estimación de σ_u^2 , es decir, $\hat{\sigma}_u^2 = \frac{\hat{u}' \Sigma^{-1} \hat{u}}{N - k}$.

4.4.5. Contraste de White.

Este contraste, propuesto por White en 1980, no precisa especificar la estructura de la heteroscedasticidad. Sus pasos son:

- Estimar el modelo original por MCO, ignorando la posible heteroscedasticidad.
- Estimar una regresión del cuadrado de los residuos del modelo MCO sobre una constante, los regresores del modelo original, sus cuadrados, y los productos cruzados de segundo orden.
- Al aumentar el tamaño muestral, el producto NR , donde N es el tamaño muestral y R^2 el coeficiente de determinación de la regresión de los residuos sigue una distribución chi-cuadrado con $p - 1$ grados de libertad, donde p es el número de regresores de la estimación de los residuos.

El tamaño muestral crece con el número de observaciones, pero el coeficiente de determinación tenderá a cero bajo la hipótesis nula de homoscedasticidad. Sólo cuando la varianza del término de error depende de las variables explicativas del modelo, el coeficiente R^2 no tiende a cero, y el producto NR permanecerá en un cierto nivel lejos de cero y debería superar al valor de chi-cuadrado para un determinado nivel de significación.

4.5. Transformación de Box-Cox.

Consideremos el modelo no lineal:

$$y_i = e^{\mathbf{x}'_i \boldsymbol{\beta}} e^{u_i}$$

donde u_i es un ruido blanco con distribución normal $N(0, \sigma_u^2)$. Este modelo tiene una distribución log-Normal (la distribución de su logaritmo es la Normal), ya que $\ln y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i$, y además,

$\text{Var}(y_i) = \text{Var}(e^{u_i}) \left(e^{\mathbf{x}'_i \boldsymbol{\beta}} \right)^2 = \left(e^{\mathbf{x}'_i \boldsymbol{\beta}} \right)^2 e^{\sigma_u^2} (e^{\sigma_u^2} - 1)$ y por tanto y_i tiene heteroscedasticidad, pero su logaritmo es homoscedástico. Por tanto, el estimador MCO aplicado al modelo logarítmico es eficiente.

La transformación logarítmica es un caso particular de una familia de transformaciones propuestas por Box y Cox en 1964 consistentes en suponer que existe un valor λ tal que

$$g(y_i, \lambda) = \frac{y_i^\lambda - 1}{\lambda} = \mathbf{x}'_i \boldsymbol{\beta} + u_i$$

donde u_i es ruido blanco con distribución Normal. Si $\lambda = 0$, tenemos que $g(y_i, \lambda) = \ln y_i$, si $\lambda = 1$ tenemos el caso lineal.

4.6. Heteroscedasticidad Condicional Autorregresiva (ARCH).

Algunas variables econométricas presentan períodos de relativa estabilidad seguidos de períodos de gran volatilidad (rentabilidad de activos financieros). En tales casos, parece adecuado especificar un modelo del tipo $y_i = \mathbf{x}'_i \boldsymbol{\beta} + u_i$, donde $u_i \sim N(0, \sigma_i^2)$, con $\sigma_i^2 = \delta_0 + \delta_1 u_{i-1}^2$, que refleja que la varianza del término de error evoluciona con cierta suavidad, alternando valores pequeños con valores elevados. El modelo es estacionario si $|\delta_1| < 1$, y además para asegurar que la varianza no sea negativa haremos $\delta_0 > 0$, $0 < \delta_1 < 1$. Bajo normalidad de u_i , el logaritmo de la verosimilitud será:

$$\ln L = -\frac{1}{2} \sum_{i=1}^N \ln(\delta_0 + \delta_1 u_{i-1}^2) - \frac{1}{2} \sum_{i=1}^N \ln \frac{u_i^2}{\delta_0 + \delta_1 u_{i-1}^2}$$

Puede probarse que la estimación MV se puede obtener por el siguiente procedimiento:

1. Estimar $\boldsymbol{\beta}$ por MCO y obtener los residuos.
2. A partir de una estimación inicial, $(\hat{\delta}_0, \hat{\delta}_1)$ obtener las variables:

$$\begin{aligned} h_i^2 &= \hat{\delta}_0 + \hat{\delta}_1 \hat{u}_{i-1}^2 & i &= 2, \dots, N \\ \mathbf{z}'_i &= \left(\frac{1}{h_i^2}, \frac{\hat{u}_{i-1}^2}{h_i^2} \right) & i &= 2, \dots, N \\ w_i &= \frac{\hat{u}_{i-1}^2}{h_i^2} - 1 & i &= 2, \dots, N \end{aligned}$$

y estimar una regresión del w_i sobre el vector \mathbf{z}'_i . Los coeficientes estimados son las correcciones a añadir a la estimación $(\hat{\delta}_0, \hat{\delta}_1)$:

$$\begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix}_j = \begin{pmatrix} \hat{\delta}_0 \\ \hat{\delta}_1 \end{pmatrix}_{j-1} + (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{W}$$

y se repite la iteración hasta que las δ converjan.

3. Una vez se dispone de $(\hat{\delta}_0, \hat{\delta}_1)$ se calcula

$$\begin{aligned} r_i^2 &= \frac{1}{h_i^2} + 2 \frac{(\hat{\delta}_1 \hat{u}_i)^2}{(h_{i+1}^2)^2} \\ c_i &= \frac{1}{h_i^2} - \frac{\hat{\delta}_1}{h_{i+1}^2} \left(\frac{\hat{u}_{i+1}^2}{h_{i+1}^2} - 1 \right) \end{aligned}$$

y se estima la regresión de $\tilde{u}_i = \frac{\hat{u}_i c_i}{r_i}$ sobre el vector $\tilde{\mathbf{x}}_i = \mathbf{x}_i r_i$ para obtener la corrección a incluir en el vector $\hat{\boldsymbol{\beta}}$: $\hat{\boldsymbol{\beta}}_j = \hat{\boldsymbol{\beta}}_{j+1} + (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{u}}$

Capítulo 5

Autocorrelación. Naturaleza y causas de la autocorrelación. Consecuencias de la autocorrelación. Contrastes de autocorrelación. Estimación de modelos con autocorrelación. Predicción.

5.1. Introducción.

En este caso examinamos modelos econométricos en los que la matriz de covarianzas del término de error no es escalar porque presenta elementos distintos de cero fuera de la diagonal principal. Esto proviene del hecho de que el término de error del modelo guarda correlación consigo mismo para distintas observaciones, por eso se llama autocorrelación al fenómeno. El objetivo es establecer una serie de contrastes que nos permitan comprobar la hipótesis nula de ausencia de autocorrelación y, en caso de rechazar esta hipótesis, estimar el modelo econométrico bajo una determinada estructura de autocorrelación.

5.2. Naturaleza y causas de la autocorrelación.

Existe autocorrelación cuando el término de error del modelo está correlacionado consigo mismo, es decir, $E(u_i u_j) \neq 0$. No es necesario que esta autocorrelación se produzca en toda la muestra, basta en algunos casos. La correlación no tiene por que producirse entre valores consecutivos. La correlación puede producirse por diversas causas:

- Existencia de ciclos o tendencias: Si la variable endógena del modelo presenta ciclos y éstos no son bien explicados por las variables exógenas del modelo, el término de error presentará autocorrelación, ya que los errores grandes tenderán a estar agrupados. Igualmente, si la variable presenta una tendencia no bien explicada por las variables explicativas, los términos de error serán negativos al principio, irán disminuyendo y se harán positivos al final.
- Variables omitidas: Si el verdadero modelo que explica el comportamiento de la variable endógena es:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

pero se estima el modelo $y_i = \beta_1 + \beta_2 x_{2i} + v_i$, entonces el término de error es $v_i = u_i + \beta_3 x_{3i}$. Si la variable x_3 está correlacionada consigo misma (tendencias, ciclos, etc...), entonces v_i presentará correlación. En este caso, la ausencia de variables en el modelo presenta otros problemas aparte de la correlación, por lo que se deberían intentar identificar si se sospecha de su presencia.

- Relaciones no lineales: Si la relación es no lineal, por ejemplo: $y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + u_i$. Si este modelo se especifica de forma lineal, nos encontraremos con una racha de residuos negativos, seguida de una racha de residuos positivos para acabar con otra racha de residuos negativos, lo que generará autocorrelación del término de error.

- Relaciones dinámicas: La mayoría de relaciones entre variables económicas se extienden a más de un período. Así, la relación entre la inflación y el crecimiento de la oferta monetaria es del tipo $\pi_t = \beta_1 + \beta_2 m_t + \beta_3 \pi_{t-1} + u_t$. Si omitimos el retardo de la variable endógena, el término de error del modelo incorporará dicha variable, mostrando autocorrelación.

5.3. Consecuencias de la autocorrelación.

Si un modelo lineal presenta autocorrelación, su estimador de mínimos cuadrados ordinarios, si bien es insesgado, ya no es el estimador lineal insesgado de mínima varianza, dicha propiedad corresponde al estimador de mínimos cuadrados generalizados. Este estimador se puede obtener de dos formas: premultiplicando las matrices de observaciones de las variables por la matriz \mathbf{V}^{-1} , siendo \mathbf{V} una matriz tal que la matriz de covarianzas del término de error del modelo se descompone como $\Sigma = \mathbf{V}\mathbf{V}'$, o resolviendo el sistema de ecuaciones normales $\hat{\beta}_{MCG} = (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{y}$. Si a pesar de todo se utiliza el estimador MCO, hay que recordar que su matriz de covarianzas es $\sigma_u^2 (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. Si no tenemos confianza en el modelo de autocorrelación y tememos que una mala especificación introduzca sesgos en el estimador MCG puede ser interesante obtener el estimador MCO para comparar. En estos casos, Newey y West han propuesto utilizar como estimador de $\frac{1}{N}\mathbf{X}'\Sigma\mathbf{X}$ la matriz

$$\mathbf{S} = \frac{1}{N} \sum_{j=0}^L \sum_{i=j+1}^N w_j \hat{u}_i \hat{u}_{i-j} [\mathbf{x}_i \mathbf{x}_{i-j}' + \mathbf{x}_{i-j} \mathbf{x}_i']$$

con $w_j = 1 - \frac{j}{L+1}$ siendo L el orden máximo de autocorrelación del término de error, que no siempre es fácil de determinar.

5.4. Contrastes de autocorrelación.

5.4.1. El contraste de Durbin-Watson.

Contrasta la hipótesis nula de ausencia de autocorrelación contra la hipótesis alternativa de autocorrelación de primer orden, del tipo $u_i = \rho u_{i-1} + \varepsilon_i$. Se utiliza el estadístico de Durbin-Watson:

$$d = \frac{\sum_{i=2}^N (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=2}^N \hat{u}_i^2}$$

La interpretación de este estadístico es la siguiente: si la correlación es positiva, valores positivos del error tenderán a estar seguidos de valores positivos del mismos, y valores negativos tenderán a estar seguidos de valores negativos. Además, dado que el estimador MCO es insesgado, los \hat{u}_i serán estimadores insesgados (aunque ineficientes) de u_i . Por tanto, las diferencias $\hat{u}_i - \hat{u}_{i-1}$ tenderán a ser menores en valor absoluto que \hat{u}_i y por tanto, $(\hat{u}_i - \hat{u}_{i-1})^2 < \hat{u}_i^2$ y el numerador de d tenderá a ser pequeño en comparación con el denominador.

Si el coeficiente de correlación fuera negativo, tendríamos tendencia a tener valores positivos seguidos de valores negativos y viceversa y por tanto el valor absoluto de $\hat{u}_i - \hat{u}_{i-1}$ tenderá a ser mayor que \hat{u}_i y el estadístico tenderá a tomar valores grandes.

Desarrollando el estadístico,

$$d = \frac{\sum_{i=2}^N \hat{u}_i^2 + \sum_{i=2}^N \hat{u}_{i-1}^2 - 2 \sum_{i=2}^N \hat{u}_i \hat{u}_{i-1}}{\sum_{i=2}^N \hat{u}_i^2}$$

Si el número de observaciones es suficientemente grande, $\sum_{i=2}^N \hat{u}_i^2 \approx \sum_{i=2}^N \hat{u}_{i-1}^2$ y entonces $d \approx 2(1 - \hat{\rho})$, ya

que $\hat{\rho} = \frac{\sum_{i=2}^N \hat{u}_i \hat{u}_{i-1}}{\sum_{i=2}^N \hat{u}_i^2}$. Como para que el término de error no sea divergente se tiene que producir que

$\rho \in [-1; 1]$, d estará entre 0 y 4, con valores próximos a cero cuando exista autocorrelación positiva de primero orden y valores próximos a 4 cuando la correlación sea negativa. Cuando no exista autocorrelación del término de error, el valor de d estará próximo a 2.

Dado que los residuos dependen de la matriz \mathbf{M} , que a su vez depende de la matriz de observaciones, \mathbf{X} , su distribución, su matriz de covarianzas, y por tanto la distribución del estadístico cambian con cada matriz de observaciones y no se pueden tabular. Hay tablas de las cotas superior e inferior de d para los niveles de

significación sobre el conjunto de todas sus posibles distribuciones. Si se quiere contrastar la hipótesis nula de ausencia de autocorrelación contra la hipótesis alternativa de autocorrelación positiva de primer orden y se obtiene un valor de d por debajo de la cota inferior, podemos afirmar que se rechaza la hipótesis nula, sean cuales sean los valores de la matriz \mathbf{X} . Si el valor de d supera la cota superior, concluiremos que no puede rechazarse la hipótesis nula. Si el valor se sitúa entre estas dos cotas, no se puede tomar ninguna decisión. Si se quiere contrastar la hipótesis nula de ausencia de autocorrelación contra la hipótesis alternativa de autocorrelación negativa de primer orden, se lleva a cabo el mismo proceso, pero para el estadístico $4 - d$. El problema de este contraste es que las cotas que se utilizan están obtenidas entre todas las distribuciones posibles de d , y por tanto son demasiado estrictas, dejando muchos casos de indefinición. Las cotas obtenidas por Durbin y Watson suponen que hay un término independiente en el modelo, y que todas las variables explicativas son deterministas. Por tanto, este supuesto no puede mantenerse si se utilizan retardos de la variable dependiente como variables explicativas. En esos casos este procedimiento produce estimaciones sesgadas del parámetro ρ , y el sesgo disminuirá el valor absoluto del estimador, lo que aumenta el riesgo de aceptación de la hipótesis nula aunque se presente autocorrelación.

5.4.2. Contraste de Wallis.

Es una extensión del contraste de Durbin-Watson para correlaciones de cuarto orden, que son habituales en datos trimestrales. En este caso, la especificación es: $u_t = \rho_4 u_{t-4} + \varepsilon_t$. Con el fin de contrastar la hipótesis nula Wallis propone un estadístico de Durbin-Watson modificado:

$$d_4 = \frac{\sum_{i=5}^N (\hat{u}_i - \hat{u}_{i-4})^2}{\sum_{i=1}^N \hat{u}_i^2}$$

De este estadístico se conocen límites superiores e inferiores para el caso de una matriz \mathbf{X} no estocástica, para modelos con término independiente y sin término independiente.

5.4.3. Contrastes de Durbin.

EL contraste de Durbin-Watson exige que la matriz de observaciones de las variables exógenas sea no estocástica, lo que no se cumple en el caso de que los regresores contengan retardos de la variable dependiente. Para el caso general Durbin desarrolló una prueba asintótica para muestras grandes. Se sigue contrastando contra la autocorrelación de primer orden, y en él es necesario especificar el conjunto completo de regresores. Así, si tenemos el modelo:

$$\begin{aligned} y_t &= \beta_1 + \beta_2 y_{t-1} + \cdots + \beta_r y_{t-r} + \beta_{r+1} x_{1t} + \cdots + \beta_{r+s} x_{st} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

con $\varepsilon_t \sim N(0, \sigma_\varepsilon^2 \mathbf{I})$. El resultado obtenido por Durbin es que bajo la hipótesis nula, $H_0 : \rho = 0$ el estadístico

$$h = \hat{\rho} \sqrt{\frac{N}{1 - N \text{Var}(\hat{\beta}_2)}} \xrightarrow{L} N(0, 1)$$

El procedimiento del contraste sería como sigue:

1. Ajustar la regresión por MCO suponiendo ausencia de heteroscedasticidad, y obtener el estimador de $\text{Var}(\hat{\beta}_2)$.
2. Partiendo de los residuos, calcular $\hat{\rho}$.
3. Calcular h , y si el valor de $N(0, 1)$ al nivel de significación elegido es menor, rechazar la hipótesis nula para ese valor de significación (en el caso de una significación del 5 %, el valor de $N(0, 1)$ para una probabilidad del 95 % es de 1,645).
4. Si la h es negativa, se puede utilizar un contraste similar para autocorrelación negativa.

Si $N \text{Var}(\hat{\beta}_2) \geq 1$ el contraste falla. Para esos casos Durbin encontró un procedimiento asintótico equivalente:

1. Estimar la regresión MCO y obtener los residuos.
2. Estimar la regresión MCO de \hat{u}_t sobre $\hat{u}_{t-1}, y_{t-1}, \dots, y_{t-r}, x_{1t}, \dots, x_{st}$.

3. Si el coeficiente de \hat{u}_{t-1} es significativamente distinto de cero utilizando el contraste de significación habitual, se rechaza la hipótesis nula.

Durbin indica que este procedimiento puede utilizarse también para contrastar perturbaciones de orden mayor que uno, añadiendo retardos del término de error a la regresión.

5.4.4. Contraste de Breusch y Godfrey.

Para llevar a cabo contrastes que supongan una estructura del término error más generales que autorregresión de primer orden, generalizando este podemos considerar el siguiente estadístico:

$$r_k = \frac{\sum_{i=1}^N \hat{u}_i \hat{u}_{i-k}}{\sum_{i=1}^N \hat{u}_i^2}$$

el caso con $k = 1$ sería la estimación MCO del parámetro ρ en el caso anterior. Los demás estadísticos serán o no significativos dependiendo de la estructura de la autocorrelación del término de error. Esta estructura puede ser del tipo $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$. El contraste introducido simultáneamente por Breusch y Godfrey considera este tipo de correlación como hipótesis alternativa.

1. Estimar el modelo por MCO y obtener los residuos. El modelo puede incluir retardos de la variable endógena sin que afecte a la validez del contraste.
2. Estimar una regresión de los residuos sobre p retardos de los mismos así como sobre las variables explicativas del modelo original. El número de retardos debe coincidir con el número de estadísticos r_k cuya significación conjunta se quiere contrastar. Obtener el valor R^2 para esta regresión.
3. Comparar NR^2 con la tabla de una distribución chi-cuadrado con p grados de libertad y rechazar la hipótesis nula de no autocorrelación si NR^2 es superior al valor de las tablas.

Dado que sabemos que los residuos son ortogonales a las variables explicativas, si se cumple la hipótesis nula de ausencia de correlación el valor de R^2 debe tender a cero mucho más rápido de lo que aumenta el tamaño de la muestra. La validez estricta del contraste se reduce al caso en el que el tamaño muestral es infinito. Este contraste es asintóticamente equivalente a la segunda versión del contraste de Durbin.

5.5. Estimación de modelos con autocorrelación.

SI el modelo presenta autocorrelación de los términos de error, los valores fuera de la diagonal de la matriz de covarianzas serán distintos de cero, por lo que dicha matriz puede contener hasta $\frac{N(N+1)}{2}$ parámetros distintos a estimar. Claramente si no tenemos ninguna información sobre la forma de la autocorrelación es una tarea sin solución. Supongamos, por ejemplo, que el término de error sigue una autocorrelación autorregresiva de primer orden: $u_t = \rho u_{t-1} + \varepsilon_t$, donde $|\rho| < 1$ es un parámetro desconocido, y ε_t es una variable aleatoria con $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2$ y $Cov(\varepsilon_t, \varepsilon_s) = 0$ si $t \neq s$. Como $|\rho| < 1$, la perturbación sólo depende de ε_t y de sus valores pasados. Además, se puede demostrar que $u_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}$, y por tanto $Var(u_t) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$; $Cov(u_t, u_s) = \sigma_{ts} = \rho^{s-t} \sigma_u^2$, con $s > t$.

$$\begin{aligned} Cov(u_{t-1}, u_t) &= \sigma_{t-1,t} = E(u_{t-1}u_t) = E[u_{t-1}(\rho u_{t-1} + \varepsilon_t)] = \rho E(u_{t-1}^2) + E(u_{t-1}\varepsilon_t) = \rho \sigma_u^2 + 0 \\ Cov(u_{t-2}, u_t) &= \sigma_{t-2,t} = E(u_{t-2}u_t) = E[u_{t-2}(\rho u_{t-1} + \varepsilon_t)] = \rho E(u_{t-2}u_{t-1}) + E(u_{t-2}\varepsilon_t) = \rho^2 \sigma_u^2 + 0 \\ &\vdots \\ Cov(u_{t-s}, u_t) &= \sigma_{t-s,t} = E(u_{t-s}u_t) = E[u_{t-s}(\rho u_{t-1} + \varepsilon_t)] = \rho E(u_{t-s}u_{t-1}) + E(u_{t-s}\varepsilon_t) = \rho^s \sigma_u^2 + 0 \end{aligned}$$

y por tanto:

$$Cov(\mathbf{u}) = \sigma_u^2 \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \dots & 1 \end{pmatrix}$$

que sólo tiene dos parámetros desconocidos, ρ y σ_u^2 . Dado que son desconocidos, será necesario estimarlos para obtener una estimación de la matriz $\mathbf{\Sigma}$, y poder estimar el modelo mediante el estimador MCG. No obstante,

hay que remarcar que el hecho de utilizar una estimación de Σ hace que este estimador no podamos afirmar que es el de mínima varianza, esto dependerá de las propiedades del estimador de Σ , $\hat{\Sigma}$. Podríamos estimar el modelo mediante MCO, y posteriormente estimar de nuevo por MCO la regresión de \hat{u}_t sobre \hat{u}_{t-1} , y de aquí obtener los estimadores $\hat{\rho}$, $\hat{\sigma}_\varepsilon^2$ y $\hat{\sigma}_u^2 = \frac{\hat{\sigma}_\varepsilon^2}{1 - \hat{\rho}^2}$.

5.5.1. Estimación mediante transformación de variables.

Tenemos nuestro modelo:

$$\begin{aligned} y_t &= \mathbf{x}'_t \boldsymbol{\beta} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

El modelo para el instante $t - 1$ es: $y_{t-1} = \mathbf{x}'_{t-1} \boldsymbol{\beta} + u_{t-1}$.

Y haciendo: $y_t - \rho y_{t-1} = \mathbf{x}'_t \boldsymbol{\beta} - \rho (\mathbf{x}'_{t-1} \boldsymbol{\beta}) + \varepsilon_t$, y definiendo como nuevas variables $y_t^* = y_t - \rho y_{t-1}$, $x_{it}^* = x_{it} - \rho x_{it-1}$ el modelo se convierte en $y_t^* = \mathbf{x}_t^{*'} \boldsymbol{\beta} + \varepsilon_t$, que no tiene heteroscedasticidad, y por tanto su estimador MCO coincide con su estimador MCG. Esta transformación ignora la primera observación muestral, lo que produce una ligera pérdida de eficiencia, irrelevante si la muestra es grande.

Podemos estimar el parámetro ρ como hemos visto antes y efectuar la transformación de variables con esta estimación.

Aplicando este método es posible que no desaparezcan las autocorrelaciones en los residuos, bien debidas a que provienen de una estimación de ρ que además es ineficiente, bien porque la estructura de la autocorrelación sea más compleja. En este último caso existe la posibilidad de iterar el método.

La transformación que se aplica equivale a descomponer la matriz de covarianzas y aplicarla a las variables del modelo para obtener términos de error esféricos.

5.5.2. Procedimiento de Cochrane-Orcutt.

Aplica la anterior transformación de modo iterativo:

- Estimar el modelo de regresión mediante MCO, ignorando la presencia de autocorrelación.
- Utilizar los residuos MCO para estimar el parámetro ρ , o bien mediante una regresión de u_t sobre u_{t-1} o bien mediante el estadístico de Durbin-Watson.
- Utilizar la estimación de ρ para obtener las variables cuasidiferenciadas.
- Estimar el modelo con las variables cuasidiferenciadas por MCO, obteniendo la estimación de los coeficientes.
- Utilizar esta estimación para generar un nuevo conjunto de residuos. Utilizar estos residuos para obtener una nueva estimación de ρ .
- Repetir hasta alcanzar el grado de convergencia deseado.

Para evitar que el procedimiento converja a un mínimo local y no global, se recomienda crear una red de búsqueda o partición del espacio paramétrico y evaluar la suma de los cuadrados de los residuos en los nodos de esa red, para tomar el valor de ρ que minimice los mismos.

Este procedimiento de estimación puede extenderse al caso de autocorrelación de orden superior a uno. Por ejemplo, si se tiene el modelo con autocorrelación $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$, las dos ecuaciones a iterar serían:

$$\begin{aligned} y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} &= (\mathbf{x}_t - \rho_1 \mathbf{x}_{t-1} - \rho_2 \mathbf{x}_{t-2})' \boldsymbol{\beta} + \varepsilon_t \\ \hat{u}_t &= \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t \end{aligned}$$

5.6. Predicción.

Una vez hemos estimado el modelo transformado la forma de hacer predicciones a partir de ese modelo. Veamos cómo se haría para una autocorrelación con un retardo. La especificación del modelo sería:

$$\begin{aligned} y_t &= \mathbf{x}'_t \boldsymbol{\beta} + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$$

con $|\rho| \leq 1$ y $Var(\varepsilon_t) = \sigma_\varepsilon^2 \mathbf{I}$. Y haciendo: $y_t - \rho y_{t-1} = \mathbf{x}'_t \boldsymbol{\beta} - \rho (\mathbf{x}'_{t-1} \boldsymbol{\beta}) + \varepsilon_t$, y definiendo como nuevas variables $y_t^* = y_t - \rho y_{t-1}$, $x_{it}^* = x_{it} - \rho x_{it-1}$ el modelo se convierte en $y_t^* = \mathbf{x}_t^{*'} \boldsymbol{\beta} + \varepsilon_t$, que no tiene

heteroscedasticidad, y por tanto su estimador MCO coincide con su estimador MCG. Aplicando el estimador así calculado para realizar una predicción puntual, tenemos:

$$\hat{y}_{t+1}^* = \mathbf{x}_{t+1}' \hat{\boldsymbol{\beta}}$$

$$\hat{y}_{t+1} = \mathbf{x}_{t+1}' \hat{\boldsymbol{\beta}} + \rho \left(y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}} \right)$$

El segundo término de la predicción es un estimador de la esperanza de u_{t+1} condicionada a u_t ya que $E(u_{t+1}|u_t) = \rho u_t = \rho(y_t - \mathbf{x}_t' \boldsymbol{\beta})$, que se estima mediante $\rho(y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}})$. La varianza de la predicción será:

$$Var(\hat{y}_t) = \hat{\sigma}_\varepsilon^2 \left(1 + \mathbf{x}_{t+1}' (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{x}_{t+1}^* \right)$$

Donde

$$\hat{\sigma}_\varepsilon^2 = \frac{\left(y_t - \mathbf{X}^{*'} \hat{\boldsymbol{\beta}} \right)' \left(y_t - \mathbf{X}^{*'} \hat{\boldsymbol{\beta}} \right)}{T - k}$$

El problema es que normalmente ρ es desconocido, y por tanto debemos estimarlo conjuntamente con $\boldsymbol{\beta}$. Por tanto, la predicción sería:

$$\hat{y}_{t+1} = \mathbf{x}_{t+1}' \hat{\boldsymbol{\beta}} + \hat{\rho} \left(y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}} \right)$$

Pero ya no conocemos exactamente las propiedades de la predicción ni podemos expresar su varianza, ya que la varianza que hemos calculado depende de ρ y no tiene en cuenta la indertidumbre al estimarlo. SI la varianza de la perturbación es mucho mayor que la varianza del estimador, se puede utilizar $\hat{\sigma}_\varepsilon^2$ como aproximación. Otra posibilidad es utilizar técnicas de bootstrapping para establecer distribuciones muestrales.

Capítulo 6

Modelos dinámicos. Justificación teórica de los modelos econométricos dinámicos. Modelos de retardos infinitos. Estimación con retardos de la variable endógena. Contraste de exogenidad de Asuman. Eficiencia relativa de los estimadores de variables instrumentales. Estimación de modelos con expectativas racionales.

6.1. Modelos dinámicos.

Trataremos en este tema de aquellos modelos econométricos en los que las relaciones entre la variable endógena y las variables explicativas no son contemporáneas, sino que aparecen

Capítulo 7

Inferencia en el modelo lineal. Contraste de hipótesis. Tratamiento general.

Contraste acerca de un coeficiente del modelo. Contraste de un subconjunto paramétrico. Contraste de significación global del modelo. Intervalos y regiones de confianza. Contraste de cambio estructural: Test de Chow. Estimación bajo restricciones. Predicción en el modelo lineal.

7.1. Inferencia en el modelo lineal.

aaa

Capítulo 8

El modelo lineal general. Especificación. Estimadores mínimo cuadrático ordinarios. Propiedades. Contraste de normalidad. Estimador máximo verosimilitud. Errores de especificación.

8.1. Introducción.