

# Índice general

<b>1. El modelo lineal general.</b>	<b>4</b>
1.1. Introducción.	4
1.2. Especificación para datos de sección cruzada y para datos en forma de serie temporal.	4
1.2.1. El modelo es lineal, estocástico y constante.	5
1.2.2. El modelo no tiene multicolinealidad.	5
1.2.3. Exogeneidad del modelo.	5
1.2.4. Muestra aleatoria.	6
1.2.5. Correlación decreciente.	6
1.2.6. Existe una relación causal entre las variables explicativas y la variable endógena:	6
1.2.7. Las variables explicativas son deterministas:	6
1.3. Estimadores mínimo cuadrático ordinarios.	6
1.4. Propiedades para muestras finitas y para grandes muestras.	7
1.4.1. Inssegadez.	7
1.4.2. Varianza del estimador MCO.	7
1.4.3. Teorema de Gauss-Markov:	8
1.4.4. Cada una de las variables explicativas es ortogonal al vector de residuos:	8
1.4.5. Expresiones de la suma residual.	8
1.4.6. El vector de residuos es una transformación lineal del término de error.	8
1.4.7. Sumas de cuadrados.	8
1.4.8. Estimación de $\sigma_u^2$ .	9
1.4.9. Bondad del ajuste.	9
1.4.10. Distribución de los estimadores.	9
1.5. Estimador del método generalizado de los momentos y del método de máxima verosimilitud.	10
1.5.1. Estimador del método generalizado de los momentos.	10
1.5.2. Estimador de máxima verosimilitud.	10
<b>2. Inferencia en el modelo lineal.</b>	<b>11</b>
2.1. Inferencia en el modelo lineal.	11
2.2. Distribución muestral de los estimadores MCO.	11
2.3. Contraste de hipótesis.	12
2.3.1. Tratamiento general del contraste de hipótesis.	12
2.3.2. Contraste de significación global del modelo.	13
2.4. Contraste acerca de un coeficiente del modelo.	13
2.5. Contraste de un subconjunto paramétrico.	13
2.6. Inferencia para grandes muestras: consistencia, eficiencia y normalidad asintótica.	14
2.6.1. Consistencia del estimador.	14
2.6.2. Normalidad asintótica	14
2.6.3. Eficiencia del estimador.	14
2.7. Contrastes para grandes muestras basados en el Multiplicador de Lagrange.	14
2.8. Predicción en el modelo lineal.	15
2.8.1. Cálculo de las predicciones.	15
2.8.2. Error de predicción.	16
2.8.3. Intervalos de confianza para la predicción.	16

<b>3. Heteroscedasticidad y autocorrelación.</b>	<b>17</b>
3.1. Introducción.	17
3.2. Causas y consecuencias de la heteroscedasticidad y de la autocorrelación.	17
3.2.1. Posibles causas de heteroscedasticidad.	17
3.2.2. Consecuencias de la heteroscedasticidad.	18
3.2.3. Posibles causas de autocorrelación.	18
3.2.4. Consecuencias de la autocorrelación.	19
3.3. Inferencia robusta a la heteroscedasticidad y a la autocorrelación.	19
3.3.1. Inferencia robusta a la heteroscedasticidad.	19
3.3.2. Inferencia robusta a la autocorrelación.	21
3.4. Estimación por mínimos cuadrados generalizados factibles.	21
3.4.1. El estimador de mínimos cuadrados generalizados.	21
3.4.1.1. Propiedades del estimador MCG.	22
3.4.1.2. Estimación del parámetro $\sigma_u^2$ .	22
3.4.1.3. El coeficiente de determinación.	22
3.4.1.4. El estimador MC ponderados?.	22
3.4.2. El estimador MCG factible.	22
3.5. Principales contrastes de heteroscedasticidad y de autocorrelación.	23
3.5.1. Contrastes de heteroscedasticidad.	23
3.5.1.1. Contraste de Breusch y Pagan.	23
3.5.1.2. Contraste de White.	23
3.5.2. Contrastes de autocorrelación.	24
3.5.2.1. Contraste de Durbin-Watson.	24
3.5.2.2. Contraste de Breusch y Godfrey.	24
<b>4. Modelo lineal: especificación.</b>	<b>25</b>
4.1. Introducción.	25
4.2. Formas funcionales lineales y no lineales en el modelo de regresión múltiple.	25
4.2.1. Formas funcionales lineales.	25
4.2.2. Formas funcionales logarítmicas.	26
4.2.3. Formas funcionales cuadráticas.	27
4.2.4. Términos de interacción.	27
4.3. Regresión múltiple con variables explicativas con información cualitativa.	28
4.3.1. Descripción de las variables cualitativas.	28
4.3.2. Interpretación de los coeficientes.	28
4.4. Regresión múltiple con variables binarias que interactúan.	28
4.4.1. Interacción entre variables binarias.	28
4.4.2. Interacción entre variables binarias y ordinarias.	29
4.5. Uso de variables proxy para variables explicativas no observables.	29
4.5.1. Uso de variables dependientes retardadas como variables proxy.	31
4.6. Errores de especificación.	31
4.6.1. Omisión de variables relevantes.	31
4.6.2. Inclusión de variables no relevantes.	32
4.6.3. Errores de especificación funcional.	32
4.7. Contraste RESET.	33
4.8. Contraste contra alternativas no anidadas.	33
<b>5. Modelo lineal con series de tiempo.</b>	<b>35</b>
5.1. Modelo lineal con series de tiempo.	35
5.1.1. Tipos de modelos con series temporales.	35
5.1.2. Propiedades del estimador MCO para muestras finitas.	37
5.2. Variables binarias para efectos temporales y variables en forma de números índice.	39
5.3. Uso de variables con tendencia en la regresión.	39
5.4. Uso de series débilmente dependientes.	39
5.5. Transformación de series altamente persistentes.	39
5.6. Tratamiento de la estacionalidad en el modelo.	39

<b>6. Modelos dinámicos.</b>	<b>40</b>
6.1. Introducción.	40
6.2. Justificación teórica de los modelos econométricos dinámicos.	40
6.3. Modelos de retardos infinitos.	40
6.4. Estimación con retardos de la variable endógena.	40
6.5. Contraste de exogeneidad de Asuman.	40
6.6. Eficiencia relativa de los estimadores de variables instrumentales.	40
6.7. Estimación de modelos con expectativas racionales.	40
<b>7. Multicolinealidad y modelos no lineales.</b>	<b>41</b>
7.1. Introducción.	41
7.2. Concepto y consecuencias.	41
7.3. Detección de la multicolinealidad.	41
7.4. Remedios contra la multicolinealidad.	41
7.5. Observaciones influyentes.	41
7.6. Especificación de modelos no lineales.	41
7.7. Aproximación lineal al modelo no lineal.	41
7.8. Mínimos Cuadrados no lineales.	41
7.9. Estimación de Máxima Verosimilitud.	41
<b>8. Datos de panel.</b>	<b>42</b>
8.1. Introducción.	42
8.2. Descripción del problema.	42
8.3. El modelo de efectos aleatorios.	42
8.4. Estimación.	42
8.5. Contraste de especificación.	42
8.6. Modelos dinámicos.	42
8.7. Identificación de efectos individuales en el estimador intragrupo.	42

# 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,  $\beta$ , 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  $\beta_k$  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  $\varepsilon_i$  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. Inssegadez.

El estimador es inssegado 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  $\beta$  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 inssegado.

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

### 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,  $Var(\epsilon) = \sigma^2 \mathbf{I}_n$ . 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  $\beta$ . Sea  $A = \tilde{A} - (X'X)^{-1}X'$ . Por tanto,

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

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  $\beta$ .

**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 variables 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 $\sigma_u^2$ .

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

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  $\text{tr}$  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  $\sigma_u^2$  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,  $\varepsilon \sim N(\mathbf{0}, \mathbf{I}_n \sigma^2)$  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  $\beta$  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  $\sigma_u^2$ .

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  $\sigma_u^2$  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.

El modelo clásico de regresión lineal propone una serie de restricciones en la distribución conjunta de las variables dependientes e independientes. Necesitamos conocer la distribución muestral de los estimadores de los parámetros del modelo, para poder realizar contrastes de hipótesis acerca de la significación de los distintos parámetros del mismo. Estos contrastes nos permitirán determinar si las hipótesis de la teoría económica se cumplen.

A modo de recordatorio, las hipótesis clásicas de partida para el modelo son las siguientes:

**Hipótesis comunes:**

- Modelo lineal.
- No multicolinealidad perfecta de las variables explicativas.
- Exogeneidad (esperanza condicionada nula).

**Hipótesis para datos de sección cruzada:**

- Muestra aleatoria.

**Hipótesis para datos de series temporales:**

- Distribución de probabilidad constante en el tiempo e independencia asintótica.

### 2.2. Distribución muestral de los estimadores MCO.

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

**Teorema 4** *Bajo los supuestos del modelo clásico de regresión lineal se cumple que:*

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

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

Y como  $\varepsilon \sim N(\mathbf{0}_n; \sigma_\varepsilon^2 \mathbf{I}_n)$  y  $\beta$  es una constante,  $\hat{\beta} \sim N(\beta; \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1})$ .

### 2.3. 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  $\sigma_\varepsilon^2$ , es decir el vector  $\varepsilon$  se distribuye según una normal multivariante  $N(\mathbf{0}_N; \sigma_\varepsilon^2 \mathbf{I}_N)$ . Como el estimador MCO es una transformación lineal del vector  $\varepsilon$ , se tiene que  $\hat{\beta}_{MCO} \sim N_k(\beta; \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1})$ .

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,  $\frac{1}{\sigma_\varepsilon^2} \varepsilon' \mathbf{M} \varepsilon \sim \chi^2(n-k)$  y como  $\varepsilon' \mathbf{M} \varepsilon = \hat{\varepsilon}' \hat{\varepsilon}$  se deduce que

$$\frac{1}{\sigma_\varepsilon^2} \hat{\varepsilon}' \hat{\varepsilon} = (N-k) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi^2(N-k).$$

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

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

Por otro lado, sabemos que  $\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \varepsilon$  y que  $\hat{\varepsilon}' \hat{\varepsilon} = \varepsilon' \mathbf{M} \varepsilon$ . 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  $\hat{\varepsilon}' \hat{\varepsilon} = (N-k) \hat{\sigma}_\varepsilon^2$  se distribuyen de forma independiente.

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

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

Con este estadístico podemos contrastar la hipótesis nula  $H_0 : \beta = \beta^0$ , sustituyendo en el estadístico el valor de  $\beta$  por  $\beta^0$  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.

#### 2.3.1. Tratamiento general del contraste de hipótesis.

El contraste que hemos visto 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,  $H_0 : \mathbf{R}\beta = \mathbf{r}$ , siendo  $\mathbf{R}$  una matriz  $q \times k$ , siendo  $q$  el número de restricciones, con los coeficientes de los parámetros  $\beta$  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  $\beta_1^0$ .

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

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

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

$$\frac{(\mathbf{R}\hat{\beta} - \mathbf{r})' [\hat{\sigma}_\varepsilon^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / q}{\frac{\hat{\varepsilon}' \hat{\varepsilon}}{N-k}} \sim F_{q, N-k}$$

O, lo que es lo mismo,  $(\mathbf{R}\hat{\beta} - \mathbf{r})' [\hat{\sigma}_\varepsilon^2 \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / q \sim F_{q, N-k}$ . 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.

### 2.3.2. 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  $\boldsymbol{\beta} = (\beta_1; \boldsymbol{\beta}_2)$ .

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

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

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

$F = \frac{\frac{SCE}{k-1}}{\frac{SCR}{N-k}} = \frac{R^2 / (k-1)}{(1-R^2) / (N-k)}$ , 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.

### 2.4. 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  $\frac{(\hat{\beta}_i - \beta_i^0)^2}{\hat{\sigma}_\varepsilon^2 a_{ii}} \sim F_{1, N-k}$ . Como

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

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  $\frac{\hat{\beta}_i}{\hat{\sigma}_\varepsilon \sqrt{a_{ii}}} \sim t_{N-k}$ , 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

$\mathbf{y} = (\mathbf{X}_1; \mathbf{X}_2) \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} + \hat{\boldsymbol{\varepsilon}} = \mathbf{X}_1 \hat{\boldsymbol{\beta}}_1 + \mathbf{X}_2 \hat{\boldsymbol{\beta}}_2 + \hat{\boldsymbol{\varepsilon}}$ . 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,

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

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

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

## 2.6. Inferencia para grandes muestras: consistencia, eficiencia y normalidad asintótica.

### 2.6.1. Consistencia del estimador.

### 2.6.2. Normalidad asintótica

Una de las hipótesis en las que nos basamos para realizar las inferencias sobre el modelo es que la distribución de los errores es normal. Esto implicaría que, para unas  $\mathbf{X}$  fijas, la distribución de las  $\mathbf{y}$  sería también normal. Esta hipótesis es muy fuerte, y no siempre se cumple.

El hecho de que no se cumpla la hipótesis de normalidad no afecta a la insesgadez ni a la eficiencia del estimador, pero nos impide obtener la distribución de los estimadores, con lo que no podemos realizar los contrastes de hipótesis que hemos visto. Es por esto que vamos a ver el siguiente teorema:

**Teorema 5** *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  $\beta$  y varianza  $\sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}$ . Además,  $\hat{\sigma}_\varepsilon^2$  es un estimador consistente de  $\sigma_\varepsilon^2$  y  $(\hat{\beta}_j - \beta_j)/\hat{\sigma}_\varepsilon a_{jj} \sim N(0; 1)$ .*

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.

La distribución del contraste es una normal, no una  $t$  de Student, porque es una aproximación, sin embargo, ya que al aumentar  $n$  la  $t$  de Student tiende a la normal tipificada, podemos utilizarla para  $n$  suficientemente grande. Por tanto, podemos realizar los contrastes basados en la  $t$  de Student exactamente igual que con el modelo clásico, siempre que  $n$  sea suficientemente grande. Hay que tener en cuenta que para aplicar esta aproximación lo que tiene que ser grande no es el tamaño de la muestra, son los grados de libertad, lo que puede dar problemas con modelos con muchas variables.

Para que este teorema se cumpla son necesarias tanto la heterocedasticidad como la exogeneidad.

### 2.6.3. Eficiencia del estimador.

## 2.7. Contrastes para grandes muestras basados en el Multiplicador de Lagrange.

Vamos a ver un contraste para muestras grandes que puede resultar útil en caso de contrastes de restricciones de exclusión múltiples. Usaremos el **estadístico del multiplicador de Lagrange**.

Bajo los supuestos de Gauss-markov, a saber:

- Linealidad del modelo.
- Observaciones obtenidas por muestreo aleatorio.
- Valor esperado condicionado nulo del término de error.
- No multicolinealidad perfecta.
- Homocedasticidad.

Supongamos que tenemos el modelo habitual, con  $n$  observaciones y  $k$  variables explicativas, y supongamos sin pérdida de generalidad que queremos contrastar la significación conjunta de las últimas  $q$  variables, es decir:

$$H_0 : \beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0$$

La hipótesis alternativa consiste en que al menos un coeficiente sea distinto de cero. La hipótesis nula impone  $q$  restricciones al modelo, así que podemos expresar el modelo restringido como:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k - qx_{k-q} + u$$

Estimamos el modelo restringido (llamamos a los estimadores restringidos  $\tilde{\beta}_i$ ). El modelo estimado será:

$$y = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \dots + \tilde{\beta}_k - qx_{k-q} + \tilde{u}$$

Si los coeficientes de las variables que hemos omitido son iguales a cero, los residuos restringidos no deberían estar correlacionados con ninguna de esas variables en la muestra. Para comprobar esto, regresamos los residuos restringidos frente a las  $k$  variables del modelo (incluimos las variables sí significativas para tener en cuenta las posibles correlaciones internas entre ellas).

Si se cumple la hipótesis nula, el  $R^2$  de la regresión debe ser muy próximo a cero, ya que los residuos estarán aproximadamente incorrelacionados con todas las variables. Bajo la hipótesis nula se puede demostrar que el tamaño muestral multiplicado por el  $R^2$  de la regresión auxiliar sigue una distribución asintótica  $\chi_q^2$ . Por tanto, el procedimiento del contraste sería:

1. Estimar el modelo restringido con  $k - q$  variables. Obtener los residuos  $\tilde{u}$ .
2. Regresar  $\tilde{u}$  sobre todas las variables independientes, y obtener el coeficiente de determinación correspondiente  $R_u^2$ .
3. Calcular el estadístico  $LM = nR_u^2$ .
4. Obtener el  $p$ -valor asociado al estadístico  $LM$  a partir de una distribución  $\chi_q^2$ . Si  $P(\chi_q^2 > LM) < \alpha$ , nivel de significatividad fijado, rechazamos la hipótesis nula, y por tanto al menos una de las  $q$  variables es significativa. Si no es así, no podemos rechazar la hipótesis nula.

En este estadístico no desempeñan ningún papel los grados de libertad, sólo importa el número de variables a contrastar. Esto se debe a que es un estadístico de naturaleza asintótica. Sin embargo, para  $n$  muy grande es posible que un valor de  $R^2$  aparentemente bajo sí refleje significatividad de las variables, debido a la forma del estadístico.

## 2.8. 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.

### 2.8.1. 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:  
 $E_T y_{T+1} = \mathbf{x}'_{T+1} \hat{\beta}_T$ .

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

### 2.8.2. Error de predicción.

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

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

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  $\beta$ .
- 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:

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

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

### 2.8.3. 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:

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

y por tanto,  $e_t(1) \sim N(0, \sigma_e^2)$ , donde  $\sigma_e^2$  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  $\lambda$  un valor apropiado (por ejemplo, tal que  $\frac{1}{\lambda^2} = 0,05$ , y suponiendo que sustituir  $\sigma_e$  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

# Heteroscedasticidad y autocorrelación.

Causas y consecuencias de la heteroscedasticidad y de la autocorrelación. Inferencia robusta a la heteroscedasticidad y a la autocorrelación. Estimación por mínimos cuadrados generalizados factibles. Principales contrastes de heteroscedasticidad y de autocorrelación.

### 3.1. Introducción.

Al analizar las propiedades de los estimadores *MCO*, entre las hipótesis de partida asumimos 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  $\sigma_u^2$ . 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. Si bien la hipótesis de matriz de covarianzas escalar no afecta a la insesgadez del estimador, sí afecta a su eficiencia y a los distintos contrastes que se realizan sobre sus términos.

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. Causas y consecuencias de la heteroscedasticidad y de la autocorrelación.

#### 3.2.1. Posibles causas de heteroscedasticidad.

La homoscedasticidad supone que la varianza del término de perturbación del modelo es constante. Si embargo, esto no es lo habitual, ni mucho menos. Hay varias razones por las que puede aparecer la heteroscedasticidad.

- Muchas variables explicativas acentúan la probabilidad de que exista una mayor variabilidad en el comportamiento de los agentes económicos. Por ejemplo, a medida que aumentan los ingresos de las familias tienen más disponibilidad de fondos una vez han cubierto las necesidades básicas, y por tanto tienen mayor margen de decisión sobre la cantidad que quieren destinar al consumo y al ahorro. En general esto ocurre porque los agentes tienen más grados de libertad en su comportamiento. Esto provoca que la variabilidad de la perturbación aumente a medida que lo hacen las variables explicativas.
- La mejora de las técnicas de recolección de datos hace que los errores disminuyan, y por tanto disminuye la variabilidad de la perturbación que corresponda a los errores de observación.
- Si los datos de los que se dispone corresponden a la agregación de datos correspondientes a distintos grupos, normalmente su variabilidad depende del tamaño del grupo que se agrega.

- Si el modelo está mal especificado, bien porque no incluya variables relevantes, bien por una transformación incorrecta de los datos (niveles vs logaritmos), también se produce heteroscedasticidad. Esta es especialmente grave, ya que puede que viole también la hipótesis de exogeneidad.

Normalmente la heteroscedasticidad es más frecuente con los datos de corte transversal, ya que se refieren a distintos individuos, aunque también puede aparecer en datos de series temporales.

### 3.2.2. Consecuencias de la heteroscedasticidad.

Si los datos presentan heteroscedasticidad, el estimador de mínimos cuadrados ordinarios sigue siendo lineal, insesgado y consistente, ya que para obtener estas propiedades no se asume en ningún momento homoscedasticidad.

Tampoco se ve afectada la validez de  $R^2$  y  $\bar{R}^2$  como medidas de la bondad del ajuste, ya que ambos son estimadores consistentes del  $R^2$  poblacional también en presencia de heteroscedasticidad.

Sin embargo, para calcular la matriz de varianzas y covarianzas del estimador sí que asumimos homoscedasticidad, por lo que los estimadores de las varianzas de los coeficientes ya no serán insesgados, y el estadístico  $t$  no se distribuirá siguiendo una  $t$  de student, con lo que no podremos calcular intervalos de confianza de los estimadores. Del mismo modo, los estadísticos  $F$  no seguirán una distribución  $F$ , y el contraste de los multiplicadores de Lagrange dejará de ser válido.

También hemos visto que el teorema de Gauss-Markov, que nos dice que el estimador de mínimos cuadrados ordinarios es el estimador lineal insesgado de mínima varianza deja de ser válido, ya que asume de forma crucial la hipótesis de homoscedasticidad. Por tanto, el estimador ya no es ELIO, ni tampoco asintóticamente eficiente.

### 3.2.3. Posibles causas de autocorrelación.

La autocorrelación afecta esencialmente a modelos con datos de series temporales, aunque también puede afectar a modelos con datos de sección cruzada; en ese caso se conoce como autocorrelación espacial. Si las observaciones con datos transversales se han generado mediante muestreo aleatorio, los datos utilizados son por definición independientes, y por tanto no puede haber autocorrelación espacial.

En el caso de series temporales es muy frecuente que el término de error esté autocorrelacionado. Algunos de los motivos son:

- 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  $v_t$  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.

### 3.2.4. Consecuencias de la autocorrelación.

Al igual que con la heteroscedasticidad, la autocorrelación no afecta a la insesgadez del estimador, siempre que las variables independientes sean exógenas.

En lo que se refiere a la consistencia, tampoco se ve afectada por la autocorrelación.

Sin embargo, el teorema de Gauss-Markov requiere de homoscedasticidad y no autocorrelación de los errores, por lo que el estimador MCO ya no será ELIO, y por tanto no será eficiente. Además, los errores estándar y los estadísticos que hemos utilizado para realizar inferencia sobre los estimadores tampoco son ya válidos, ni siquiera asintóticamente.

En cuanto a las medidas de bondad del ajuste,  $R^2$  y  $\bar{R}^2$ , no se ven afectados siempre que los datos sean estacionarios y débilmente dependientes.

## 3.3. Inferencia robusta a la heteroscedasticidad y a la autocorrelación.

### 3.3.1. Inferencia robusta a la heteroscedasticidad.

Puesto que los contrastes de hipótesis acerca de los coeficientes del modelo son fundamentales en cualquier análisis econométrico, y dado que esta inferencia basada en el estimador MCO es incorrecta si aparece heteroscedasticidad, parece que tendremos que abandonar este estimador. Sin embargo, se han desarrollado una serie de métodos para ajustar los contrastes en presencia de heteroscedasticidad de forma desconocida. Estos métodos se conocen como métodos *robustos a la heteroscedasticidad*, dado que al menos para muestras grandes son válidos aun si no tenemos homoscedasticidad.

Tenemos el modelo lineal:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

Y se cumple que  $\text{Var}(u_i|\mathbf{x}_i) = \sigma_i^2$ , con los  $\sigma_i^2$  no necesariamente iguales, y  $\text{Cov}(u_i, u_j|\mathbf{x}_i) = 0$  si  $i \neq j$ . Entonces, sabemos que:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + \left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i u_i \right) \\ \text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \text{Var}(\mathbf{u}) \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

Y, por la ley de los grandes números:

$$\begin{aligned}\left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right) &\xrightarrow{p} E(\mathbf{x}'\mathbf{x}) \\ \left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i u_i \right) &\xrightarrow{p} E(\mathbf{x}'\mathbf{u}) = \mathbf{0}\end{aligned}$$

Como  $E(\mathbf{x}'\mathbf{x})$  es no singular, por las hipótesis del modelo, aplicando las propiedades de la convergencia en probabilidad,

$$\left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \xrightarrow{p} E(\mathbf{x}'\mathbf{x})^{-1}$$

Por tanto,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left( n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i \right)^{-1} \left( n^{-1/2} \sum_{i=1}^n \mathbf{x}'_i u_i \right)$$

Como  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1} \xrightarrow{p} E(\mathbf{x}'\mathbf{x})^{-1}$ ,  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1} - E(\mathbf{x}'\mathbf{x})^{-1} \xrightarrow{p} 0$ . Por tanto,  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1} - E(\mathbf{x}'\mathbf{x})^{-1} = o_p(1)$  y  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1} = E(\mathbf{x}'\mathbf{x})^{-1} + o_p(1)$ . Como además  $(n^{-1/2} \sum_{i=1}^n \mathbf{x}'_i u_i) = O_p(1)$ , podemos escribir:

$$\sqrt{n}(\hat{\beta} - \beta) = (E(\mathbf{x}'\mathbf{x}))^{-1} \left( n^{-1/2} \sum_{i=1}^n \mathbf{x}'_i u_i \right) + o_p(1)$$

Por el teorema central del límite, podemos decir que

$$\left( n^{-1/2} \sum_{i=1}^n \mathbf{x}'_i u_i \right) \xrightarrow{d} N(\mathbf{0}; E(u^2 \mathbf{x}'\mathbf{x}))$$

Y por tanto,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}; (E(\mathbf{x}'\mathbf{x}))^{-1} E(u^2 \mathbf{x}'\mathbf{x}) (E(\mathbf{x}'\mathbf{x}))^{-1})$$

Así, la varianza asintótica de nuestro estimador será:

$$Avar(\hat{\beta}) = n^{-1} (E(\mathbf{x}'\mathbf{x}))^{-1} E(u^2 \mathbf{x}'\mathbf{x}) (E(\mathbf{x}'\mathbf{x}))^{-1}$$

Como  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1} \xrightarrow{p} E(\mathbf{x}'\mathbf{x})^{-1}$ ,  $(n^{-1} \sum_{i=1}^n \mathbf{x}'_i \mathbf{x}_i)^{-1}$  será un estimador consistente de  $E(\mathbf{x}'\mathbf{x})^{-1}$ .

Como por la ley de los grandes números  $(n^{-1} \sum_{i=1}^n u_i^2 \mathbf{x}'_i \mathbf{x}_i) \xrightarrow{p} E(u^2 \mathbf{x}'\mathbf{x})$ ,  $(n^{-1} \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}'_i \mathbf{x}_i)$  será un estimador consistente de  $E(u^2 \mathbf{x}'\mathbf{x})$ , donde hemos sustituido los valores de la perturbación, desconocidos, por los residuos del modelo MCO.

Y un estimador consistente de la varianza asintótica será:

$$\hat{Avar}(\hat{\beta}) = [n/(n-k)] (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}'_i \mathbf{x}_i \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Donde hemos añadido el multiplicador  $[n/(n-k)]$  como una corrección de grados de libertad. Este elemento asegura que, si los  $\hat{u}_i^2$  fuesen constantes para todas la  $i$  (un evento casi imposible, pero la mayor evidencia de homocedasticidad que se puede encontrar), con este método obtendríamos los errores estándar del estimador MCO. Hay cierta evidencia que esto permite mejorar el comportamiento del estimador en muestras de tamaño finito. Hay otras formas de ajustar la ecuación para conseguir esto, pero si  $n$  es muy grande con respecto a  $k$ , estos ajustes son en general equivalentes.

Los elementos de la diagonal principal de este estimador serán las varianzas robustas a la heterocedasticidad de cada coeficiente del modelo:

$$Var(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_i^2}{STC_j^2}$$

Donde  $\hat{r}_{ij}$  es el residuo de regresar  $x_j$  sobre el resto de variables explicativas, y  $STC_j$  es la suma de cuadrados de los residuos de esta regresión. La raíz cuadrada de esta cantidad es el error estándar para  $\hat{\beta}_j$  robusto a la heterocedasticidad.

A partir de estos errores estándar es sencillo obtener un estadístico  $t$  robusto a la heterocedasticidad. La forma general de un estadístico  $t$  es:

$$t = \frac{\text{valor estimado} - \text{valor a contrastar}}{\text{error estándar}}$$

Y puesto que el valor estimado lo obtenemos a partir de los estimadores MCO, el valor a contrastar viene de la hipótesis nula, y el error estándar robusto a la heterocedasticidad lo acabamos de calcular, podemos construir el estadístico. Hay que tener en cuenta que la distribución de este estadístico es asintótica, por lo que no valdría para muestras pequeñas.

En presencia de la heterocedasticidad, los estadísticos  $F$  para el contraste de hipótesis múltiples y  $LM$  dejan de ser válidos. El estadístico robusto a la heterocedasticidad equivalente al  $F$  se llama *estadístico de Wald*, y se calcula así:

La forma matricial de la hipótesis nula es:

$$H_0 : \mathbf{R}\beta = \mathbf{r}$$

Donde  $\mathbf{R}$  es una matriz  $q \times k$ , siendo  $q$  el número de restricciones a contrastar, y  $\mathbf{r}$  es  $q \times 1$

A partir de lo que hemos visto es fácil definir que  $\mathbf{R}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}; \mathbf{R}' Avar(\hat{\beta}) \mathbf{R})$ , y si se cumple la hipótesis nula,  $\mathbf{R}\beta = \mathbf{r}$ , y por tanto,

$$\mathbf{R}\hat{\beta} - \mathbf{r} \xrightarrow{d} N(\mathbf{0}; \mathbf{R}' Avar(\hat{\beta}) \mathbf{R})$$

Y, por tanto,

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left[ \mathbf{R}' \text{Avar}(\hat{\boldsymbol{\beta}}) \mathbf{R} \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \xrightarrow{d} \chi_q^2$$

También tenemos que el estimador consistente de la varianza es:

$$\hat{\text{Avar}}(\hat{\boldsymbol{\beta}}) = [n/(n-k)] (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i' \mathbf{x}_i \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Y llegamos a la conclusión de que:

$$[n/(n-k)] (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left[ \mathbf{R}' (\mathbf{X}'\mathbf{X})^{-1} \left( \sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i' \mathbf{x}_i \right) (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R} \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \xrightarrow{d} \chi_q^2$$

Que recibe el nombre de estadístico de Wald,  $W$ , es robusto a la heterocedasticidad para muestras grandes. Con él podemos contrastar la significación de un grupo de coeficientes del modelo.

### 3.3.2. Inferencia robusta a la autocorrelación.

Puesto que los contrastes de hipótesis acerca de los coeficientes del modelo son fundamentales en cualquier análisis econométrico, y dado que esta inferencia basada en el estimador MCO es incorrecta si aparece autocorrelación, parece que tendremos que abandonar este estimador. Sin embargo, se han desarrollado una serie métodos para ajustar los contrastes en presencia de autocorrelación de forma desconocida.

Consideremos el modelo de regresión lineal múltiple estándar:

$$y = \mathbf{x}\boldsymbol{\beta} + u$$

Nos interesa obtener un error estándar robusto a la autocorrelación para los  $\hat{\beta}_i$ .

## 3.4. Estimación por mínimos cuadrados generalizados factibles.

### 3.4.1. El estimador de mínimos cuadrados generalizados.

Vamos ahora a calcular un estimador del modelo que no dependa ni de la homocedasticidad ni de la ausencia de correlación. En este caso, la matriz de varianzas y covarianzas de la perturbación será de la forma:

$\text{Var}(\mathbf{u}) = \sigma_u^2 \boldsymbol{\Sigma}$ , donde  $\boldsymbol{\Sigma}$  es una matriz simétrica y definida positiva (por ser una matriz de varianzas).

Sabemos que el estimador MCO bajo el supuesto de  $\text{Var}(\mathbf{u}) = \sigma_u^2 \mathbf{I}_n$  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  $n \times n$ :

$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{u}$ , y denotamos  $\mathbf{y}^* = \mathbf{P}\mathbf{y}$ ,  $\mathbf{X}^* = \mathbf{P}\mathbf{X}$  y  $\mathbf{u}^* = \mathbf{P}\mathbf{u}$ . 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:  $\text{Var}(\mathbf{u}^*) = \text{Var}(\mathbf{P}\mathbf{u}) = \sigma_u^2 \mathbf{P}\boldsymbol{\Sigma}\mathbf{P}'$ .

Como la matriz  $\boldsymbol{\Sigma}$  es simétrica y definida positiva, sabemos que siempre existe una matriz cuadrada no singular  $\mathbf{V}$  de modo que  $\boldsymbol{\Sigma} = \mathbf{V}\mathbf{V}'$ , o equivalentemente,  $\mathbf{V}^{-1}\boldsymbol{\Sigma}(\mathbf{V}^{-1})' = \mathbf{I}_n$ ,  $\boldsymbol{\Sigma}^{-1} = (\mathbf{V}^{-1})' \mathbf{V}^{-1}$ . 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:

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

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

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

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

**3.4.1.1. 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:  $\hat{\beta}_{MCG} = \beta + (X^{*'} X^*)^{-1} X^{*'} u^*$  y como  $E(u^*) = \mathbf{0}_N$ , entonces  $E(\hat{\beta}_{MCG}) = \beta$ .
- Matriz de covarianzas:  $Var(\hat{\beta}_{MCG}) = \sigma_u^2 (X^{*'} X^*)^{-1} = \sigma_u^2 (X' \Sigma^{-1} X)^{-1}$ .
- Ecuaciones normales: el estimador MCG satisface un sistema de ecuaciones normales. Si  $(X^{*'} X^*) \hat{\beta} = X^{*'} y^*$  es el sistema de ecuaciones normales que satisface el estimador MCO del modelo transformado, para el estimador MCG tendremos:

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

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

- Hay dos formas equivalentes de obtener el estimador MCG: descomponiendo la matriz  $\Sigma$  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  $\beta$ , 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.

**3.4.1.2. Estimación del parámetro  $\sigma_u^2$ .**

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

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

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

**3.4.1.3. 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.

**3.4.1.4. El estimador MC ponderados?.****3.4.2. El estimador MCG factible.**

El problema que presenta el estimador MCG es que hay que conocer la forma de la varianza del término de error, y esto normalmente no es así, salvo casos muy excepcionales.

Sin embargo, lo que sí es más habitual es poder especificar un modelo para la función de la heterocedasticidad, y desconocer sus parámetros. Así, podremos usar los datos para estimar los parámetros de este modelo. Esto dará como resultado una estimación de cada elemento de la diagonal principal de la matriz de varianzas y covarianzas,  $h_i$  que denominaremos  $\hat{h}_i$ . Si utilizamos el estimador en la transformación del modelo, obtendremos el estimador de mínimos cuadrados generalizados **factible (MCGF)**.

Si en lugar de la estimación utilizásemos el verdadero valor de las  $h_i$  para transformar el modelo, el estimador sería insesgado, consistente y ELIO. El haber tenido que utilizar las  $\hat{h}_i$ , estimadas con los mismos datos, implica que el estimador ya no es insesgado (por lo tanto no es ELIO). Sin embargo el estimador MCGF es consistente y asintóticamente más eficiente que el MCO, convirtiéndole en una buena alternativa para muestras grandes si hay evidencia de que la heterocedasticidad aumenta los errores estándar del estimador MCO.

### 3.5. Principales contrastes de heterocedasticidad y de autocorrelación.

#### 3.5.1. Contrastes de heterocedasticidad.

Se han propuesto una serie de contrastes para detectar la presencia de heterocedasticidad en la varianza del término de error. Estos contrastes se deben utilizar para contestar a la pregunta de si la varianza condicionada es función de las variables explicativas, y no para decidir que estimador debemos usar, principalmente porque nuestra hipótesis nula es la ausencia de heterocedasticidad, lo que no nos asegura que esta no esté presente aunque no podamos rechazar la hipótesis nula.

##### 3.5.1.1. Contraste de Breusch y Pagan.

Partimos del modelo lineal general, con el conjunto de supuestos clásicos (que no incluyen la homocedasticidad), y queremos contrastar la hipótesis nula  $H_0 : Var(u_i) = E(u_i^2) = \sigma^2$ , de que el modelo presenta homocedasticidad. Si  $H_0$  se cumple, el valor esperado del cuadrado del término de error no puede estar relacionado de ninguna forma con las variables independientes. Si la hipótesis nula no se cumple,  $E(u^2)$  puede ser cualquier función de las variables independientes.

Un método simple es suponer una función lineal:

$$u^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v$$

Y en este caso, la hipótesis nula de homocedasticidad equivale a:

$$H_0 : \delta_1 = \delta_2 = \cdots + \delta_k = 0$$

Si se cumple la hipótesis nula, el término de error,  $v$ , es independiente de las  $x_i$ , y por tanto, tanto el estadístico  $F$  como el  $LM$  de significatividad global de las variables independientes son válidos. Estos estadísticos tendrán una justificación asintótica aunque  $u$  no siga una distribución normal.

Como no conocemos los errores reales de nuestra muestra, vamos a usar los residuos como estimadores de los mismos. Por tanto, estimaremos la ecuación

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v$$

Y calcularemos los estadísticos  $F$  y  $LM$  para la significatividad conjunta de  $x_1, \dots, x_k$ . Se puede demostrar que el uso de los residuos no afecta a la distribución de los estadísticos para muestras grandes.

$$F = \frac{R_{\hat{u}^2}^2/k}{(1 - R_{\hat{u}^2}^2)/(n - k - 1)} \sim F_{k;n-k-1} \quad \text{Bajo la hipótesis nula}$$

$$LM = n \cdot R_{\hat{u}^2}^2 \sim \chi_k^2 \quad \text{Bajo la hipótesis nula}$$

A la versión  $LM$  del contraste se la llama contraste de Breusch y Pagan.

##### 3.5.1.2. Contraste de White.

El supuesto de homocedasticidad puede reemplazarse, sin que dejen de tener validez los errores estándar ni los contrastes de significatividad, por el supuesto más débil de que el error al cuadrado,  $u^2$ , está incorrelacionado con las variables independientes, sus cuadrados, y sus productos cruzados. Esto llevó a White a proponer un contraste de heterocedasticidad que añade estos términos, para contrastar todas las formas de heterocedasticidad que invalidan los errores estándar y los estadísticos.

Por tanto, siguiendo el mismo principio que con el contraste de Breusch y Pagan, hay que regresar los cuadrados de los residuos del modelo MCO contra las variables independientes, sus cuadrados, y sus productos cruzados, y contrastar la significación global del modelo. El contraste de White utiliza el estadístico  $LM$ , que es asintóticamente válido.

El problema de este contraste es que los cuadrados y los productos cruzados hacen que aumenten mucho los grados de libertad del modelo, incluso para un número moderado de variables independientes.

Para soslayar este problema, podemos usar el hecho de que los valores ajustados por MCO son función de las variables independientes:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$$

Por tanto, elevando al cuadrado las  $\hat{y}_i$  obtendremos todos los productos cruzados. Así, si ajustamos el modelo:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + v$$

Y usamos los estadísticos  $F$  y  $LM$  para contrastar la hipótesis nula de que  $\delta_1 = \delta_2 = 0$  estaremos realizando un contraste equivalente al de White.

Dado que  $\hat{y}$  es un estimador de  $E(y|x)$ , esta versión del contraste es útil en aquellas situaciones en las que se espera que la varianza cambie con el nivel de valor esperado.

### 3.5.2. Contrastes de autocorrelación.

En general, los contrastes de autocorrelación se basan en el hecho de que si los errores están correlacionados también lo estarán los residuos. Por eso, el contraste más intuitivo consiste en regresar los residuos estimados por MCO respecto de los mismos residuos retardados un período, y usar el estadístico  $t$  para contrastar la significación del modelo.

Este contraste solo será válido si los regresores no están correlacionados con los errores en ningún instante, y por tanto no es correcto si existen retardos de la variable endógena.

#### 3.5.2.1. Contraste de Durbin-Watson.

Solo es válido de forma estricta si se cumplen las hipótesis de modelo lineal clásico y el modelo tiene término constante. Se basa en el estadístico de Durbin-Watson:

$$DW = \frac{\sum_{i=2}^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=2}^n \hat{u}_i^2} = 2(1 - \hat{\rho}) - \frac{\hat{u}_1^2 + \hat{u}_n^2}{\sum_{i=2}^n \hat{u}_i^2}$$

Si la muestra es suficientemente grande, podemos despreocupar el último término, y tenemos  $DW \approx 2(1 - \hat{\rho})$ .

Este contraste es conceptualmente igual al de significación de la regresión de los residuos sobre sí mismos. Si no hay autocorrelación,  $\hat{\rho} = 0$  y  $DW = 2$ . Si el valor es inferior a dos, contrastaremos correlación positiva, si es superior a dos correlación negativa. El estadístico está tabulado, y depende del nivel de significación, del tamaño de la muestra y del número de variables explicativas. Las tablas muestran dos valores,  $d_i$  o valor inferior y  $d_s$  o valor superior.

- Si  $DW < d_i$ , rechazamos la hipótesis nula de no existencia de autocorrelación, y la autocorrelación será positiva.
- Si  $DW > d_s$ , no podemos rechazar la hipótesis nula.
- Si  $d_i < DW < d_s$ , el contraste no es concluyente.

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  $4 - DW$ . Este contraste exige unas hipótesis muy fuertes, que no se cumplen con demasiada frecuencia, y por otro lado tiene una amplia región de indecisión. Es por esto que en general se prefieren otros contrastes.

#### 3.5.2.2. Contraste de Breusch y Godfrey.

Este contraste tiene como hipótesis nula la ausencia de autocorrelación frente a la hipótesis alternativa de presencia de autocorrelación lineal de orden  $q$  en el término de error. Se puede utilizar cuando el vector de regresores incorpora retardos de la variable endógena, lo cual es muy habitual en modelos de series temporales. El contraste consiste en hacer la regresión de los residuos como sigue:

$$\hat{u}_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_q \hat{u}_{t-q} + \varepsilon_t$$

Para contrastar la ausencia de autocorrelación contrastamos la significatividad de los términos asociados a los residuos retardados mediante el contraste LM. El estadístico que obtenemos es:

$$BG_{LM} = (n - q)R_u^2$$

que tiene una distribución  $\chi_q^2$ , y una región crítica de forma  $C = \{BG_{LM} > c\}$ . La razón de incluir entre los regresores a las variables independientes es que nos permite usar el contraste con variables explicativas que no sean estrictamente exógenas.

El contraste exige homocedasticidad de los residuos. También se puede contrastar la significatividad mediante un contraste  $F$ , en este caso si hubiera heterocedasticidad se podría usar la versión robusta a la heterocedasticidad del mismo.



## Capítulo 4

### Modelo lineal: especificación.

Formas funcionales lineales y no lineales en el modelo de regresión múltiple. Regresión múltiple con variables explicativas con información cualitativa. Regresión múltiple con variables binarias que interactúan. Uso de variables proxy para variables explicativas no observables. Errores de especificación. Contraste RESET. Contraste contra alternativas no anidadas.

#### 4.1. Introducción.

#### 4.2. Formas funcionales lineales y no lineales en el modelo de regresión múltiple.

##### 4.2.1. Formas funcionales lineales.

Veamos el significado de los coeficientes del modelo lineal. Si tenemos el modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

Supongamos que la variable  $x_2$  varía en una unidad manteniendo el resto de coeficientes constantes. Entonces:

$$\begin{aligned}\hat{y}_1 &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \\ \hat{y}_2 &= \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 (x_2 + 1) + \cdots + \hat{\beta}_k x_k \\ \hat{y}_2 - \hat{y}_1 &= \Delta \hat{y} = \hat{\beta}_2\end{aligned}$$

Por tanto, el coeficiente de cada variable nos dice lo que varía la estimación de la variable dependiente por cada unidad de dicha variable, manteniéndose el resto de variables independientes constantes.

EL problema con esta interpretación es que los coeficientes dependen de las unidades de medida de cada variable, y por tanto no podemos saber la importancia relativa de las variables independientes si sus unidades de medida son difíciles de interpretar (por ejemplo, las puntuaciones de un examen).

EN ese caso, podemos tipificar tanto las variables independientes como la variable dependiente. Para ello les restamos su media muestral y las dividimos por su desviación típica muestral. Así, si el modelo estimado es:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i$$

Si promediamos la ecuación para todas las unidades de la muestra, teniendo en cuenta que los residuos tienen de media cero, tenemos que:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \cdots + \hat{\beta}_k \bar{x}_k$$

Y restando esta ecuación a las  $n$  ecuaciones del modelo:

$$y_i - \bar{y} = \hat{\beta}_1 (x_{i1} - \bar{x}_1) + \hat{\beta}_2 (x_{i2} - \bar{x}_2) + \cdots + \hat{\beta}_k (x_{ik} - \bar{x}_k) + \hat{u}_i$$

Si llamamos  $\hat{\sigma}_y$  a la desviación típica muestral de la variable dependiente y  $\hat{\sigma}_i$  a la desviación típica muestral de la variable independiente  $x_i$ , tras unos cálculos sencillos llegamos a la siguiente ecuación:

$$(y_i - \bar{y})/\hat{\sigma}_y = (\hat{\sigma}_1/\hat{\sigma}_y)\hat{\beta}_1[(x_{i1} - \bar{x}_1)/\hat{\sigma}_1] + (\hat{\sigma}_2/\hat{\sigma}_y)\hat{\beta}_2[(x_{i2} - \bar{x}_2)/\hat{\sigma}_2] + \cdots + (\hat{\sigma}_k/\hat{\sigma}_y)\hat{\beta}_k[(x_{ik} - \bar{x}_k)/\hat{\sigma}_k] + (\hat{u}_i/\hat{\sigma}_y)$$

Y ajustando este modelo llegamos a los coeficientes tipificados o *coeficientes beta*,  $\hat{b}_i$ , cuya fórmula es:  
 $\hat{b}_i = (\hat{\sigma}_i / \hat{\sigma}_y) \hat{\beta}_i$ .

EL significado de estos coeficientes *beta* es el siguiente: si  $x_i$  aumenta en una desviación típica, entonces la variable dependiente,  $y$ , aumentará en  $\hat{b}_i$  desviaciones típicas. Así, podemos medir la influencia de cada variable independiente en la variable dependiente sin tener en cuenta las unidades de medida, lo que puede ser interesante.

El modelo lineal sólo tiene que ser lineal en los coeficientes para que podamos estimarlo. Así, en muchos casos es interesante transformar las variables por distintas razones.

#### 4.2.2. Formas funcionales logarítmicas.

Veamos una serie de casos en las que las variables se transforman tomando logaritmos. Por ejemplo, el modelo:

$$\log y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

Veamos la interpretación que tiene el coeficiente  $\beta_1$ : Si  $x_1$  varía manteniéndose el resto de variables constantes, tendremos:

$$\begin{aligned} \log \hat{y}_1 &= \hat{\beta}_0 + \hat{\beta}_1 \log x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \\ \log \hat{y}_2 &= \hat{\beta}_0 + \hat{\beta}_1 \log (x_1 + \Delta x_1) + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k \\ \log \hat{y}_2 - \log \hat{y}_1 &= \hat{\beta}_1 (\log (x_1 + \Delta x_1) - \log x_1) \end{aligned}$$

Y como  $\log x - \log y = \log x/y$ , y  $\log (1 + x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} x^n \approx x$  si  $|x|$  es suficientemente pequeño, podemos decir:

$$\begin{aligned} \log \hat{y}_2 - \log \hat{y}_1 &= \log \hat{y}_2 / \hat{y}_1 \approx \frac{y_2 - y_1}{y_1} \\ \log x_1 + \Delta x_1 - \log x_1 &= \log x_1 + \Delta x_1 / x_1 \approx \frac{\Delta x_1}{x_1} \end{aligned}$$

Y por tanto,

$$\beta_1 = \frac{\Delta y / y}{\Delta x_1 / x_1}$$

Es decir, el coeficiente es igual a la elasticidad de  $y$  respecto a  $x$ . Por tanto, por cada incremento de  $x_1$  en un 1 %,  $y$  se incrementa un  $\beta_1$  %. Esta aproximación solo es válida si tanto  $\Delta y / y$  como  $\Delta x_1 / x_1$  son

suficientemente pequeños como para considerar  $\sum_{n=2}^{\infty} \frac{(-1)^{n-1}}{n} x^n$  despreciable.

Si realizamos el mismo ejercicio para el coeficiente  $\beta_2$ , tenemos que:

$$\log \hat{y} + \Delta y - \log \hat{y} = \hat{\beta}_2 \Delta x_2$$

Aplicando la misma aproximación tenemos que

$$\% \hat{\Delta} y \approx 100 \hat{\beta}_2 \Delta x_2$$

Es decir, que por cada incremento en una unidad de  $x_2$ ,  $y$  se incrementa un  $100 \hat{\beta}_2$  %. Sin embargo, esta aproximación no es válida para cambios grandes en el logaritmo. EN estos casos, el valor exacto será:

$$\% \hat{\Delta} y = 100 [e^{\hat{\beta}_2 \Delta x_2} - 1]$$

y si  $x_2$  varía en una unidad,  $y$  varía un  $100 [e^{\hat{\beta}_2} - 1]$  %. Hay que tener en cuenta que este estimador no es insesgado, aunque sí es consistente. Esto se debe a que la esperanza no es invariante ante funciones no lineales, aunque la convergencia en probabilidad es invariante ante funciones continuas.

Si tuviésemos el modelo:

$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

Siguiendo los razonamientos anteriores, podremos decir que el coeficiente  $\beta_1$  nos dice que si  $x_1$  varía un 1 %,  $y$  varía  $100 \hat{\beta}_1$  unidades, siempre que la variación de  $x_1$  no sea muy grande.

Aparte de las interpretaciones de los coeficientes, hay que tener en cuenta que para los coeficientes de las variables logarítmicas son invariantes ante cambios de escala, con lo cual también son independientes de la unidades en que estén expresadas las mismas.

Si  $y > 0$ ,  $\log(y)$  puede hacer que la variable cumpla mejor las condiciones del modelo, reduciendo o incluso eliminando la heterocedasticidad y asimetrías en la variable. También ocurre que tomar logaritmos reduce el rango de variación de la variable, por lo que las estimaciones serán menos sensibles a valores extremos tanto de las variables independientes como de la dependiente. En general se suelen tomar logaritmos cuando las variables toman valores positivos muy grandes. Si la variable empieza en cero, se puede transformar mediante  $\log(1 + y)$ , sin que cambie la interpretación de los coeficientes, salvo en el caso de que  $y = 0$ .

Un inconveniente que tiene el tomar logaritmos de la variable  $y$  es que esto nos permite estimar el valor de  $\log y$ , no el de  $y$ . Además, dos modelos, uno en  $\log y$  y el otro en  $y$  no se pueden comparar mediante sus  $R^2$ .

### 4.2.3. Formas funcionales cuadráticas.

Las funciones cuadráticas son muy útiles para reflejar efectos marginales decrecientes de  $x$  sobre  $y$ . Por ejemplo, en el modelo  $y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$ , tenemos que  $\partial y / \partial x = \beta_1 + 2\beta_2 x$ , y si  $\beta_2 < 0$ , el efecto de  $x$  sobre  $y$  decrecerá a medida que aumente  $x$ .

Por tanto, para incrementos pequeños tenemos que

$$\Delta \hat{y} \approx (\hat{\beta}_1 + 2\hat{\beta}_2 x) \Delta x$$

El problema de emplear una forma funcional cuadrática es que llega un momento en que la relación entre  $x$  e  $y$  se invierte, y esto no siempre es lógico desde el punto de vista del modelo. Es decir, si  $x = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$ , el efecto de la  $x$  sobre la  $y$  es nulo, y a partir de ese punto pasa a ser de signo contrario al que tenía antes. Si esto no tiene lógica en la situación de nuestro modelo se puede deber a varios factores:

- El punto en el que el efecto se hace cero está fuera del rango de valores posibles de la  $x$ .
- Nuestra muestra tiene muy pocos valores en ese intervalo.
- Hay algún tipo de sesgo en nuestros datos.
- El modelo está infraespecificado: hay alguna otra variable que explica el comportamiento.

Para el caso de modelos del tipo:

$$\log y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$$

Se interpreta como que el efecto del cambio en la  $x$  afecta al incremento porcentual de  $y$  de forma creciente/decreciente, según el valor de los coeficientes.

Si el modelo es del tipo:

$$\log y = \beta_0 + \beta_1 \log x + \beta_2 [\log x]^2 + u$$

Se interpreta como que la elasticidad de  $y$  con respecto a  $x$  es variable, de la forma  $\hat{\beta}_1 + \hat{\beta}_2 \log x$ .

También se pueden estimar modelos con formas polinómicas de grado superior. La interpretación de los coeficientes en estos casos es más compleja.

### 4.2.4. Términos de interacción.

Veremos ahora modelos en los que las variables independientes interactúan entre ellas. Por ejemplo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + u$$

En este caso, el efecto parcial de  $x_2$  sobre  $y$ , manteniendo constantes el resto de variables será:

$$\frac{\Delta \hat{y}}{\Delta x_2} = \hat{\beta}_2 + \hat{\beta}_3 x_1$$

Si  $\beta_3 > 0$ , el efecto de  $x_2$  sobre  $y$  será mayor cuanto mayor sea  $x_1$ , es decir, las variables interactúan. Para resumir esta interacción podemos calcular el efecto de  $x_2$  sobre  $y$  para valores interesantes de  $x_1$  como su media, su mediana, su moda o algún cuantil que consideremos interesante.

Si reparametrizamos el modelo de la siguiente forma:

$$y = \alpha_0 + \delta_1 x_1 + \delta_2 x_2 + \beta_3 (x_1 - \mu_1)(x_2 - \mu_2) + u$$

Siendo  $\mu_1$  y  $\mu_2$  las medias de  $x_1$  y  $x_2$ , los coeficientes  $\delta_1$  es el efecto parcial de  $x_1$  en el valor medio de  $x_2$ , y  $\delta_2$  es el efecto parcial de  $x_2$  en el valor medio de  $x_1$ . Esto lo podemos reproducir para cualquier valor interesante de  $x_1$  y  $x_2$  que queramos considerar.

### 4.3. Regresión múltiple con variables explicativas con información cualitativa.

En muchos casos nuestros modelos tienen variables explicativas cualitativas, es decir, no numéricas, como por ejemplo el sexo de un individuo, la provincia en la que está situada una vivienda, el partido político al que se ha votado... Vamos a ver como incluir estas variables en nuestros modelos.

#### 4.3.1. Descripción de las variables cualitativas.

Si tenemos una variable cualitativa que puede tomar  $r$  valores o categorías, debemos definir  $r - 1$  variables ficticias.

Es decir, sea la característica  $z$  que puede tomar los valores  $a_1, a_2, \dots, a_r$ , definimos las variables ficticias  $z_1, z_2, \dots, z_{r-1}$ , tales que

$$z_i = \begin{cases} 1 & \text{si } z = a_i \\ 0 & \text{si } z \neq a_i \end{cases}$$

No definimos la variable  $z_k$ , porque dado que el individuo siempre ha de pertenecer a una categoría, tendríamos que  $z_k = 1 - \sum_{i=1}^{r-1} z_i$ , y esto introduciría colinealidad en el modelo.

Los valores 1 y 0 son arbitrarios, y se podrían haber elegido otros. Se eligen estos porque facilitan la interpretación de los coeficientes del modelo.

Así, el modelo quedaría:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \delta_1 z_1 + \dots + \delta_{r-1} z_{r-1} + u$$

Y podríamos realizar la estimación del modelo como habitualmente.

#### 4.3.2. Interpretación de los coeficientes.

Si se cumple el supuesto de esperanza condicionada nula para el término de error, cada coeficiente  $\delta_i$  se puede expresar como:

$$\delta_i = E(y|x_1, \dots, x_k, z_1, \dots, z_i = 1, \dots, z_{r-1}) - E(y|x_1, \dots, x_k, z_1, \dots, z_i = 0, \dots, z_{r-1})$$

Es decir, es la diferencia entre la esperanza de la variable dependiente si el individuo presenta el valor  $a_i$  en la característica  $z$  o si presenta el valor  $a_r$  (ya que el resto de los  $z_j$ , permanecen constantes y por lo tanto valdrán cero). Así, vemos que la categoría que excluimos de las variables actúa como “categoría base” con la que se comparan todas las demás. Es conveniente tener esto en cuenta a la hora de elegir dicha categoría base. En el caso de que la variable dependiente se presente en forma logarítmica, los coeficientes  $\delta$  multiplicados por 100 se interpretan como la diferencia en porcentaje entre la categoría base y la categoría del coeficiente. En el caso de que los valores que puede tomar nuestra variable sean demasiados como para desglosarlos en variables ficticias, podemos agrupar los mismos en categorías, especialmente si estamos hablando de variables ordinales.

### 4.4. Regresión múltiple con variables binarias que interactúan.

#### 4.4.1. Interacción entre variables binarias.

Puede haber modelos en los que aparezcan dos características binarias que pueden interactuar. Para reflejar estos casos definimos modelos del tipo:

$$y = \beta_0 + \delta_1 z_1 + \delta_2 z_2 + \delta_3 z_1 z_2 + u$$

En este caso, si  $z_1 = 1$ , la diferencia entre ambas categorías de la característica que refleja  $z_2$  es de  $\delta_2 + \delta_3$ , y si  $z_1 = 0$  la diferencia es de  $\delta_2$ , y de forma análoga la influencia de  $z_1$  depende de  $z_2$ . Contrastar la significatividad de  $\delta_3$  implica contrastar que las características no interactúan.

#### 4.4.2. Interacción entre variables binarias y ordinarias.

Supongamos que tenemos un modelo con la característica  $z$ , binaria, que modelamos mediante la variable  $z_1$ , y una característica numérica que modelamos mediante la variable  $x_1$  modelamos según el modelo:

$$y = \beta_0 + \delta_0 z_1 + \beta_1 x_1 + \delta_1 z_1 x_1 + u$$

En este caso, el coeficiente  $\delta_0$  refleja la diferencia entre los términos constantes de ambas categorías de la característica  $z$ , y el coeficiente  $\delta_1$  refleja la diferencia de pendiente entre las categorías. Es decir, la pertenencia a una u otra categoría también influye en el impacto de la variable  $x_1$  sobre la variable dependiente.

Para contrastar que la característica  $z$  no influye en el impacto de la variable  $x_1$ , tendremos que hacer un contraste de significación sobre el coeficiente  $\delta_1$ . Para contrastar que la característica  $z$  no influye en la variable  $y$ , tenemos que contrastar la significatividad conjunta de  $\delta_0$  y  $\delta_1$ .

Este mismo sistema se puede utilizar si queremos contrastar si hay diferencia en el modelo de regresión para dos grupos distintos de una población que se distinguen en una característica. Para ello, definimos la variable  $z_1$  tal que  $z_1 = 1$ ,  $z_1 = 2$ , refleja la pertenencia a cada grupo. Estimamos el modelo:

$$y = \beta_0 + \delta_0 z_1 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \delta_i x_i z_1 + u$$

Y para contrastar la hipótesis nula de que ambos modelos son iguales, contrastamos la significación conjunta de las variables  $\delta_i$ , es decir,  $H_0 : \delta_0 = \delta_1 = \dots = \delta_k = 0$ , frente a  $h_1 : \exists i / \delta_i \neq 0$ .

Esto también se puede contrastar usando el contraste de cambio estructural de Chow, que consiste en estimar el modelo restringido, considerando ambos grupos como una sola población, y por otro lado el modelo sin restringir, en el que consideramos ambos grupos como dos poblaciones y estimamos un modelo para cada uno de ellos. De esta forma, la suma de cuadrados de los residuos del modelo sin restringir será simplemente la suma de cuadrados de los residuos de ambos modelos, y por tanto el estadístico del contraste será:

$$F = \frac{SCR_T - (SCR_1 + SCR_2)}{(SCR_1 + SCR_2)} \frac{n - 2(k + 1)}{k + 1} \sim F_{k+1, n-2(k+1)}$$

A este contraste se le llama contraste de Chow. Como es de tipo  $F$ , requiere que exista homocedasticidad. Si realizamos un análisis asintótico no es necesaria la normalidad de los residuos.

En el contraste de Chow la hipótesis nula no permite ninguna diferencia entre los grupos, y muchas veces es útil contrastar que las pendientes no dependen de la pertenencia a un grupo, aunque el término constante sí pueda depender. Para contrastar esta situación se puede realizar, bien contrastando la significatividad conjunta únicamente de los términos de interacción. La otra es formando un estadístico  $F$  en el que la regresión del modelo restringido incluye una variable binaria ficticia en la que distinguimos entre los dos grupos.

#### 4.5. Uso de variables proxy para variables explicativas no observables.

Hay ocasiones en las que no podemos incluir en el modelo alguna variable explicativa, debido a que ésta es no observable (por ejemplo, la influencia de la habilidad en el salario de los trabajadores). El problema de esta exclusión es que hace que los estimadores de los coeficientes sean sesgados, ya que si dividimos la matriz de variables independientes,  $\mathbf{X} = (\mathbf{X}_0; \mathbf{Z})$ , siendo  $\mathbf{X}_0$  matriz  $n \times r$  de variables observadas, y  $\mathbf{Z}$  matriz  $n \times (k - r)$  matriz de variables no observadas y omitidas, y siendo  $\beta' = (\beta_0'; \beta_z')$ , siendo  $\beta_0$  vector  $r \times 1$  de coeficientes de las variables observadas, y  $\beta_z$  matriz  $(k - r) \times 1$  vector de coeficientes de las variables omitidas, el estimador de mínimos cuadrados ordinarios para el modelo con las variables omitidas será:

$$\hat{\beta}_0 = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{y}$$

Y como la expresión real del modelo es:

$$\mathbf{y} = [\mathbf{X}_0; \mathbf{Z}] \begin{bmatrix} \beta_0 \\ \beta_z \end{bmatrix} + \mathbf{u}$$

Tenemos que:

$$\hat{\beta}_0 = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' [\mathbf{X}_0; \mathbf{Z}] \begin{bmatrix} \beta_0 \\ \beta_z \end{bmatrix} + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{u}$$

Y como:

$$(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' [\mathbf{X}_0; \mathbf{Z}] = [(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_0; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Z}] = [\mathbf{I}_r; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Z}]$$

Como consecuencia:

$$E(\hat{\beta}_0) = [\mathbf{I}_r; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Z}] \begin{bmatrix} \beta_0 \\ \beta_Z \end{bmatrix} = \beta_0 + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{Z} \beta_Z$$

Con lo que si las variables omitidas son significativas el estimador será sesgado.

En general, lo que nos suele interesar es conocer el efecto de las variables observadas sobre nuestra variable dependiente. El término independiente no nos suele interesar, y dado que las variables omitidas no son observables, no tiene mucho sentido conocer su efecto sobre la variable dependiente, dado que en general no sabremos interpretar ese valor.

Para eliminar o reducir el sesgo sobre las variables omitidas, intentaremos sustituir las variables no observables por variables que sí se puedan observar y mantengan una correlación con ellas. A estas variables las llamaremos **variables proxy**.

Supongamos que tenemos el modelo:

$$\mathbf{y} = [\mathbf{X}_0; \mathbf{X}_1] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \mathbf{u}$$

visto anteriormente, con  $k - r$  variables  $\mathbf{X}_1$  no observables. Si podemos encontrar  $k - r$  variables proxy que denotamos por  $\mathbf{Z}$ , que sí podamos observar y estén relacionadas con las variables no observables, podremos escribir para cada una de ellas:

$$x_{1i} = \delta_{0i} + \delta_{1i} z_i + v_i$$

Donde  $v_i$  es la perturbación resultante de que la correlación entre ambas variables no es perfecta, y si  $\delta_{1i} = 0$ , no tendremos una buena variable proxy. El método consiste en sustituir las variables originales con las proxy. A este procedimiento se le conoce como **solución por sustitución del problema de variables omitidas**. Como estamos reemplazando unas variables por otras que no son iguales, vamos a ver cómo deben ser estas variables proxy para que los estimadores sean consistentes.

Para que los estimadores de los coeficientes de las variables no omitidas sean consistentes se tiene que cumplir:

- Si se cumple las hipótesis del modelo lineal, el error  $u$  estará incorrelacionado con las  $\mathbf{X}_0$  y  $\mathbf{X}_1$ . Además,  $u$  también debe estar incorrelacionado con las  $\mathbf{Z}$ . Esto quiere decir que si el modelo incluye las variables  $\mathbf{X}_0$  y  $\mathbf{X}_1$ , entonces las  $\mathbf{Z}$  son irrelevantes para el modelo, lo cual es lógico, dado que sólo deben influir en el modelo por su relación con las  $\mathbf{X}_1$ . Por tanto, esta hipótesis es bastante razonable.
- Los errores  $v_i$  están incorrelacionados con  $\mathbf{X}_0$ ,  $\mathbf{X}_1$  y  $\mathbf{Z}$ . Para que esto ocurra, las  $\mathbf{Z}$  deben ser buenas aproximaciones de las  $\mathbf{X}_1$ , y no deben depender de las  $\mathbf{X}_0$ .

Si estimamos el modelo sustituyendo las  $\mathbf{X}_1$  por las  $\mathbf{Z}$ :

$$\hat{\beta} = \left( \begin{bmatrix} \mathbf{X}_0' \\ \mathbf{Z}' \end{bmatrix} [\mathbf{X}_0; \mathbf{Z}] \right)^{-1} \begin{bmatrix} \mathbf{X}_0' \\ \mathbf{Z}' \end{bmatrix} [\mathbf{X}_0; \mathbf{X}_1] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \left( \begin{bmatrix} \mathbf{X}_0' \\ \mathbf{Z}' \end{bmatrix} [\mathbf{X}_0; \mathbf{Z}] \right)^{-1} \begin{bmatrix} \mathbf{X}_0' \\ \mathbf{Z}' \end{bmatrix} \mathbf{u}$$

Por tanto, si ponemos el modelo en forma de ecuaciones:

$$y = \beta_0 + \sum_{i=1}^r \beta_i x_i + \sum_{i=r+1}^k \beta_i x_i + u$$

Y sustituimos las  $x_i$  no observadas por su expresión en función de las  $z_i$ :

$$y = \beta_0 + \sum_{i=1}^r \beta_i x_i + \sum_{i=r+1}^k \beta_i (\delta_{0i} + \delta_{1i} z_i + v_i) + u$$

Y operando y reordenando, obtenemos:

$$y = \beta_0 + \sum_{i=r+1}^k \beta_i \delta_{0i} + \sum_{i=1}^r \beta_i x_i + \sum_{i=r+1}^k \beta_i \delta_{1i} z_i + u + \sum_{i=r+1}^k \beta_i v_i$$

Con lo cual vemos que si sustituimos las variables no observadas por sus variables proxy, obtenemos un modelo con estas características:

- Si se cumplen las dos hipótesis que hemos fijado, cumple las hipótesis de un modelo lineal, ya que  $u + \sum_{i=r+1}^k \beta_i v_i$  están incorrelacionadas con las variables exógenas. Por tanto, los estimadores serán insesgados o al menos consistentes.
- El estimador de los coeficientes correspondientes a las variables observadas no varía.
- Tanto el término independiente como los coeficientes de las variables no observadas no se estiman, pero ya hemos visto que estos en cualquier caso no nos interesan. Obtenemos en su lugar estimaciones del término independiente y los coeficientes para las variables proxy, que pueden o no tener interés para nuestro estudio.

Es importante tener en cuenta que si las variables proxy no cumplen las hipótesis que les hemos exigido, los estimadores que obtengamos serán sesgados.

#### 4.5.1. Uso de variables dependientes retardadas como variables proxy.

Hay veces en las que tenemos claro cual es la variable independiente no observable que afecta a nuestro modelo, y podemos encontrar una variable proxy adecuada. Sin embargo, en otras ocasiones bien no tenemos claro que factor puede afectar al modelo, o no podemos encontrar una variable proxy. En estos casos, se suele utilizar un retardo de la variable dependiente como mecanismo de control de las variables ocultas.

Esto aumenta los requisitos sobre los datos si estamos trabajando sobre datos de corte transversal (necesitamos datos de dos periodos), pero nos permite tener en cuenta factores históricos que pueden tener influencia en el valor actual de la variable dependiente. De hecho, es de suponer los factores no observables que tengan influencia sobre el modelo en la actualidad también la tendrían en periodos pasados.

Este es un método que está bastante extendido en la práctica, y aunque no es perfecto, ayuda a mejorar la estimación en muchos casos.

### 4.6. Errores de especificación.

Entendemos como error de especificación aquel que se comete cuando el modelo que se quiere estimar no coincide con el modelo real que siguen los datos. Estos errores pueden ser de tres tipos: por omisión de variables relevantes, por inclusión de variables irrelevantes y por mala especificación funcional del modelo.

#### 4.6.1. Omisión de variables relevantes.

Supongamos que tenemos un modelo del tipo:

$$\mathbf{y} = [\mathbf{X}_0; \mathbf{X}_1] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \mathbf{u}$$

Pero que, debido a las razones que sean, nosotros estimamos asumiendo que el modelo es:

$$\mathbf{y} = \mathbf{X}_0 \beta_0 + \mathbf{u}$$

El estimador MC que obtendremos será:

$$\hat{\beta} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{y}$$

Si sustituimos las  $\mathbf{y}$  por su expresión según el modelo correcto:

$$\hat{\beta} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' [\mathbf{X}_0; \mathbf{X}_1] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{u}$$

Y como:

$$(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' [\mathbf{X}_0; \mathbf{X}_1] = [(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_0; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_1] = [\mathbf{I}_r; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_1]$$

Como consecuencia:

$$E(\hat{\beta}) = [\mathbf{I}_r; (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_1] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \beta_0 + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X}_1 \beta_1$$

Con lo que si las variables omitidas son significativas el estimador será sesgado. El sesgo se compone del producto de dos factores:

- Una matriz que tiene por columnas los estimadores por mínimos cuadrados de los coeficientes de las regresiones de las variables excluidas sobre las no excluidas.
- El vector de coeficientes que tienen las variables excluidas en el modelo verdadero.

Por tanto, si las variables excluidas son ortogonales a las no excluidas, es decir, si  $\mathbf{X}_0' \mathbf{X}_1 = \mathbf{0}_{r \times (k-r)}$ . Evidentemente, esto es altamente improbable.

Por otro lado, dado que la estimación de la varianza de la perturbación es  $\hat{\sigma}_u^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - k}$ , si definimos  $\mathbf{M}_0 = \mathbf{I}_N - \mathbf{X}_0(\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0'$ , la estimación que calcularemos será:

$$E\left(\frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{N - r}\right) = \sigma^u + \frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{M}_0 \mathbf{X} \boldsymbol{\beta}}{N - r}$$

Y la estimación de la varianza tendrá un sesgo positivo.

#### 4.6.2. Inclusión de variables no relevantes.

En este caso, la matriz de observaciones de las variables exógenas que usaremos para la estimación será  $\mathbf{X}_0 = [\mathbf{X}; \mathbf{X}_1]$ , donde  $\mathbf{X}_1$  es una matriz  $N \times s$  de variables no relevantes para el modelo. Por tanto, el estimador será:

$$\hat{\boldsymbol{\beta}}_0 = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{y}$$

Si sustituimos las  $\mathbf{y}$  por su expresión según el modelo correcto:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{X} \boldsymbol{\beta} + (\mathbf{X}_0' \mathbf{X}_0)^{-1} \mathbf{X}_0' \mathbf{u}$$

Y como:

$$(\mathbf{X}_0' \mathbf{X}_0)^{-1} (\mathbf{X}_0' \mathbf{X}_0) = \mathbf{I}_{k+s} = (\mathbf{X}_0' \mathbf{X}_0)^{-1} (\mathbf{X}_0' \mathbf{X}; \mathbf{X}_0' \mathbf{X}_1)$$

Por tanto,

$$(\mathbf{X}_0' \mathbf{X}_0)^{-1} (\mathbf{X}_0' \mathbf{X}) = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{s \times k} \end{pmatrix}$$

$$E \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k \\ \mathbf{0}_{s \times k} \end{pmatrix} \boldsymbol{\beta}_0 = \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{0}_s \end{pmatrix}$$

Por tanto, el estimador de los coeficientes de las variables relevantes es insesgado, y el de las variables irrelevantes tiene esperanza cero, por lo que cabe esperar que los coeficientes de estas últimas resultasen ser no significativos.

Se puede demostrar que el estimador de la varianza de la perturbación también es insesgado.

Por tanto, podría parecer que la estrategia a seguir debería ser incluir cuantas más variables sea posible en el modelo, dado que esto no introduce sesgo en los estimadores. Sin embargo, el incluir más variables de las necesarias aumenta la varianza del estimador por lo que no solo perdemos precisión, sino que nos puede hacer creer de forma errónea que un coeficiente no es significativo.

#### 4.6.3. Errores de especificación funcional.

Un modelo de regresión lineal adolece de errores de mala especificación cuando la forma funcional que proponemos no coincide con la forma funcional que relaciona las variables en la realidad.

Si, por ejemplo, tenemos un fenómeno cuya relación entre variables es  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + u$ , pero nosotros estimamos el modelo  $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$ , ya hemos visto que al omitir una variable relevante, los estimadores serán sesgados, con un tamaño para el sesgo que depende del valor de  $\beta_3$  y de la correlación entre las variables. Sin embargo, las consecuencias sobre el estimador del efecto de  $x_2$  en el modelo son aún peores, pues aunque obtuviésemos un estimador insesgado para  $\beta_2$  nos estaríamos dejando el efecto de la variable al cuadrado y por tanto el efecto total sería  $\beta_2 + 2\beta_3 x_2$ . Vemos que la diferencia entre el efecto real y el de nuestro modelo será mayor cuanto mayor sea  $x_2$ .

Este mismo problema se introduce cuando en la relación funcional aparece un término de interacción entre dos variables: no solo los coeficientes que obtenemos son sesgados, sino que los efectos de las variables que aparecen en la interacción tendrán una interpretación errónea.



Otro posible mala especificación ocurrirá si usamos la variable en forma lineal en lugar de en forma logarítmica. Entonces los estimadores de los efectos parciales no serán ni insesgados ni consistentes. Veremos más adelante algunos contrastes para contemplar estos casos.

En general, estas malas especificaciones funcionales se pueden intuir incluyendo términos cuadráticos en el modelo y realizando un contraste  $F$  de significación conjunta. Si los términos cuadráticos son significativos, pueden incluirse en el modelo, aunque a veces pueden reflejar otros problemas con la especificación funcional, como el uso de la variable en nivel en vez de en logaritmos o viceversa. Puede ser difícil identificar la causa por la que la forma funcional que se ha seleccionado es incorrecta. En muchos casos, para detectar estas relaciones no lineales es suficiente con usar logaritmos de ciertas variables y añadir términos al cuadrado.

## 4.7. Contraste RESET.

Se han propuesto varios contrastes para detectar la mala especificación funcional. El contraste de RESET (regression specification error test) de Ramsey ha resultado muy útil en este sentido.

El contraste es muy sencillo: si el modelo original,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

Cumple la hipótesis de que la media de la perturbación condicionada a las variables independientes es cero, entonces ninguna función no lineal que añadamos al modelo debería ser significativa.

Si añadiésemos las variables en forma cuadrática, nos encontraríamos con el problema de que si tenemos muchas variables los grados de libertad del modelo disminuyen mucho, además de que no estamos contemplando otras posibles formas funcionales. Para evitar esto, se añaden polinomios de las predicciones del modelo original ajustadas por MCO.

Tenemos que decidir cuántas funciones de los valores ajustados vamos a incluir en el contraste. No hay una respuesta válida para todos los casos, pero con los términos al cuadrado y al cubo suele ser suficiente. Por tanto, la ecuación ampliada quedará:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + u$$

La hipótesis nula es que el modelo original es correcto. Por tanto, contrastamos  $H_0 : \delta_1 = \delta_2 = 0$  en el modelo ampliado. Para eso utilizamos el estadístico  $F$ , que si es significativo nos indicará un posible problema de especificación funcional. Bajo la hipótesis nula,  $F \sim F_{2;n-k-3}$  aproximadamente para muestras grandes.

También podemos utilizar la versión LM del contraste, o utilizar los métodos apropiados para que el contraste sea robusto a la heterocedasticidad.

El problema de este contraste es que no nos dice nada de la especificación funcional correcta si se rechaza la hipótesis nula.

A veces se utiliza RESET como un contraste mucho más general, para detectar variables omitidas e incluso heterocedasticidad. Sin embargo, la potencia de RESET para detectar variables omitidas es muy baja si éstas tienen esperanzas que son lineales en las variables incluidas en el modelo, y es incapaz de detectar la heterocedasticidad si el modelo está bien especificado. Por tanto, solamente debe usarse para detectar mala especificación funcional.

## 4.8. Contraste contra alternativas no anidadas.

Obtener contrastes para otras clases de mala especificación funcional (por ejemplo, decidir si una variable ha de aparecer en nivel o en logaritmo) no lleva a contrastes que se salen del marco habitual del contraste de hipótesis. Así, para contrastar el modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

contra el modelo

$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 \log x_2 + u$$

Nos encontramos con el problema de que los modelos no están anidados y por tanto no podemos usar el contraste estándar de la  $F$ .

Una posibilidad que se ha propuesto es proponer un modelo general que contenga a cada uno de los dos como un caso particular, del tipo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \log x_1 + \beta_4 \log x_2 + u$$

y contrastar las restricciones que nos lleven a esos modelos. Por ejemplo, primero se puede contrastar  $H_0 : \beta_3 = \beta_4 = 0$ , que de cumplirse nos llevará al primer modelo y después contrastar  $H_0 : \beta_1 = \beta_2 = 0$  que nos define el segundo modelo. Este modelo fue propuesto por Mizon y Richard.

Otro método es el propuesto por Davidson y MacKinnon. Estos autores destacan que si el primer modelo es el correcto, entonces el valor ajustado del segundo no debería ser significativo e el primero. Por tanto, se estima el segundo modelo por MCO y obtenemos los valores ajustados para las  $y$ . A estos valores les llamamos  $\hat{y}$ . El **contraste de Davidson-MacKinnon** se basa en el estadístico  $t$  correspondiente a  $\hat{y}$  en el modelo:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \theta_1 \hat{y} + u$$

Un estadístico significativo (contra una alternativa de dos colas) supone un rechazo del primer modelo.

Del mismo modo, si llamamos  $\hat{y}$  a los valores ajustados de la estimación del primer modelo, el contraste para el segundo modelo se basa en el estadístico  $t$  de  $\hat{y}$  en el modelo

$$y = \beta_0 + \beta_1 \log x_1 + \beta_2 \log x_2 + \theta_1 \hat{y} + u$$

Y un  $t$  significativo rechaza el segundo modelo.

Ambos contrastes sirven para contrastar cualquier par de modelos no anidados con la misma variable dependiente.

Hay varios problemas a la hora de utilizar estos contrastes. Para empezar, puede que ninguno de los modelos se rechace. En este caso, podemos usar el  $R^2$  ajustado para decidirnos por uno de ellos. Otra posibilidad es que se rechacen los dos, en cuyo caso habrá que buscar una alternativa. Sin embargo, en la práctica, si los efectos de las variables independientes más importantes sobre la variable dependiente son similares, no tendrá importancia que modelo usar.

Otro problema es que rechazar un modelo no implica que el otro sea correcto, ya que el rechazo puede deberse a diversas formas de mala especificación funcional.

Por último, un problema más difícil se produce cuando la duda en la especificación está en la variable dependiente (por ejemplo, nivel frente a logaritmo). En este caso los contrastes que hemos visto no son válidos.

## Capítulo 5

# Modelo lineal con series de tiempo.

**Variables binarias para efectos temporales y variables en forma de números índice. Uso de variables con tendencia en la regresión. Uso de series débilmente dependientes. Transformación de series altamente persistentes. Tratamiento de la estacionalidad en el modelo.**

### 5.1. Modelo lineal con series de tiempo.

La principal diferencia que presentan los modelos de series temporales respecto a los modelos de corte transversal es que los datos provenientes de un conjunto de series temporales vienen ordenados respecto al tiempo, es decir, los datos tienen un orden natural. Esto es importante, porque los datos pasados pueden afectar a los datos futuros, pero no al revés.

Otra distinción fundamental es que, mientras que los datos de corte transversal se suponen provenientes de una muestra aleatoria tomada de una población mayor, en el caso de series temporales hemos de suponer que nuestros datos provienen de un proceso estocástico o proceso de series temporales, es decir, una sucesión de variables aleatorias cuyo índice es el tiempo. Así, un conjunto de datos de series temporales no es más que una realización del proceso estocástico subyacente. Si fuésemos capaces de retorceder en el tiempo y volver a tomar los datos probablemente la realización de ese proceso sería distinta, es por esto que consideramos a los datos como un conjunto de variables aleatorias. Por tanto, el conjunto de todas las posibles realizaciones de un proceso estocástico equivale a la población total para un análisis transversal, y el tamaño muestral será el número de periodos para los que disponemos de observaciones.

#### 5.1.1. Tipos de modelos con series temporales.

Vamos a ver dos tipos de modelos con series temporales que se pueden estimar fácilmente por mínimos cuadrados ordinarios.

**Modelos estáticos.** En estos modelos, la relación entre las variables exógenas y la variable dependiente es contemporánea, esto es, tenemos datos de series temporales para un conjunto de variables, fechados de forma contemporánea y el modelo que las relaciona es del tipo  $y_t = \beta_0 + \sum_{i=1}^k \beta_i x_{ti} + u_t$ .

Estos tipos de modelos se formulan cuando se entiende que los cambios en las variables exógenas tendrán un efecto inmediato en la variable dependiente, o cuando estamos interesados en conocer la relación de intercambio entre las variables.

**Modelos de retardos distribuidos finitos.** En un modelo de retardos distribuidos finitos, una o más variables exógenas afectan a la variable dependiente con algún retardo. Un ejemplo podría ser:

$$y_t = \beta_0 + \sum_{j=0}^q \delta_j x_{t-j} + u_t$$

Este modelo se utiliza para variables que influyen en la variable dependiente pero no de forma simultánea, sino con algún retardo. El modelo que hemos puesto como ejemplo es un modelo con retardos distribuidos finitos (RDF) de orden  $q$ .

Para interpretar este modelo, supongamos que  $X_0$  es constante igual a una cantidad  $c$ , en el instante  $t$  aumenta hasta  $c + 1$  y en el instante  $t + 1$  vuelve a bajar hasta  $c$ . En ese caso, el valor esperado de  $y$  en cada

instante sería:

$$\begin{aligned}
y_{t-1} &= \beta_0 + \sum_{j=0}^q \delta_j c = y_{t-1} \\
y_t &= \beta_0 + \delta_0(c+1) + \sum_{j=1}^q \delta_j c = y_{t-1} + \delta_0 \\
y_{t+1} &= \beta_0 + \delta_0 c + \delta_1(c+1) + \sum_{j=2}^q \delta_j c = y_{t-1} + \delta_1 \\
y_{t+2} &= \beta_0 + \delta_0 c + \delta_1 c + \delta_2(c+1) + \sum_{j=3}^q \delta_j c = y_{t-1} + \delta_2 \\
&\dots \\
y_{t+q} &= \beta_0 + \sum_{j=0}^{q-1} \delta_j c + \delta_q(c+1) = y_{t-1} + \delta_q \\
y_{t+q+1} &= \beta_0 + \sum_{j=0}^q \delta_j c = y_{t-1}
\end{aligned}$$

Por tanto, podemos ver que  $\delta_0$  es el efecto inmediato que un cambio en  $x_0$  tiene en  $y$ . Normalmente se denomina **propensión al impacto** o **modificador de impacto**. Por otro lado,  $\delta_1$  es el efecto en  $y$  de un cambio en  $x_0$  un período después de que el cambio se produzca,  $\delta_2$  es el efecto dos periodos después del cambio, etc. En el momento  $t+q+1$   $y$  vuelve a su valor inicial, debido a que en nuestro modelo hemos supuesto  $q$  retardos. Si realizamos un gráfico de  $\delta_j$  respecto a  $j$ , obtenemos su distribución de retardos, que muestra el efecto que tiene sobre  $y$  un cambio temporal en  $x_0$ .

Si el cambio en  $x_0$  fuese permanente, es decir, si  $x_0$  pasa de valer  $c$  a valer  $c+1$  para  $t, t+1, \dots$ , el efecto sería el siguiente:

$$\begin{aligned}
y_{t-1} &= \beta_0 + \sum_{j=0}^q \delta_j c = y_{t-1} \\
y_t &= \beta_0 + \delta_0(c+1) + \sum_{j=1}^q \delta_j c = y_{t-1} + \delta_0 \\
y_{t+1} &= \beta_0 + \delta_0(c+1) + \delta_1(c+1) + \sum_{j=2}^q \delta_j c = y_{t-1} + \delta_0 + \delta_1 \\
y_{t+2} &= \beta_0 + \delta_0(c+1) + \delta_1(c+1) + \delta_2(c+1) + \sum_{j=3}^q \delta_j c = y_{t-1} + \delta_0 + \delta_1 + \delta_2 \\
&\dots \\
y_{t+q} &= \beta_0 + \sum_{j=0}^q \delta_j(c+1) = y_{t-1} + \sum_{j=0}^q \delta_j \\
y_{t+q+1} &= \beta_0 + \sum_{j=0}^q \delta_j(c+1) = y_{t-1} + \sum_{j=0}^q \delta_j
\end{aligned}$$

Por tanto, vemos que  $\sum_{j=0}^q \delta_j$  es el cambio a largo plazo que experimenta  $y$  tras un aumento permanente de una unidad en  $x_0$  y se denomina **propensión a largo plazo (PLP)** o **multiplicador a largo plazo** y es a menudo de interés en estos modelos.

Como a menudo existe una correlación elevada entre los retardos de la variable independiente, estos modelos pueden presentar problemas de multicolinealidad, lo que hace que las estimaciones de los  $\delta_i$  individuales sean muy imprecisas. Veremos que aún en este caso, a menudo podemos obtener buenos estimadores de la PLP. Estos modelos pueden tener más de una variable con retardos, variables contemporáneas, etc. Puede ocurrir que el objetivo al estimar el modelo sea contrastar si la variable independiente tiene efecto retardado sobre la variable dependiente.

**5.1.2. Propiedades del estimador MCO para muestras finitas.**

Veremos ahora cómo debemos modificar las hipótesis clásicas del modelo para poderlas aplicar a un modelo de series temporales, y que propiedades se deducen de ellas para muestras finitas.

**Insesgadez de los estimadores.** La primera hipótesis clásica establece que las variables se relacionan según un modelo lineal en los parámetros.

**Hipótesis 1 Linealidad en los parámetros:** El proceso estocástico  $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t), t = 1, 2, \dots, n\}$  sigue el modelo lineal

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t$$

donde  $\{u_t, t = 1, 2, \dots, n\}$  es la sucesión de términos de error o perturbaciones.

De igual modo que hemos hecho para el modelo con datos de corte transversal, podemos expresarlo en forma matricial, es decir,  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ .

**Hipótesis 2 Media condicionada nula:** Para cada instante  $t$  el valor esperado del término de error dadas las variables explicativas en todos los periodos temporales es cero, es decir,

$$E(u_t | \mathbf{X}) = 0$$

o, matricialmente:

$$E(\mathbf{u} | \mathbf{X}) = \mathbf{0}$$

Esta hipótesis es crucial a la hora de deducir las propiedades que hacen al estimador MCO conveniente, y se puede interpretar en términos de incorrelación. Es decir, la hipótesis de esperanza condicionada nula implica que el término de error está incorrelacionado con las variables explicativas en cada uno de los periodos temporales. El hecho de que esté expresado en función de la esperanza condicionada implica que la relación entre la variable independiente y las variables explicativas debe estar completamente especificada en el modelo. Si las  $u_t$  son independientes de las  $\mathbf{X}$  y  $E(u_t) = 0$  el supuesto se cumple automáticamente.

Si lo que se cumple es  $E(u_t | \mathbf{x}_t) = 0$ , decimos que tenemos exogeneidad contemporánea, pero la hipótesis exige exogeneidad estricta.

En el caso de datos de corte transversal, dado que suponemos muestreo aleatorio, la perturbación será automáticamente independiente de las variables explicativas de otras observaciones. En el caso de series temporales esto no se produce, es por esto que debemos exigir explícitamente la exogeneidad estricta.

La exogeneidad estricta puede incumplirse por varias causas: por una mala especificación del modelo que no incluya todas las variables independientes implicadas, por la comisión de errores de medida al medir los regresores, o porque el valor de la variable independiente pueda influir en valores futuros de alguna variable exógena. Para que una variable explicativa sea estrictamente exógena, no puede verse influida por valores pasados de la variable endógena. Por desgracia, esta circunstancia se da en muchos casos, y muchas variables explicativas violan la hipótesis de exogeneidad estricta.

Muchos análisis de modelos estáticos y de retardos distribuidos finitos suponen esta hipótesis mediante la hipótesis más fuerte de que las variables explicativas son deterministas, es decir, sin componente aleatorio, pero esto es obviamente falso si tenemos observaciones de series temporales.

**Hipótesis 3 No multicolinealidad perfecta:** En nuestro modelo ninguna variable independiente es constante o combinación lineal exacta de las demás.

En el caso de que no se cumpla esta hipótesis, la matriz  $\mathbf{X}$  será singular, y por tanto tendremos infinitos estimadores para los coeficientes del modelo.

**Teorema 6** Bajo las anteriores hipótesis (linealidad, media condicionada nula y ausencia de multicolinealidad) los estimadores de mínimos cuadrados de los coeficientes del modelo son estimadores insesgados de los mismos.

Si se cumplen las hipótesis de linealidad y no multicolinealidad, el estimador MCO existe y su fórmula es  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ . Para comprobar si es insesgado, calculemos su esperanza. Para ello, primero calculamos la esperanza condicionada a  $\mathbf{X}$ :

Entonces:

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

Ya que  $\beta$  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, siempre que se cumpla la hipótesis de esperanza condicionada nula.

Para completar las hipótesis de Gauss-Markov para el caso de series temporales, necesitamos dos hipótesis adicionales.

**Hipótesis 4 Homoscedasticidad:** La varianza de la perturbación condicionada a  $\mathbf{X}$  es constante para cualquier valor de  $t$ :  $Var(u_t|\mathbf{X}) = Var(u_t) = \sigma^2$ ,  $t = 1, 2, \dots, n$ .

Por tanto, la varianza de la perturbación no puede depender de las variables exógenas (para lo que es suficiente que  $u_t$  y  $\mathbf{X}$  sean independientes) y esta varianza debe ser constante en el tiempo. Si esta hipótesis no se cumple, diremos que el modelo presenta *heteroscedasticidad*, o bien que los errores son *heteroscedásticos*. Para que esta hipótesis se cumpla, las variables no observadas que afecten a la variable dependiente deben tener una variabilidad constante, y que la varianza de la variable endógena no dependa del valor de ninguna variable exógena.

**Hipótesis 5 Ausencia de autocorrelación:** Condicionando a las variables exógenas, los errores en dos periodos de tiempo están incorrelacionados:  $Corr(u_t, u_s|\mathbf{X}) = 0$  para todo  $t \neq s$ .

Si esta hipótesis no se cumple, diremos que el término de error presenta **autocorrelación**.

Si un modelo presenta homoscedasticidad y ausencia de autocorrelación, la matriz de varianzas-covarianzas del término de error será:

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}_n$$

Estas cinco hipótesis son las hipótesis de Gauss-Markov aplicadas al modelo con datos de series temporales.

**Teorema 7 Varianza del estimador MCO:** Bajo las hipótesis de Gauss-Markov (linealidad, media condicionada nula, ausencia de multicolinealidad, homoscedasticidad y ausencia de autocorrelación) la matriz de varianzas-covarianzas de los estimadores condicionada a  $\mathbf{X}$  es:

$$Var(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

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

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 un 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.

**Teorema 8 Estimador insesgado de  $\sigma^2$ :** Bajo las hipótesis de Gauss-Markov (linealidad, media condicionada nula, ausencia de multicolinealidad, homoscedasticidad y ausencia de autocorrelación) el estimador

$$Var(\hat{\sigma}^2) = \frac{\mathbf{u}'\mathbf{u}}{n - k}$$

es un estimador insesgado de  $\sigma^2$ .

**Teorema 9 Teorema de Gauss-Markov:** *Bajo las hipótesis de Gauss-Markov (linealidad, media condicionada nula, ausencia de multicolinealidad, homocedasticidad y ausencia de autocorrelación) los estimadores de mínimos cuadrados son los estimadores lineales insesgados de mínima varianza, condicionados a  $\mathbf{X}$ .*

## 5.2. Variables binarias para efectos temporales y variables en forma de números índice.

Sección 10.4,

## 5.3. Uso de variables con tendencia en la regresión.

Sección 10.5

## 5.4. Uso de series débilmente dependientes.

Sección 11.1

## 5.5. Transformación de series altamente persistentes.

Sección 11.3

## 5.6. Tratamiento de la estacionalidad en el modelo.

Sección 10.5

## Capítulo 6

### Modelos dinámicos.

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

#### 6.1. [Introducción.](#)

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

#### 6.2. [Justificación teórica de los modelos econométricos dinámicos.](#)

#### 6.3. [Modelos de retardos infinitos.](#)

#### 6.4. [Estimación con retardos de la variable endógena.](#)

#### 6.5. [Contraste de exogeneidad de Asuman.](#)

#### 6.6. [Eficiencia relativa de los estimadores de variables instrumentales.](#)

#### 6.7. [Estimación de modelos con expectativas racionales.](#)



## Capítulo 7

# Multicolinealidad y modelos no lineales.

Concepto y consecuencias. Detección de la multicolinealidad. Remedios contra la multicolinealidad. Observaciones influyentes. Especificación de modelos no lineales. Aproximación lineal al modelo no lineal. Mínimos Cuadrados no lineales. Estimación de Máxima Verosimilitud.

- 7.1. [Introducción.](#)
- 7.2. [Concepto y consecuencias.](#)
- 7.3. [Detección de la multicolinealidad.](#)
- 7.4. [Remedios contra la multicolinealidad.](#)
- 7.5. [Observaciones influyentes.](#)
- 7.6. [Especificación de modelos no lineales.](#)
- 7.7. [Aproximación lineal al modelo no lineal.](#)
- 7.8. [Mínimos Cuadrados no lineales.](#)
- 7.9. [Estimación de Máxima Verosimilitud.](#)

## Capítulo 8

### Datos de panel.

Descripción del problema. El modelo de efectos aleatorios. Estimación. Contraste de especificación. Modelos dinámicos. Identificación de efectos individuales en el estimador intragrupo.

8.1. [Introducción.](#)

8.2. [Descripción del problema.](#)

8.3. [El modelo de efectos aleatorios.](#)

8.4. [Estimación.](#)

8.5. [Contraste de especificación.](#)

8.6. [Modelos dinámicos.](#)

8.7. [Identificación de efectos individuales en el estimador intragrupo.](#)