NOVA

# ECONOMETRÍA EMPRESA

## Modelo de Regresión Lineal Múltiple

## GRADO DE ADE (UB)

NOVA

www.academianovaonline.com

# ÍNDICE

# TEMA 0: INTRODUCCIÓN A LA ECONOMETRÍA

Todo estudio econométrico se centra en dos pilares básicos: la teoría y los hechos. La teoría permite derivar un modelo (el modelo económico) que sintetiza la incógnita relevante sobre el fenómeno (la variable endógena) objeto del análisis y del cual deriva el modelo econométrico que permite medirlo y contrastarlo empíricamente.

Los hechos se concretan en una serie de datos que denominaremos información muestral. La muestra, a su vez, consiste en una lista ordenada de valores numéricos de las variables objeto de estudio. En una muestra de corte transversal, diversos agentes económicos de una naturaleza similar proporcionan información solicitada en un mismo instante de tiempo. Alternativamente, el investigador económico trabaja en ocasiones con datos de series temporales, en las que se dispone de información acerca de unidad económica, como puede ser un país, una empresa, a lo largo de tiempo; estas muestras pueden tener frecuencia diaria, mensual, anual, según frecuencia de observación de los datos.

Una vez que se especifica el modelo y se dispone de la información estadística convenientemente tratada, se llega a la etapa siguiente del trabajo econométrico: la etapa de estimación. Los resultados de esta etapa de estimación permiten medir y contrastar las relaciones sugeridas por la teoría económica.

## OBJETIVOS

• Conocer la estructura del MRLM.

• Familiarizarse con las hipótesis básicas del MRLM y entender su importancia.

• Conocer los métodos de estimación del MRLM, el método de mínimos cuadrados ordinarios (MCO).

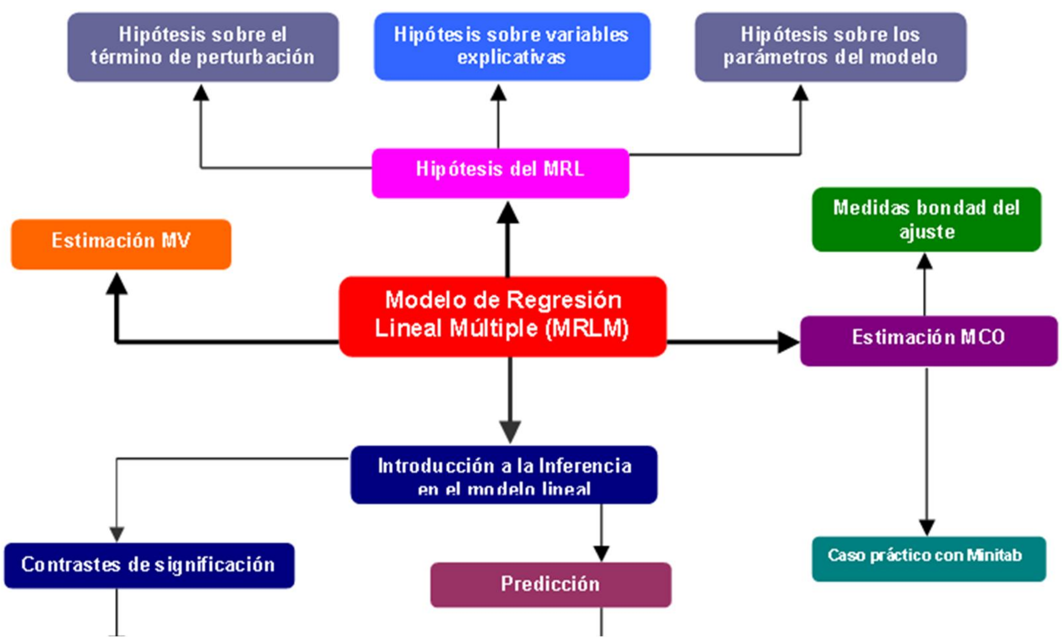
• Introducir es uso de un software econométrico para el análisis.

• Saber cuantificar e interpretar bondad del ajuste del modelo.

• Evaluar la contribución de cada variable exógena en explicar el comportamiento de la variable endógena; contrastar la significación individual de un parámetro y la global del modelo.

• En base de la estimación de MRLM, realizar predicciones puntuales y por intervalo de la variable endógena.

## ESQUEMA DE CONTENIDO



# TEMA 1. MODELO DE REGRESIÓN LINEAL MÚLTIPLE (MRLM)

Mediante un modelo de regresión lineal múltiple (MRLM) tratamos de explicar el comportamiento de una determinada variable que denominaremos variable a explicar, variable endógena o variable dependiente, (y representaremos con la letra Y) en función de un conjunto de k-1 variables explicativas $X_2$, $X_3$ ,..., $X_k$ mediante una relación de dependencia lineal (suponiendo $X_1$= 1):

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_K x_{ki} + u_i \qquad \text{para i = 1, 2, 3,…, n}$$

y donde $u_i$ es el término de perturbación del modelo

Para determinar el modelo anterior, es necesario hallar (estimar) el valor de los coeficientes $\beta_1$, $\beta_2$,..., $\beta_k$. La linealidad en parámetros posibilita la interpretación correcta de los parámetros del modelo. Los parámetros miden la intensidad media de los efectos de las variables explicativas sobre la variable a explicar y se obtienen al tomar las derivadas parciales de la variable a explicar respecto a cada una de as variables explicativas.

Nuestro objetivo es asignar valores numéricos a los parámetros. Es decir, $\beta_1$, $\beta_2$,..., $\beta_k$ trataremos de estimar el modelo de manera que, los valores ajustados de la variable endógena resulten tan próximos a los valores realmente observados como sea posible. A fin de poder determinar las propiedades de los estimadores obtenidos al aplicar distintos métodos de estimación y realizar diferentes contrastes, hemos de especificar un conjunto de hipótesis sobre el MRLM que hemos formulado. Existen tres grupos de hipótesis siguientes:

- las hipótesis sobre el término de perturbación,
- las hipótesis sobre las variables explicativas, y
- las hipótesis sobre los parámetros del modelo

.
Para una muestra de n observaciones (cada observación estará formada por una tupla con los valores de $X_2$, $X_3$,..., $X_k$ y el valor de Y asociado), tendremos el siguiente sistema de n ecuaciones lineales:

$$\begin{aligned}
&i = 1 \rightarrow y_1 = \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \ldots + \beta_K x_{K1} + u_1 \\
&i = 2 \rightarrow y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \ldots + \beta_K x_{K2} + u_2 \\
&i = 3 \rightarrow y_3 = \beta_1 + \beta_2 x_{23} + \beta_3 x_{33} + \ldots + \beta_K x_{K3} + u_3 \\
&\ldots \\
&i = n \rightarrow y_n = \beta_1 + \beta_2 x_{2n} + \beta_3 x_{3n} + \ldots + \beta_K x_{Kn} + u_n
\end{aligned}$$

En notación matricial:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} 1 & x_{21} & x_{31} & \ldots & x_{K1} \\ 1 & x_{22} & x_{32} & \ldots & x_{K2} \\ 1 & x_{23} & x_{33} & \ldots & x_{K3} \\ \ldots & & & & \\ 1 & x_{2n} & x_{3n} & \ldots & x_{Kn} \end{pmatrix} \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \ldots \\ \beta_K \end{pmatrix} \qquad U = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \ldots \\ u_n \end{pmatrix}$$

Así pues:

$Y = X\beta + U$ que es la forma matricial o condensada del modelo.

## 1.1 HIPÓTESIS/PREMISAS DEL MODELO DE REGRESIÓN LINEAL MÚLTIPLE

- **Hipótesis de linealidad**: debe existir una relación lineal entre la variable endógena Y, los regresores X y los parámetros β. Si no tiene linealidad se dice que tenemos un error de especificación.

- **Hipótesis de permanencia estructural o estabilidad estructural**: la única hipótesis que haremos acerca de los parámetros del modelo es la hipótesis de permanencia estructural, lo cual quiere decir que los parámetros poblacionales, $\beta_j$, se mantienen constantes a lo largo de toda la muestra.

- **Hipótesis de término de perturbación**: En estas condiciones, las hipótesis del MRLM se resumen en la esfericidad del término de perturbación:

- El término de perturbación es una variable aleatoria.
- El término de perturbación se distribuye según Ley Normal. ***Hipótesis de normalidad***.
- La esperanza del término de perturbación es nula:
  $E[u_i] = 0, \forall i$
- La varianza del término de perturbación es constante a lo largo del tiempo. Es la ***hipótesis de homoscedasticidad***.
  $E[u_i^2] = Var(u_i) = \sigma_u^2, \forall t$
- **Hipótesis de No autocorrelación**. $\operatorname{cov}(u_i u_j) = E(u_i u_j) = 0, \forall\ i \neq j$

Entonces:

$$Var(U) = E(UU') = \begin{pmatrix} \sigma_u^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & .. & 0 \\ \dots & & \sigma_u^2 & & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \sigma_u^2 \end{pmatrix}_{n\times n} = \sigma_u^2 I_n$$

RESUMEN:

$u_i \sim \mathrm{N}(0, \sigma_u^2)\ \forall\ i$ $\qquad U \sim \mathrm{N}(0, \sigma_u^2 I_n)\ \forall i$ **Término de Perturbación Esférico**

- **Hipótesis de regresores no estocásticos** (hipótesis sobre las variables explicativas)
  - Las variables explicativas son fijas o deterministas.
  - La variables explicativas están no correlacionadas con la perturbación aleatoria.
    $\mathrm{E}(\mathrm{X}'U) = 0$ regresores estrictamente exógenos.
  - En el modelo no se excluyen las variables relevantes y tampoco no se incluyen las variables irrelevantes, a la hora de explicar el comportamiento de la variable endógena.

- **Hipótesis de Ausencia total de Multicolinealidad**: (hipótesis sobre las variables explicativas): las variables explicativas deben ser linealmente independientes
  - Las variables explicativas no presentan ninguna relación lineal entre si. Son incorreladas.

- **Hipótesis de grados de libertad del modelo positivos**: el requisito mínimo es que para obtener solución, el número de observaciones N tiene que ser igual al número de parámetros K.

$gl_{modelo}$= N-K $\qquad \mathrm{N} - \mathrm{K} \geq 0 \qquad \mathrm{N} \geq K$

## 1.2 ESTIMACIÓN DE LOS PARÁMETROS DEL MODELO

En el modelo de regresión especificado existe un conjunto de parámetros desconocido ($\beta_j$ y $\sigma_u^2$ ). Por ello, en primer lugar, se tratará de su estimación.

Existen diversos métodos para estimar los parámetros del modelo, muchos de los cuales se basan en los residuos o errores, que se definen como la diferencia entre el valor real de variable dependiente y el estimado por el modelo para dicha variable.

$$e_i = y_i - \hat{y}_i \qquad \text{i=1,2,...,n}$$

Entre los métodos que estiman los parámetros del modelo a partir de los residuos, el más sencillo es el método de Mínimos Cuadrados Ordinarios (MCO), que hace mínima la suma de los cuadrados de los residuos.

Partiendo de $Minimizar \sum_{i=1}^{n} e_i^2$

El criterio de ajuste impone: $\sum e_i^2 = Mínimo$ . ¿Cómo traduzco esta expresión al lenguaje matricial

$$e = \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \dots \\ e_n \end{pmatrix} \rightarrow Vector\ residuos\ ajuste$$

con lo que:

$$e'e = (e_1\ e_2\ e_3\ \dots\ e_n) \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \dots \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + e_3^2 + \dots + e_n^2 = \sum e_i^2$$

y como queremos que: $\sum e_t^2 = Mínimo$ derivamos e igualamos a cero.

$$\frac{\partial (e'e)}{\partial \beta} = 0$$

y depejamos: $\hat{\beta} = (X'X)^{-1} X'Y$

### 1.2.1 Estimación de los $\hat{\beta}_j$

Se obtiene un sistema de ecuaciones (ecuaciones normales) que permite obtener los estimadores mínimo cuadrático ordinarios de los parámetros $\beta_j$ a partir de la expresión:

$$\hat{\beta} = (X'X)^{-1} X'Y = \begin{pmatrix} b_1 \\ b_2 \\ . \\ b_k \end{pmatrix} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ . \\ \hat{\beta}_k \end{pmatrix}$$

Si en el modelo existiera término independiente, estas matrices se simplificarían con las siguientes expresiones:

$$X'X = \begin{pmatrix} n & \sum_{i=1}^{n} X_{i2} & \dots & \sum_{i=1}^{n} X_{ik} \\ \sum_{i=1}^{n} X_{i2} & \sum_{i=1}^{n} X_{i2}^2 & \dots & \sum_{i=1}^{n} X_{i2}X_{ik} \\ \dots\dots & \dots\dots & \dots & \dots\dots \\ \sum_{i=1}^{n} X_{ik} & \sum_{i=1}^{n} X_{ik}X_{i2} & \dots & \sum_{i=1}^{n} X_{ik}^2 \end{pmatrix} \qquad X`Y = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_{i2}Y_i \\ \dots\dots \\ \sum_{i=1}^{n} X_{ik}Y_i \end{pmatrix}$$

Estos estimadores MCO son estimadores lineales, insesgados y óptimos (ELIO) en el modelo de regresión lineal, normal, clásico.

Cada uno de los coeficientes $\hat{\beta}_j$ representa el efecto de la variable independiente sobre la variable explicada; es decir el valor estimado de $\beta_j$ indica la variación que experimenta la variable dependiente cuando la variable independiente $X_j$ varía en una unidad y todas las demás permanecen constantes.

Propiedades de los estimadores $\hat{\beta}_j$:

- Lineales
- Insesgados
- Óptimos o eficientes
- Consistentes

**Matriz de varianzas y covarianzas de los $\hat{\beta}_j$**

Es una matriz simétrica formada por la varianza de los estimadores y las covarianzas.

$$Var(\hat{\beta}) = \sigma^2 (X'X)^{-1} = \begin{pmatrix} \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1\hat{\beta}_2) & \text{cov}(\hat{\beta}_1\hat{\beta}_3) & \dots & \text{cov}(\hat{\beta}_1\hat{\beta}_k) \\ & \text{var}(\hat{\beta}_2) & \text{cov}(\hat{\beta}_2\hat{\beta}_3) & \dots & \text{cov}(\hat{\beta}_2\hat{\beta}_k) \\ & & \text{var}(\hat{\beta}_3) & \dots & \text{cov}(\hat{\beta}_3\hat{\beta}_k) \\ & & \dots & & \dots \\ & & & \dots & \dots \\ & & & & \text{var}(\hat{\beta}_K) \end{pmatrix}_{k\times k}$$

## 1.2.2 Estimación de la varianza del término de perturbación $\hat{\sigma}_u^2$

El estimador de la varianza de la perturbación no se deduce del sistema de ecuaciones normales; se calcula a partir de la fórmula:

$$\hat{\sigma}_u^2 = S_u^2 = \frac{\sum e_i^2}{n-k} = \frac{\sum (y_i - \hat{y}_i)^2}{n-k} \frac{\overline{VE}}{n-k} = \frac{SC_R}{n-k} = \frac{e'e}{n-k}$$

y se puede comprobar que es el estimador insesgado $E(\hat{\sigma}_u^2) = E(S_u^2) = \sigma_u^2$ de la varianza de la perturbación. $\dfrac{e'e}{\sigma_u^2} \sim \chi^2_{\text{n-k}}$

Dado que desconocemos el valor de $\sigma^2$, sustituimos ésta por si estimación:

$V\hat{a}r(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$ Así pues:

$$V\hat{a}r(\hat{\beta}) = \begin{pmatrix} \hat{\text{var}}(\hat{\beta}_1) & \hat{\text{cov}}(\hat{\beta}_1\hat{\beta}_2) & \hat{\text{cov}}(\hat{\beta}_1\hat{\beta}_3) & \dots & \hat{\text{cov}}(\hat{\beta}_1\hat{\beta}_k) \\ & \hat{\text{var}}(\hat{\beta}_2) & \hat{\text{cov}}(\hat{\beta}_2\hat{\beta}_3) & \dots & \hat{\text{cov}}(\hat{\beta}_2\hat{\beta}_k) \\ & & \hat{\text{var}}(\hat{\beta}_3) & \dots & \hat{\text{cov}}(\hat{\beta}_3\hat{\beta}_k) \\ & & \dots & & \dots \\ & & & \dots & \dots \\ & & & & \hat{\text{var}}(\hat{\beta}_K) \end{pmatrix}$$

Se llama error estándar de un estimador a su desviación estándar estimada:

$$er.es(\hat{\beta}_j) = es(\hat{\beta}_j) = S_{\beta_j} = \sqrt{V\hat{a}r(\hat{\beta}_j)}$$

### 1.2.3 Inferencia de los parámetros del modelo

El método de estimación expuesto permite obtener estimaciones puntuales de los parámetros del modelo. La inferencia permite completar esta estimación puntual, mediante la estimación por intervalos y los contrastes de hipótesis.

Los primeros posibilitan la obtención de un intervalo dentro del cual, con un determinado nivel de confianza, oscilará el verdadero valor de un parámetro.

- Intervalo de confianza para el parámetro $\beta_j$

  Su cálculo se realiza mediante: $\beta_j \in \left(\hat{\beta}_j \pm t_{n-k} \cdot S_{\beta_j}\right)$

$$p\left(\hat{\beta}_j - t_{n-k} \cdot S_{\beta_j} \le \beta_j \le \hat{\beta}_j + t_{n-k} \cdot S_{\beta_j}\right) = 1 - \alpha$$

- Contrastes de hipótesis de $\beta_j$

  - Contraste de significación global o conjunta del modelo.
  - Contraste de significación individual.
  - Contraste individual.
  - Contraste de restricciones lineales

## 1.3 ANÁLISIS DE LA VARIANZA

El modelo de regresión se plantea para explicar el comportamiento de la variable dependiente (Y). En dicho estudio será interesante analizar la variación que experimenta esta variable y, dentro de esta variación, estudiar qué parte está siendo explicada por el modelo de regresión y qué parte es debida a los errores o residuos.

**Análisis de la varianza**

Una forma complementaria de presentar algunos de los contrastes de hipótesis anteriores consiste en realizar un análisis de la varianza de Y, tal como se recoge en el cuadro siguiente:

| Fuente de Variación | Suma de Cuadrados | Grados libertad | SC/gl | F |
|---|---|---|---|---|
| Explicada | $SC_{EXPL} = \sum(\hat{y}_i - \bar{y})^2$ | *k-1* | $SC_{EXPL}/k-1$ | $\hat{F} = \dfrac{SC_{EXPL}/k-1}{SC_{RESID}/n-k}$ |
| No Explicada Residuos | $SC_{RESID} = \sum(y_i - \hat{y}_i)^2 = \sum e_i^2$ | *n-k* | $SC_{RESID}/n-k = \sigma_u^2$ | |
| Total | $SC_{TOT} = \sum(y_i - \bar{y}_i)^2$ | *n-1* | | |

## 1.4 MEDIDAS DE BONDAD DE AJUSTE

Las estimaciones por MCO que hemos realizado todavía no nos permiten evaluar la calidad de ajuste del modelo. Para ello, de aquí a delante iremos viendo las medidas de bondad de ajuste.

Comenzaremos por la suma de los cuadrados de errores, SCRESID, que puede expresarse de varias formas:

$$SC_{RESID} = \sum (y_i - \hat{y}_i)^2 = \sum e_i^2 = e'e$$
$$e'e = Y'Y - \hat{\beta}X'Y = Y'Y - \hat{Y}'\hat{Y} = \sum Y_i^2 - \sum \hat{Y}_i^2$$

Despejando la suma de cuadrados de la variable endógena, queda:

$$Y'Y = \hat{Y}'\hat{Y} + e'e \qquad \text{o bien} \qquad \sum Y_i^2 = \sum \hat{Y}_i^2 + \sum e_i^2$$

Restando a ambos lados la cantidad $n\overline{Y}^2$, obtenemos:

$$\sum Y_i^2 - n\overline{Y}^2 = \sum \hat{Y}_i^2 - n\overline{Y}^2 + \sum e'e \qquad \text{o bien} \qquad \underbrace{\sum \left(Y_i - \overline{Y}\right)^2}_{VT} = \underbrace{\sum \left(\hat{Y}_i - \overline{Y}\right)^2}_{VE} + \underbrace{\sum e_i^2}_{VE}$$

La parte izquierda representa suma de cuadrados totales ($SC_T$) y no es sino la suma de cuadrados de las desviaciones respecto a su media aritmética.

$$SC_{TOTALES} = VT = \sum \left(Y_i - \overline{Y}\right)^2$$

Por otra parte, si el modelo tiene término independiente, a la cantidad se le denomina suma de cuadrados de la regresión o explicados ($SC_E$) .

$$SC_{EXPLIC} = VE = \sum \hat{Y}_i^2 - n\overline{Y}^2 = \sum \left(\hat{Y}_i - \overline{Y}\right)^2$$

En resumen, la variabilidad total de la variable endógena ($SC_T$) puede descomponerse en dos partes: la parte que podemos explicar mediante el modelo especificado ($SC_E$) y la parte que no podemos explicar, la suma de cuadrados de los errores ($SC_R$).

### 1.4.1 Coeficiente de determinación o coeficiente de bondad de ajuste

Una vez estimado el modelo es conveniente obtener una medida acerca de la bondad del ajuste realizado. Un estadístico que facilita esta medida es el coeficiente de determinación ($R^2$), que se define:

$$R^2 = \left(r_{y\hat{y}}\right)^2 = 1 - \frac{\overline{VE}}{VT}$$

Si el modelo tiene término independiente, entonces se cumple la igualdad $SC_T = SC_E + SC_R$, y el coeficiente de determinación podrá expresarse como:

$$R^2 = \frac{VE}{VT}$$

El coeficiente de determinación indica que proporción de variabilidad total queda explicadapor la regresión. Si el modelo tiene término independiente, entonces $R^2$ toma valores entre 0 y 1. Este coeficiente permite, además, seleccionar entre modelos clásicos que tengan el mismo número de regresores, ya que la capacidad explicativa de un modelo es mayor cuanto más elevado sea el valor que tome este coeficiente.

Por otra parte el valor coeficiente de determinación crece con el número de regresores del modelo.

Por ello, si los modelos que se comparan tienen distinto número de regresores, no puede establecerse comparación entre sus $R^2$.

### 1.4.2 Coeficiente de determinación o coeficiente de bondad de ajuste

En práctica, el uso de $R^2$ presenta algunas limitaciones a la hora de comparar varios modelos desde la perspectiva de bondad del ajuste. En efecto, cuantas más variables explicativas incorporamos al modelo, mayor será el coeficiente de determinación, pues la $SC_E$ disminuye conforme aumenta el número de variables explicativas.

Como el coeficiente de determinación crece con el número de regresores del modelo, necesitamos una medida de bondad de ajuste que tenga en cuenta el ajuste en función del número de variables.

Por tanto, cuando queremos llevar a cabo un análisis comparativo entre varios modelos restringidos, utilizamos $R^2$ corregido:

Para comparar modelos restringidos con distinto número de variables explicativas debe emplearse el coeficiente de determinación corregido $\overline{R}^2$, que depura el incremento que experimenta el coeficiente de determinación cuando el número de regresores es mayor.

$$R^2_{corr} = R^2_{adj} = \overline{R}^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right) < R^2$$

# TEMA 2: CONTRASTRES DEL MRLM

Los contrastes de hipótesis pueden clasificarse en distintas categorías según la naturaleza de la hipótesis planteada.

- Contrates paramétricos de los parámetros $\beta_j$
- Contrates para verificar las especificaciones del modelo.

## 2.1 CONTRASTE DE SIGNIFICACIÓN GLOBAL

- Formulación de la hipótesis

$H_0 : \beta_2 = \beta_3 = ... = \beta_K = 0$ MODELO GLOBALMENTE NO SIGNIFICATIVO

$H_A : NoH_0$ MODELO GLOBALMENTE SIGNIFICATIVO

- Estadístico de prueba: $\hat{F}_{\exp} = \dfrac{VE/(k-1)}{\overline{VE}/(n-k)} = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)}$

- Estadístico teórico $F_{k-1,n-k,\alpha}$ valor en tablas

- Regla de decisión: si $\hat{F}_{\exp} > F$ Se rechaza la hipótesis nula y el modelo es globalmente significativo.

El estadístico se distribuye bajo la hipótesis nula con una distribución F de Snedecor con k-1 grado de libertad en el numerador y n-k grados de libertad en el denominador. La regla de decisión utilizada para contrastar la significación global del modelo es la siguiente:

• Si $\hat{F}_{\exp} \geq F_{k-1,n-k,\alpha}$, el estadístico de contraste cae fuera de la región de aceptación, con lo que rechazamos la hipótesis nula. Por tanto, el modelo es globalmente significativo.
• Si $\hat{F}_{\exp} < F_{k-1,n-k,\alpha}$, el estadístico de contraste cae dentro de la región de aceptación, de modo que ahora la hipótesis nula no la rechazamos. En consecuencia, podemos afirmar que el modelo no es globalmente significativo.

## 2.2 CONTRASTE DE SIGNIFICACIÓN INDIVIDUAL

- Formulación de la hipótesis $H_0: \beta_j = 0$, $H_A: \beta_j \neq 0$
- Estadístico de prueba: $\hat{t}_{\exp} = \dfrac{\hat{\beta}_j}{S_{\beta_j}}$
- Estadístico teórico $t = t_{n-k(\alpha/2)}$
- Regla de decisión: si $\left|\hat{t}_{\exp}\right| > t$ Se rechaza la hipótesis nula y el parámetro es significativo

## 2.2 CONTRASTE INDIVIDUAL

- Formulación de la hipótesis $H_0: \beta_j = \beta_j^*$, $H_A: \beta_j \neq \beta_j^*$
- Estadístico de prueba: $\hat{t}_{\exp} = \dfrac{\hat{\beta}_j - \beta_j^*}{S_{\beta_j}}$
- Estadístico teórico $t = t_{n-k(\alpha/2)}$
- Regla de decisión: si $\left|\hat{t}_{\exp}\right| > t$ Se rechaza la hipótesis nula

## 2.3 CONTRASTES DE RESTRICCIONES LINEALES

En las restricciones lineales no aparece el término independiente. Son restricciones del tipo:

$$a\beta_1 + b\beta_2 + c\beta_3 + \ldots + d\beta_k = A$$

Es decir, es una combinación lineal de parámetros.

Ejemplo práctico: Contrastar $\beta_1 + \beta_2 = 1$

$$\left.\begin{array}{l} H_0 : \beta_1 + \beta_2 = 1 \\ H_a : \beta_1 + \beta_2 \neq 1 \end{array}\right\}$$

Se realiza el test de la t. Se requiere, para ello, la **matriz de varianzas y covarianzas**.

En la hipótesis nula pueden aparecer varias restricciones lineales.

$$H_0 : \left\{\begin{array}{l} \beta_2 = \beta_3 \\ \beta_2 + 5\beta_4 = 3 \\ \beta_5 = 2 \end{array}\right\}$$

$$H_a : NoH_0$$

Se realiza el test de la F. Se requiere, para ello, la **suma de cuadrados residuales**.

- Formulación de la hipótesis $H_0 : R\beta = r$ Se cumplen las restricciones
  $H_A : NoH_0$
- Estadístico de prueba: $$\hat{F}_{\exp} = \frac{\left(R\hat{\beta} - r\right)' \cdot \left[R\left(X'X\right)^{-1} R'\right]^{-1} \cdot \left(R\hat{\beta} - r\right) / q}{\overline{VE} / n-k}$$

$$\hat{F}_{\exp} = \frac{\left(\overline{VE}_R - \overline{VE}\right) / q}{\overline{VE} / n-k} = \frac{\left(e'_R e_R - e'e\right) / q}{e'e / n-k} \qquad \hat{F}_{\exp} = \frac{\dfrac{R^2 - R_R^2}{q}}{\dfrac{1-R^2}{n-k}}$$

- Estadístico teórico $F = F_{(q,\, n-k, \alpha)}$
- Regla de decisión: si $\hat{F}_{\exp} > F$ Se rechaza la hipótesis nula y no se cumplen las restricciones

Los estimadores MCR (Mínimos Cuadrados restringidos) son insesgados sólo si se cumplen las restricciones lineales.

Tanto si se cumplen las restricciones como si no se cumplen:

$$Var\left(\hat{\beta}_{MCR}\right) \leq Var\left(\hat{\beta}_{MCO}\right)$$

## 2.4 CONTRASTE DE NORMALIDAD

- Formulación de la hipótesis $H_0 : u_i \sim \text{Normal}$
  $H_A : NoH_0$
- Estadístico de prueba: $BJ = N\left(\frac{\hat{a}^2}{6} + \frac{(\hat{c}-3)^2}{24}\right)$
- Estadístico teórico $\chi^2 = \chi^2_{(2,\alpha)}$
- Regla de decisión: si $BJ > \chi^2$ Se rechaza la hipótesis nula

## 2.5 CONTRASTE DE LINEALIDAD

El análisis econométrico se basa en contrastes lineales. Es conveniente un contraste sobre la forma funcional de la ecuación del modelo para verificar que la especificación lineal es adecuada. Se emplea el contraste RESET.

Diremos que se comete un error en la forma funcional cuando se especifica una relación (que puede ser lineal, cuadrática, cúbica, exponencial, logarítmica, etc.) y la verdadera relación es diferente de la especificada.

Una especificación incorrecta en la forma funcional del modelo puede considerarse, en algunos casos, como la omisión de variables relevantes. Entonces, las consecuencias son las mismas que las que provoca la omisión de variables relevantes, es decir, los estimadores serán sesgados e inconsistentes (ver ejemplo 1 de la parte práctica con software).

En general, un error en la forma funcional nos puede llevar a obtener término de perturbación no esférico (i.e., con heteroscedasticidad y/o autocorrelación), así como al hecho de que la distribución se aleje de la distribución del término de perturbación del modelo correctamente especificado. En consecuencia, es importante disponer de algún método para detectar un posible error en la especificación de la forma funcional. Uno de los contrastes más utilizados
es el contraste RESET.

Supongamos el modelo:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t$$

Se quiere contrastar:

$$H_0 : Linealidad$$
$$H_a : No\ linealidad$$

Para ello se estima el modelo anterior por MCO. A continuación, se plantea un nuevo modelo, llamado “regresión auxiliar”:

$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \gamma \hat{y}_t^2 + u_t$ que se estima por MCO.

Y se contrasta:

$$H_0 : \gamma = 0 \rightarrow Linealidad$$
$$H_a : \gamma \neq 0 \rightarrow No\ Linealidad$$

Se realiza un contraste de significación individual de $\gamma$ en este último modelo para verificar la relevancia de $\hat{y}_t^2$. Esto es:

$$\hat{t}_{\exp} = \frac{\hat{\gamma}}{er.es(\hat{\gamma})} \sim t_{n-k}$$

## 2.6 CONTRASTE DE PERMAMNENCIA ESTRUCTURAL O TEST DE CHOW

Sea el modelo Y = Xβ + U, que se extiende a los periodos t = 1, 2, 3, $N_1$, $N_{1+1}$, $N_{1+2}$,…, N.

Supongamos que existen motivos para sospechar que hubo un cambio estructural al pasar del periodo $N_1$ al $N_{1+1}$. Para verificar si tal sospecha es aceptable se efectúa el test de CHOW, cuyo estadístico de prueba es un derivación del test de Razón de Verosimilitud.

Se divide la muestra original en 2 submuestras:

1ª submuestra) de n = 1, 2, 3, …, $N_1$.
2ª submuestra) de n = $N_{1+1}$, $N_{1+2}$,…, N

Se realiza el contraste:

$$H_0 : Permanencia\ estructural$$
$$H_A : Cambio\ estructural$$

La permanencia estructural implica estabilidad instrumental de los parámetros del modelo.

El estadístico de prueba es:

$$\hat{F}_{\exp} = \frac{\frac{\overline{VE}_N - (\overline{VE}_1 + \overline{VE}_2)}{k}}{\frac{\overline{VE}_1 + \overline{VE}_2}{N-2k}} = \frac{\frac{e'e_N - (e'e_1 + e'e_2)}{k}}{\frac{e'e_1 + e'e_2}{N-2k}} = \frac{\frac{SCR_N - (SCR_1 + SCR_2)}{k}}{\frac{SCR_1 + SCR_2}{N-2k}} \sim F_{(k,N-2k)}$$

## 2.8 ERRORES DE ESPECIFICACIÓN

Aquí se analiza como se ven afectadas las propiedades de los estimadores como consecuencia de incurrir en una especificación incorrecta de las variables exógenas que intervienen en el modelo de regresión.

Los dos errores de especificación que analizaremos son:

**1. Omisión de variables relevantes**

Sus consecuencias son:

i) Estimaciones sesgadas e inconsistentes de los coeficientes que afectan a las variables exógenas.

ii) Sobreestimación de la varianza del término de perturbación.

iii) Estimador de la varianza del término de perturbación sesgado.

iv) Invalidación del procedimiento inferencial.

**2. Inclusión de variables superfluas o irrelevantes**

Sus consecuencias son:

i) Estimaciones insesgadas de los coeficientes que afectan a las variables exógenas.

ii) Sobreestimación de la varianza de los estimadores de estos coeficientes.

iii) Estimador de la varianza del término de perturbación insesgado.

## 2.9 PREDICCIÓN EN EL MODELO DE REGRESIÓN

Una vez estimado y validado el modelo, una de sus aplicaciones más importantes consiste en poder realizar predicciones acerca del valor que tomaría la variable dependiente en el futuro o para una unidad extramuestral.

Esta predicción se puede realizar tanto para un valor individual como para un valor medio, o esperado, de la variable dependiente, siendo posible efectuar una predicción puntual o por intervalos. Su cálculo se realiza mediante las expresiones que figuran a continuación.

$$IC_{Y_f} \in \left[\hat{Y}_f \pm t_{n-k} S_y\right] \qquad IC_{Y_f} \in \left[\hat{Y}_f \pm t_{n-k}\hat{\sigma}_u \sqrt{1 + X'_f (X'X)^{-1} X_f}\right]$$