

Tema 3.2: Modelo lineal general: inferencia y predicción

Universidad Complutense de
Madrid

2013

Contraste de hipótesis paramétricas en el MLG (I)

- Necesitamos añadir a la lista de hipótesis del MLG, la H.8. El error sigue una distribución normal. Las hipótesis 7 del MLG fijan los momentos de primer y segundo orden de las perturbaciones, pero no su distribución.

$$\varepsilon \sim N(0, \sigma^2 I)$$

Muchos estudios econométricos tienen como objetivo responder a una pregunta que puede plantearse estadísticamente como un contraste de hipótesis. Ejemplos de contrastes en un modelo econométrico hay muchos:

- (1) Contrastar si un cambio legal afecta o no al precio medio de la vivienda. En este caso, se contrasta que la ley no haya tenido efecto y que el precio medio de la vivienda (descontados otros factores) es igual antes y después de la ley.

Contraste de hipótesis paramétricas en el MLG (II)

- (2) Contrastar si la productividad marginal del trabajo coincide con el salario (considerado como exógeno) y la empresa se encuentra en equilibrio.
- (3) Contrastar si hay discriminación salarial por sexo.
- (4) Contrastar si la publicidad afecta al consumo de cigarrillos, etc.
- (5) Contrastar la existencia de rendimientos constantes a escala en una función de producción de tipo Cobb-Douglas.
- (6) Contrastar si la elasticidad demanda-precio de un bien determinado es unitaria.

Contraste de hipótesis paramétricas en el MLG (III)

- Una hipótesis se contrasta comparando nuestras predicciones con la realidad. Si coinciden, dentro de un margen de error admisible, mantenemos la hipótesis. En caso contrario, la rechazamos y buscamos otras que sean capaces de explicar los datos
- Planteamos una hipótesis nula a priori creíble y sólo la rechazamos cuando existe suficiente evidencia en los datos en contra de la misma. Si rechazamos la nula, implícitamente no rechazamos otra hipótesis llamada alternativa.

Contraste de hipótesis paramétricas en el MLG (IV)

Suponiendo que la nula es simple, es decir, del tipo, $H_0 : \beta_i = S$ siendo S un valor conocido, la hipótesis alternativa puede ser de distintos tipos:

(A) Si rechazamos la nula, pero no sabemos en qué dirección puede ser falsa, por lo que la $H_1 : \beta_i \neq S$ (contraste bilateral o de dos colas).

(B) Si rechazamos la nula, pero sabemos forzosamente que se cumple que $H_1 : \beta_i > S$ (contraste unilateral o de una sola cola).

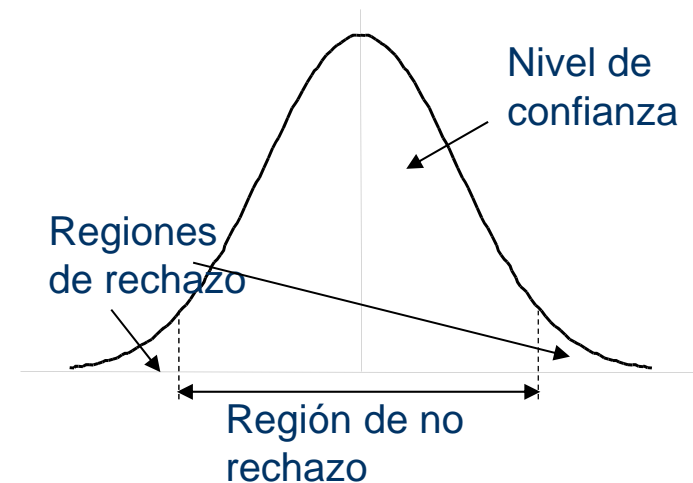
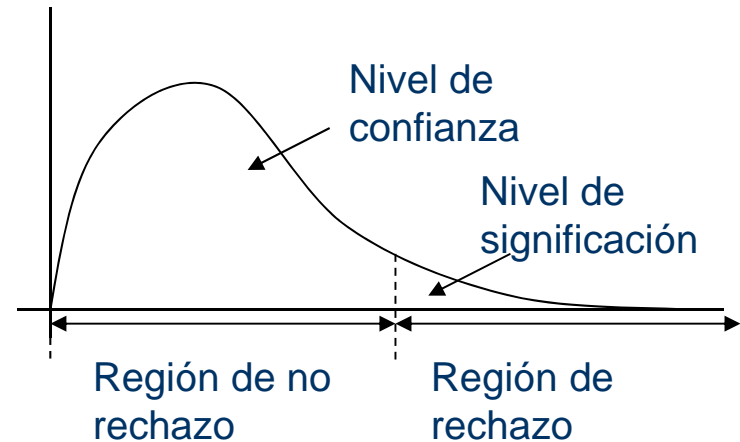
Contraste de hipótesis paramétricas en el MLG (IV)

Etapas del contraste: Son tres:

- (1) Definir claramente la hipótesis nula y la alternativa.
- (2) Definir una medida de discrepancia entre la hipótesis nula y los datos. La medida de discrepancia debe tener una distribución conocida cuando la nula es cierta. Una discrepancia es grande cuando tiene una probabilidad de ocurrir muy pequeña, siendo la nula cierta. La discrepancia es pequeña cuando es “probable” siendo la hipótesis nula cierta.
- (3) Definir una regla para decidir qué discrepancias son demasiado grandes como para atribuir las a errores de medida y qué discrepancias son pequeñas. Esto es definir **la zona de rechazo y de no rechazo de la nula.**

Contraste de hipótesis paramétricas en el MLG (V)

- Si el valor del estadístico calculado con la muestra está:
En la región de rechazo, se rechaza la nula en favor de la alternativa
Fuera de la región de rechazo, no se rechaza la nula
- Una decisión errónea es rechazar la nula, siendo cierta (error de tipo I). La probabilidad de cometer este error es el nivel de significación del contraste (que se suele escoger pequeño, del 1%, 5% ó del 10%).
- Se puede tener una o dos zonas de rechazo, dependiendo de si la distribución del estadístico es de una o de dos colas.
- Nivel de confianza = $1 - \text{Nivel de significación}$



Estadístico t (contraste bilateral) (I)

Supongamos que se ha estimado la función de producción de una empresa y se quiere contrastar si está en equilibrio igualando la productividad marginal del trabajo ($PmgL$) al salario (S). Es decir:

$$H_0 : \beta_i = S$$

$$H_1 : \beta_i \neq S$$

donde $\beta_i = PmgL$ y este parámetro se puede estimar por MCO en un modelo. Una respuesta errónea es comparar directamente la estimación del parámetro directamente con S . Esto es incorrecto por varias razones: en primer lugar, es muy difícil que coincida exactamente con S , aunque la nula sea cierta. Además, estamos suponiendo que ha sido estimado sin incertidumbre.

Estadístico t (contraste bilateral) (II)

Aprovechando la información que tenemos sobre $\hat{\beta}$ (estimador MCO del vector de parámetros) y que es una combinación lineal de los errores, sabemos que:

$$\hat{\beta} \sim N[\beta, \sigma^2 (X^T X)^{-1}]$$

y un estimador concreto, como el i -ésimo:

$$\hat{\beta}_i \sim N[\beta_i, \sigma^2 (X^T X)^{-1}_{ii}]$$

donde $(X^T X)^{-1}_{ii}$ es el elemento i -ésimo de la diagonal principal de la matriz $(X^T X)^{-1}$ y se asume la hipótesis de normalidad de las perturbaciones. Por insesgadez, la media de su distribución es β_i

Estadístico t (contraste bilateral) (III)

Estandarizando esta normal, se obtiene el ratio:

$$\frac{\hat{\beta}_i - \beta_i}{\sqrt{\sigma^2 (X^T X)^{-1}_{ii}}} \sim N(0,1)$$

y apoyándonos en la función de distribución de una normal estándar podríamos derivar el intervalo de valores en el que con una determinada probabilidad se encuentra el verdadero β_i . No obstante, esto sólo es posible si conocemos el valor de σ^2 (varianza de los errores)

Como lo más frecuente en la práctica es que sea un parámetro desconocido, debemos usar su estimador y su distribución (conocida a partir de la normalidad del error). ¹⁰

Estadístico t (contraste bilateral) (IV)

- En el caso MCO, sabemos que $\frac{n-k}{\sigma^2} \hat{\sigma}_{MCO}^2 \sim \chi_{n-k}^2$
- donde $\hat{\sigma}_{MCO}^2 = \frac{SR}{n-k}$
- Combinando esta distribución chi-cuadrado con la normal estándar anterior y sabiendo que ambas distribuciones son independientes, se obtiene una distribución t de Student. El **estadístico de contraste** es el ratio:

$$\frac{N(0,1)}{\sqrt{\chi_{n-k}^2 / n-k}} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\hat{\sigma}_{MCO}^2 (X^T X)^{-1}_{ii}}} = \frac{\hat{\beta}_i - \beta_i}{d\hat{t}(\hat{\beta}_i)} \sim t_{n-k}$$

Estadístico t (contraste bilateral) (V)

- Las dos maneras de contrastar esta hipótesis nula son:
- La primera, construir el intervalo de confianza del $\alpha\%$ para el parámetro y apoyándonos en la distribución t de Student, comprobar si el valor S cae dentro del intervalo (en cuyo caso, no podemos rechazar la nula) o no (en cuyo caso rechazamos la nula en favor de la alternativa a ese nivel de significación). Es decir, teniendo en cuenta la distribución t de Student:

$$1 - \alpha = \text{Pr ob} \left[-t_{\alpha/2} \leq \frac{\hat{\beta}_i - \beta_i}{d\hat{t}(\hat{\beta}_i)} \leq t_{\alpha/2} \right]$$

donde $t_{\alpha/2}$ es un valor crítico que deja en ambas colas de la distribución $\alpha/2\%$ masa de probabilidad.

Estadístico t (contraste bilateral) (VI)

O bien:

$$1 - \alpha = \text{Pr ob}[\hat{\beta}_i - t_{\alpha/2} d\hat{t}(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + t_{\alpha/2} d\hat{t}(\hat{\beta}_i)]$$

La segunda forma es evaluar el estadístico de contraste en la nula y compararlo (en valor absoluto) con el valor crítico que deja a la derecha de la distribución $\alpha/2$ % masa de probabilidad.

Si $\left| \frac{\hat{\beta}_i - S}{d\hat{t}(\hat{\beta}_i)} \right| \geq t_{\alpha/2} \rightarrow$ se rechaza la nula a favor de la alternativa

Si $\left| \frac{\hat{\beta}_i - S}{d\hat{t}(\hat{\beta}_i)} \right| < t_{\alpha/2} \rightarrow$ no se rechaza la nula a ese nivel de significación

Estadístico t (contraste unilateral) (I)

- Por alguna razón, normalmente porque así lo dicta la teoría económica, podemos descartar valores de un parámetro negativos (o positivos). Si queremos contrastar que:
$$H_0 : \beta_i = 0 \quad H_1 : \beta_i > 0$$
- Bajo la hipótesis alternativa se espera que el valor del estadístico t sea positivo. Buscamos un valor de la t positivo y suficientemente grande como para rechazar la nula en favor de la alternativa. Valores negativos de la t no proporcionan evidencia en favor de la H_1 , sino en favor de la H_0 .
- Si $t \geq t_\alpha$ se rechaza la nula, donde t_α deja a la derecha el $\alpha\%$ masa de probabilidad. Si $t < t_\alpha$ no podemos rechazar la nula a ese nivel de significación. ¹⁴

Contraste de significación individual de un parámetro: el ratio t (I)

- Este contraste es un caso particular del anterior. Dado un modelo lineal general de regresión:

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_i x_{ti} + \dots + \beta_k x_{tk} + \varepsilon_t$$

- se quiere contrastar $H_0 : \beta_i = 0$ $H_1 : \beta_i \neq 0$
- Queremos saber si la variable x_{ti} tiene efecto sobre la variable endógena una vez que ésta ha sido explicada por todos los regresores, salvo la x_{ti}
- El estadístico de contraste es conocido como el ratio t y es el resultado de dividir la estimación puntual del parámetro entre su desviación típica estimada.

Contraste de significación individual de un parámetro: el ratio t (II)

- El ratio t es:

$$t = \frac{\hat{\beta}_i}{d\hat{t}(\hat{\beta}_i)} \sim t_{n-k}$$

- El estadístico tiene el mismo signo que el coeficiente estimado, ya que la desviación típica siempre es positiva
- Un valor de la estimación del parámetro muy alejado del cero, proporcionará evidencia en contra de la nula y a la inversa. No obstante, hay un error de estimación del parámetro medido por su desviación típica, por lo que la distancia de la estimación al cero hay que ponderarla por su error estándar. **El estadístico t , de hecho, mide en cuántas desviaciones típicas está alejado el estimador MCO del parámetro del cero.**

Contraste de significación individual de un parámetro: el ratio t (III)

Dada la hipótesis alternativa, queremos saber si x_{ti} tiene efecto sobre la y , sin especificar si el efecto es positivo o negativo (contraste bilateral).

Si $|t| > t_{\alpha/2} \longrightarrow$ se rechaza la nula al α %

Si $|t| \leq t_{\alpha/2} \longrightarrow$ no se rechaza la nula a ese nivel

Si se rechaza la hipótesis nula, decimos que x_{ti} es una variable individualmente significativa a ese nivel de significación y si no se puede rechazar, decimos que no es estadísticamente significativa a ese nivel de significación. Estos contrastes se llevan a cabo para todos los parámetros. Son parte de la diagnosis de un modelo.

Cálculo e interpretación del p-value

(I)

- Hay un componente de arbitrariedad en el enfoque clásico de contrastar una hipótesis escogiendo previamente el valor del nivel de significación del contraste. En realidad, no existe un nivel de significación correcto (se suele trabajar al 1%, 5% ó al 10%).
- Además, puede ser engañoso trabajar con un nivel de significación prefijado. Por ejemplo, queremos contrastar la significatividad individual de una variable en un modelo, teniendo 40 grados de libertad y siendo el valor del estadístico $t=1.85$. En este caso, el valor crítico de la distribución es igual a 2.021 al 5%, pero es igual a 1.684 al 10%. Por tanto, la hipótesis nula no se rechaza al 5%, pero sí al 10% (ya que $t=1.85 < 2.021$ y $t=1.85 > 1.684$)

Cálculo e interpretación del p-value (II)

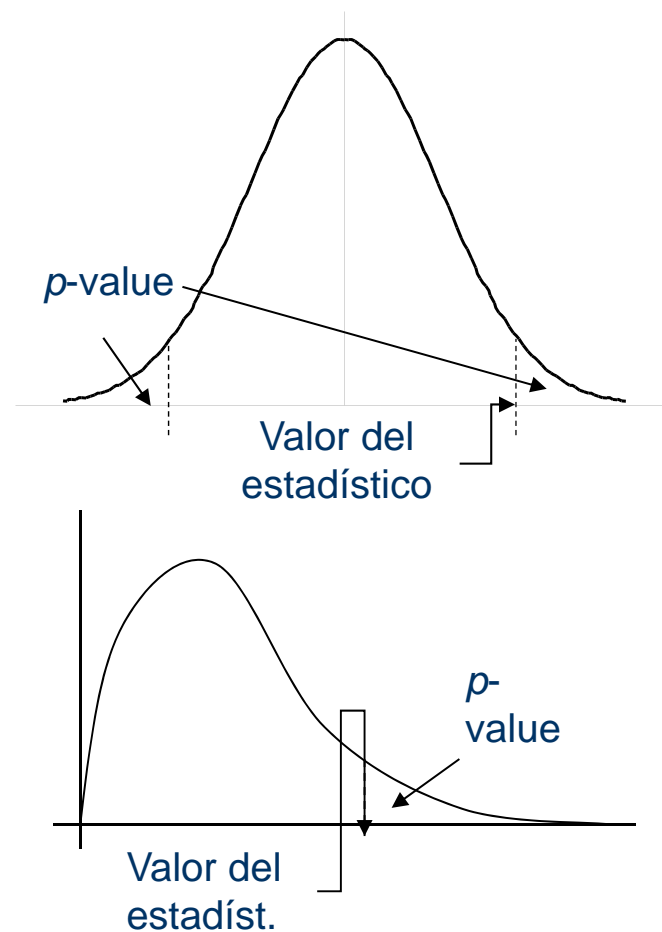
- En lugar de llevar a cabo un contraste usando distintos niveles de significación prefijados, **es más conveniente calcular el nivel de significación más pequeño al que se rechazaría la nula en cada caso (éste es el valor p o p -value)**. En el contraste de significación individual de un parámetro, el p -value se define como:

$$p = \text{Pr ob} [|t_{n-k}| > |t| / H_0 \text{ es cierta}]$$

- donde t_{n-k} representa una variable aleatoria que sigue una distribución t de Student con $n-k$ grados de libertad y t es el valor del estadístico de contraste bajo la nula.

Cálculo e interpretación del p-value (III)

- El **p-value** se puede interpretar como una medida del riesgo de rechazar la nula erróneamente.
- **Cuanto más pequeño es el p-value, con mayor contundencia rechazan los datos la nula a favor de la alternativa.**
- En el ejemplo anterior, el p-value del contraste es 0.0718
Como $0.0718 > 0.05$, no se rechaza la nula al 5%, pero sí al 10% ya que $0.0718 < 0.10$



Cálculo e interpretación del p-value (IV)

En el caso de un **contraste de una sola cola**, como por ejemplo $H_0 : \beta_i = 0$ frente a la $H_1 : \beta_i > 0$

El **p-value del contraste se define como:**

$$p = \text{Pr ob}[t_{n-k} > t / H_0 \text{ es cierta}]$$

Y en el contraste $H_0 : \beta_i = 0$ frente a la $H_1 : \beta_i < 0$

$$p = \text{Pr ob}[t_{n-k} < t / H_0 \text{ es cierta}]$$

donde t_{n-k} representa una variable aleatoria que sigue una distribución t de Student con $n-k$ grados de libertad y t es el valor del estadístico de contraste bajo la nula

Contraste general de hipótesis: el estadístico F (I)

- Supongamos que se desea contrastar un conjunto de m restricciones lineales sobre los parámetros del modelo:

$$Y = X\beta + \varepsilon$$

- En este caso, no es posible usar la distribución t de Student (sólo posible cuando $m=1$), sino una distribución F de Snedecor, que se define como el ratio de dos distribuciones chi-cuadrado independientes entre sí y ponderadas por sus grados de libertad:

$$F = \frac{\chi_m^2 / m}{\chi_{n-k}^2 / n - k} \sim F_{m, n-k}$$

- siendo $n-k$ los grados de libertad del modelo y m el nº de restricciones bajo la hipótesis nula.

Contraste general de hipótesis: el estadístico F (II)

En este caso general, la hipótesis nula se puede escribir como: $H_0 : A\beta = c$ donde A es una matriz conocida de tamaño $(m \times k)$ y c es un vector conocido de tamaño $(m \times 1)$. **Ejemplo:** supongamos que en un modelo de regresión como

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t$$

se quiere contrastar: $H_0 : \beta_1 = 0, \beta_2 = 0$

La hipótesis nula en forma matricial se escribe como:

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}; c = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Contraste general de hipótesis: el estadístico F (III)

Es decir, la matriz A tiene tantas filas como restricciones hay en la nula y tantas columnas como parámetros hay en el modelo. Lo mismo pasa con el vector columna c .

Cualquier restricción (es) lineal (es) entre los parámetros se puede escribir en la forma matricial anterior.

Contraste general de hipótesis: el estadístico F (IV)

La idea del contraste es medir la distancia a la que está el estimador MCO de satisfacer la hipótesis nula.

Esta distancia está definida por el vector $d = A\hat{\beta} - c$

Si la discrepancia es “grande” se rechazará la nula y si es “pequeña” no se podrá rechazar la hipótesis nula.

No obstante, hay que conseguir una forma de medir el tamaño de ese vector d (distancia) que sea escalar, adimensional y que tenga una distribución conocida bajo la nula.

A partir de la normalidad del estimador MCO de β y de la distribución chi-cuadrado del estimador de σ^2 (varianza de las perturbaciones).

Contraste general de hipótesis: el estadístico $F(V)$

Se sabe que: $\frac{d^T [\text{var}(d)]^{-1} d}{m} \sim \frac{\chi_m^2}{m}$

$$\frac{d^T [\text{var}(d)]^{-1} d}{m} = \frac{(A\hat{\beta} - c)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - c)}{m\sigma^2} \sim \frac{\chi_m^2}{m}$$

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi_{n-k}^2}{n-k}$$

donde $d = A\hat{\beta} - c$ y bajo la hipótesis nula su varianza es:

$$\text{var}(d) = E\{(A\hat{\beta} - A\beta)(A\hat{\beta} - A\beta)^T\} = E\{A(\hat{\beta} - \beta)(\hat{\beta} - \beta)^T A^T\}$$

y estimando por MCO $\text{var}(d) = A \text{var}(\hat{\beta}) A^T = A\sigma^2 (X^T X)^{-1} A^T$

Contraste general de hipótesis: el estadístico F (VI)

Dividiendo ambas distribuciones chi-cuadrado independientes entre sí, obtenemos:

$$F = \frac{d^T [\text{var}(d)]^{-1} d}{m\hat{\sigma}^2 / \sigma^2} = \frac{(A\hat{\beta} - c)^T [A(X^T X)^{-1} A^T]^{-1} (A\hat{\beta} - c)}{m\hat{\sigma}^2} \sim F(m, n - k)$$

Para calcular el estadístico F de contraste es necesario: (1) Estimar el modelo por MCO, (2) Calcular el vector distancia d y (3) Invertir la matriz $A(X^T X)^{-1} A^T$ de dimensión m .

Contraste general de hipótesis: el estadístico F (VII)

A partir de la distribución F , se escoge un nivel de significación tal que F_α es un valor crítico que deja a la izquierda el $(1-\alpha)\%$ de probabilidad.

Si $F \geq F_\alpha \rightarrow$ se rechaza la nula en favor de la alternativa

Si $F < F_\alpha \rightarrow$ no se rechaza la hipótesis nula a ese nivel de significación

Es decir, se rechaza cuando el tamaño de ese vector distancia es suficientemente grande con respecto a la distribución F de Snedecor.

El estadístico F y su relación con sumas de cuadrados de residuos (I)

La forma más sencilla de calcular el estadístico de contraste que sigue una distribución F , consiste en comparar las sumas de cuadrados de los residuos MCO de dos modelos: (1) el **modelo libre** (o modelo bajo la hipótesis alternativa) y (2) el **modelo restringido** (o modelo bajo la hipótesis nula).

Ejemplo: supongamos que en el siguiente modelo lineal

(1)

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_{1t}$$

se quiere contrastar la significatividad individual de la primera pendiente, es decir, $H_0 : \beta_1 = 0$ frente a la

$$H_1 : \beta_1 \neq 0$$

El estadístico F y su relación con sumas de cuadrados de residuos (II)

El modelo libre es el (1) donde al estimar por MCO se obtendrá la suma de cuadrados de residuos, denotada por SRL . El modelo restringido (o bajo la hipótesis nula) es aquél en el que no aparece el regresor x_{t1} . Es decir:

(2)

$$y_t = \beta_0 + \beta_2 x_{t2} + \varepsilon_{2t}$$

donde la estimación por MCO devolverá una suma de cuadrados de residuos diferente, denotada por SRR . El numerador del estadístico F se puede calcular como la distancia entre la SRR y la SRL ponderada por la varianza residual de ese modelo:

$$F = \frac{(SRR - SRL) / m}{SRL / n - k} \sim F_{m, n-k}$$

El estadístico F y su relación con sumas de cuadrados de residuos (III)

La idea del contraste mediante sumas residuales es sencilla.

(1) Supongamos que la hipótesis nula es cierta. Entonces, la variable no es significativa y no puede empeorar mucho el ajuste al usar el modelo restringido, en donde no aparece esta variable. Si la nula es cierta, ambas sumas de cuadrados de residuos no se diferenciarán mucho y el numerador del estadístico F es “pequeño”, indicando que la distancia al cumplimiento de la nula es “pequeña” y no se rechazará.

(2) Cuando la nula es falsa, la variable x_{t1} es relevante y por tanto, al eliminarla en el modelo restringido, habrá pérdida en términos de ajuste (la SRR será mayor que la SRL). Por ello, cuanto mayor es la diferencia entre la SRR y la SRL , más fácil es rechazar la nula en favor de la alternativa.

Contraste de significación conjunta de los parámetros (I)

Es importante realizar algún **contraste de significación conjunta de todos los parámetros** (además de los contrastes de significación individual). Es otro contraste que sirve para validar el modelo estimado.

El más habitual en la práctica es contrastar la significación conjunta de todas las pendientes del modelo (es decir, de todos los parámetros menos el término constante).

Vemos a continuación, que usar el estadístico F en función de sumas de cuadrados de residuos es también la forma más sencilla de realizar este contraste.

Contraste de significación conjunta de los parámetros (II)

Sea el MLG $y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_{1t}$

donde se quiere contrastar que $H_0 : \beta_1 = 0$
 $\beta_2 = 0$
.
 $\beta_k = 0$

Nótese que se excluye la constante de la hipótesis nula

El modelo restringido, o bajo la nula, es $y_t = \beta_0 + \varepsilon_{2t}$

El estadístico de contraste es:

$$F = \frac{(SRR - SRL) / k}{SRL / n - k} = \frac{(R_L^2 - R_R^2) / k}{(1 - R_L^2) / n - k} = \frac{R_L^2 / k}{(1 - R_L^2) / n - k}$$

Contraste de significación conjunta de los parámetros (III)

Obsérvese que se han dividido las dos sumas de cuadrados de residuos por la suma total ST (que es la misma en los modelos libre y restringido)

R_L^2 es el coeficiente de determinación del modelo libre,
 R_R^2 es el del modelo restringido y

se sabe que en el modelo restringido $y_t = \beta_0 + \varepsilon_{2t}$
el coeficiente de determinación es nulo.

Como puede observarse, este contraste de significación global tan sólo necesita el cálculo del R -cuadrado del modelo libre (Gretl lo calcula siempre por defecto al estimar por MCO cualquier modelo lineal)

Contraste de significación conjunta de los parámetros (IV)

Nótese que el estadístico F en función de las sumas de cuadrados de residuos de los dos modelos (libre y restringido) se puede usar siempre.

Si la(s) restricción(es) que se contrasta(n) no cambia(n) la ST del modelo restringido, se puede usar el estadístico F en función de los R -cuadrados de ambos modelos.

El estadístico F en función del R -cuadrado del modelo libre sólo sirve para contrastar la significación conjunta de las pendientes del modelo.

Predicción en el MLG (I)

Uno de los usos más frecuentes del MLG es la predicción de los valores futuros de la variable de interés. Las predicciones se usan para tomar decisiones. Suponiendo que tenemos una muestra de tamaño N , **la mejor relación entre las variables en el período $N+1$, será:**

$$y_{N+1} = x_{N+1}^T \beta + \varepsilon_{N+1}$$

donde $x_{N+1}^T = \left[x_{N+1,1} \quad x_{N+1,2} \quad \dots x_{N+1,k} \right]$

ε_{N+1} es el error del modelo en el instante $N+1$, que **seguimos asumiendo que tiene esperanza nula, varianza constante, ausencia de autocorrelación con las demás perturbaciones y una distribución normal.**

Predicción en el MLG (II)

- La predicción óptima de la variable y , usando datos hasta el instante N , a horizonte un período hacia adelante, es la esperanza condicionada a ese conjunto de información. Es decir:

$$\hat{y}_{N+1} = E_N[y_{N+1}] = E_N[x_{N+1}^T \beta + \varepsilon_{N+1}] = E_N[x_{N+1}^T \beta] + E_N[\varepsilon_{N+1}]$$

- Para calcular esta predicción puntual es necesario:

- (1) Calcular el término $E_N[x_{N+1}^T \beta] = E_N[x_{N+1}^T] \hat{\beta}$

donde $\hat{\beta}$ es el estimador MCO obtenido con N observaciones.

Predicción en el MLG (III)

(2) Predecir el valor futuro de las exógenas x_{N+1}

Si las variables explicativas son variables de control (por una empresa o por un banco central), se puede considerar conocido el valor futuro de las mismas. Por ejemplo, es posible predecir el volumen de ventas de una empresa asumiendo un determinado gasto futuro en publicidad. En caso contrario, habría que predecir el valor de las exógenas, para después predecir el valor de la endógena. Este es el problema más importante de la predicción en un MLG.

(3) Calcular el término $E_N[\varepsilon_{N+1}]$

Puesto que el error del modelo no tiene autocorrelación, el error ε_{N+1} es independiente de lo que haya pasado hasta el instante N y su historia pasada no ayuda a predecir el futuro. Por tanto, la esperanza condicional es idéntica a la esperanza incondicional e igual a cero.

Predicción en el MLG (IV)

La predicción puntual de la endógena del modelo a horizonte un período y considerando conocido el valor futuro de las exógenas es igual a:

$$\hat{y}_{N+1} = x_{N+1}^T \hat{\beta}$$

donde $\hat{\beta}$ es el estimador MCO usando N observaciones.

Predicción por intervalo Es más interesante proporcionar un intervalo de valores en el que, con una determinada probabilidad, caerá el valor futuro de la variable endógena. Para ello, definimos el error de predicción a horizonte 1 período, como la distancia entre el verdadero valor de la endógena en ese período y la predicción puntual del mismo:

$$\hat{\varepsilon}_N(1) = y_{N+1} - \hat{y}_{N+1} = -x_{N+1}^T (\hat{\beta} - \beta) + \varepsilon_{N+1}$$

Predicción en el MLG (V)

El error de predicción tiene esperanza nula y varianza igual a:

$$\text{var}[\hat{\varepsilon}_N(1)] = E\left\{ [-x_{N+1}^T(\hat{\beta} - \beta) + \varepsilon_{N+1}][\varepsilon_{N+1} - (\hat{\beta} - \beta)^T x_{N+1}] \right\}$$

donde la esperanza de los dos términos cruzados es nula; la esperanza de los términos al cuadrado dependen de la varianza del error futuro y de la matriz de varianzas del estimador MCO de β , es decir,

$$\text{var}[\hat{\varepsilon}_N(1)] = E\left\{ \varepsilon_{N+1}\varepsilon_{N+1}^T \right\} + x_{N+1}^T E\left\{ (\hat{\beta} - \beta)(\hat{\beta} - \beta)^T \right\} x_{N+1}$$

$$\text{var}[\hat{\varepsilon}_N(1)] = \sigma^2 + x_{N+1}^T \sigma^2 (X^T X)^{-1} x_{N+1}$$

Predicción en el MLG (VI)

- La distribución del error de predicción es una variable normal con media nula y la varianza que acabamos de derivar:

$$y_{N+1} - \hat{y}_{N+1} \sim N[0, \sigma^2 + x_{N+1}^T \sigma^2 (X^T X)^{-1} x_{N+1}]$$

- Estandarizando esta normal y usando el estimador y la distribución de σ^2 ,

$$\frac{\hat{\sigma}^2}{\sigma^2} \sim \frac{\chi_{n-k}^2}{n-k}$$

- se llega al ratio

$$\frac{y_{N+1} - \hat{y}_{N+1}}{\sqrt{x_{N+1}^T \hat{\sigma}^2 (X^T X)^{-1} x_{N+1} + \hat{\sigma}^2}} \sim t_{n-k}$$

Predicción en el MLG (VII)

El intervalo de valores en el que estará el valor futuro de la variable, con una confianza del $(1-\alpha)\%$ es:

$$\hat{y}_{N+1} - t_{\alpha/2} d\hat{t}[\hat{\varepsilon}_N(1)] \leq y_{N+1} \leq \hat{y}_{N+1} + t_{\alpha/2} d\hat{t}[\hat{\varepsilon}_N(1)]$$

donde $t_{\alpha/2}$ es un valor crítico de la distribución t de Student que deja en ambas colas el $\alpha/2\%$ de masa de probabilidad y la desviación típica estimada del error de previsión es:

$$d\hat{t}[\hat{\varepsilon}_N(1)] = \sqrt{x_{N+1}^T \hat{\sigma}^2 (X^T X)^{-1} x_{N+1} + \hat{\sigma}^2}$$

El intervalo está centrado en la previsión puntual y su anchura depende de la desviación típica de ese error.

Predicción en el MLG (VIII)

Nótese que **cuanto mayor sea la desviación típica del error de previsión**, más ancho será el intervalo de confianza de la previsión y menos fiable (y menos útil en la práctica será la misma).

La **previsión por intervalo siempre está centrada en la previsión puntual** de la variable.

$$\hat{y}_{N+1} - t_{\alpha/2} d\hat{t}[\hat{\varepsilon}_N(1)] \leq y_{N+1} \leq \hat{y}_{N+1} + t_{\alpha/2} d\hat{t}[\hat{\varepsilon}_N(1)]$$

Especificación: de lo general a lo particular y viceversa (I)

- En cualquier modelo se supone que existe un conjunto de variables explicativas que determinan a la “y”. Incluso aunque sólo se esté interesado en medir el efecto de una de las “x” sobre la “y”, **es importante no excluir del modelo variables *a priori***.
- **Razón:** cambios en estas variables excluidas pueden hacer variar a la “y” . Por tanto, todos los cambios en la “y” serán atribuidos a cambios en las “x’s” incluidas. Por otro lado, la lista de variables que “posiblemente” tengan influencia sobre la “y” puede ser muy larga....
- **La cuestión es ¿cuántas variables explicativas hay que incluir?**

Especificación: de lo general a lo particular y viceversa (II)

- En el procedimiento *bottom-up*, se comienza con el modelo más simple. Es decir, se explica a la “y” a través de una constante y una “x”. A continuación, se contrasta la significación individual del parámetro asociado a esa “x”. Si la hipótesis nula se rechaza, se incluye el segundo regresor y se vuelve a contrastar la significatividad individual de esta nueva variable.
- Por tanto, las variables son añadidas al modelo una a una y se dejan de incluir variables cuando alguna deja de ser individualmente significativa.

Especificación: de lo general a lo particular y viceversa (III)

- En el procedimiento denominado *top-down*, se comienza con el modelo más completo posible (se incluyen la constante y k variables explicativas)
- Se contrasta que $H_0 : \beta_k = 0$ frente a que $H_1 : \beta_k \neq 0$
- Si no se rechaza la nula, se vuelve a contrastar la hipótesis nula conjunta dada por $H_0 : \beta_k = \beta_{k-1} = 0$ y así, sucesivamente.
- Las variables se van eliminando hasta que el siguiente regresor sale significativo. Ambos métodos tienen ventajas e inconvenientes.