Libre Empresa vol. 17, No. 2, julio - diciembre de 2020, p. 121-138. (ISSN 1657-2815 / e-ISSN 2538-9904)

Modelo de máxima verosimilitud *

Maximum likelihood model

Modelo de probabilidade máxima

Alberto Gómez-Mejía
Investigador y profesor de Economía. Universidad Libre Seccional Cali, Cali - Colombia
alberto.gomezm@unilibre.edu.co https://orcid.org/0000-0002-0312-2236

Resumen

El objetivo de este artículo es hacer una introducción al modelo de Máxima Verosimiltiud (MV), ampliamente usado desde hace décadas en estadística, biometría, ingeniería y en econometría. A pesar de su utilidad, los cursos básicos de econometría siguen haciendo énfasis en los Mínimos Cuadrados Ordinarios (MCO) por lo fácil de su matemática y comprensión conceptual, y dejan el MV para los ejercicios con el software comercial que lo incluye por configuración , por la superioridad de los resultados frente a los del MCO. El MV es bastante utilizado para las regresiones no lineales y muestras grandes, por ejemplo, modelos de variables dependientes dicotómicas como Logit y Probit; heterocedasticidad condicional como GARCH y EGARCH, modelos censurados y truncados, etc. Es de prever que con el desarrollo de la inteligencia artificial en el data science y el machine learning, el MCO será descartado.

Palabras clave

Optimización; maximización; eficiencia asintótica; insesgamiento; grandes muestras; restricciones; Prueba de Wald; Multiplicador de Lagrange.

Codigo JEL

C50, C51, C52

Abstract

The objective of this article is to make an introduction to the Maximum Likelihood (MV) model, widely used for decades in statistics, biometrics, engineering and econometrics. Despite its usefulness, basic econometrics courses continue to emphasize Ordinary Least Squares (OLS) due to its ease of mathematics and conceptual understanding and leave the MV for exercises with commercial software that does include it by default, due to the superiority of the results compared to those of the OLS. MV is widely used for non-linear regressions and large samples, for example, models of dichotomous dependent variables such as Logit and Probit; conditional heteroscedasticity such as GARCH and EGARCH, censored and truncated models, etc. It is expected that as artificial intelligence develops in data science and machine learning, OLS will be discarded.

Keywords

Optimization; maximization; asymptotic efficiency; unbiased; large samples; constraints; Wald test; Lagrange multiplier.

F.R. 01/09/2020 F.A. 18/12/2020

* **Como citar:** Gomez-Mejia, A. (2021). Modelo de máxima verosimilitud. Libre Empresa, 17(2). 121-138. https://doi.org/10.18041/1657-2815/libreempresa.2020v17n2.8027

Resumo

O objetivo deste artigo é introduzir o modelo de Máxima Probabilidade (ML), que tem sido amplamente utilizado por décadas em estatística, biometria, engenharia e econometria. Apesar de sua utilidade, os cursos básicos de econometria ainda enfatizam os mínimos quadrados ordinários (OLS) por causa de sua fácil compreensão matemática e conceitual, e deixam a MV para exercícios com software comercial que a inclui por configuração, por causa da superioridade dos resultados em relação ao OLS. O OLS é amplamente utilizado para regressões não lineares e grandes amostras, por exemplo, modelos variáveis dicotômicas dependentes como Logit e Probit; heterocedasticidade condicional como GARCH e EGARCH, modelos censurados e truncados, etc. Espera-se que com o desenvolvimento da inteligência artificial na ciência dos dados e na aprendizagem de máquinas, o OLS seja descartado.

Palavras-chave

Otimização; maximização; eficiência assimptótica; imparcialidade; grandes amostras, restrições; teste de Wald; multiplicador de Lagrange

1. Introducción

Las estimaciones de MCO se derivan de ajustar un modelo lineal que representa el promedio de los datos. El modelo se obtiene al minimizar la suma de las desviaciones o errores al cuadrado. La función de verosimilitud indica qué tan probable es reproducir la muestra observada en función de los posibles valores de los parámetros, es decir, el fin es maximizar la función de verosimilitud con base en los parámetros que tienen más probabilidades de producir los datos observados. Técnicamente, se recomienda MV para muestras grandes por su adaptabilidad a los diferentes tipos de datos, y producir estimaciones más precisas.

Las ventajas de MV sobre el MCO son:

1. Mejores estimaciones de los parámetros de distribución
2. Menor varianza de los parámetros (eficiencia)
3. Confiabilidad en la medición de los intervalos de confianza y en las pruebas de hipótesis de los parámetros.
4. En los modelos de datos censurados y truncados, el MV usa todos los datos incluido los censurados mientras que MCO los excluye.

2. Marco teórico

El tema es tratado ampliamente en los libros de econometría: Chow (1983), Judge, George G., Carter R. Hill, William E. Griffiths, Helmut Lutkepohl y Tsoung-Chao Lee (1984), Lütkepohl, H. (1991), Maddala, G.S. (1992), Johnston, J. y Di Nardo (1996), Davidson (2000) Pindyck, Robert y Daniel Rubinfeld(2002), Gujarati y Porter (2010), Wooldridge,(2010). Greene (2018).

3. MV para regresión simple y distribución normal

El método de máxima verosimilitud (MV) o máxima probabilidad, busca estimar los parámetros que maximicen la probabilidad de observar las Y dadas, para lo cual es necesario maximizar la función.

Partiendo de la función de distribución normal para una variable

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} . e\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma\sqrt{2\pi}} . e\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

se escribe la multiplicatoria, es decir, multiplicando esta expresión "n" veces,

$$f(X) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} . e\left[-\frac{1}{2\sigma^2}(X-\mu)^2\right]$$

adicionalmente, con base en el modelo de regresión lineal con una sola variable independiente ($Y_i = \beta_1 + \beta_2 X_i + U_i$) donde Y_i es una variable independiente e idénticamente distribuida (i.i.d.), con distribución normal y varianza de los residuos σ^2, se aplica a una función de probabilidad conjunta de ($Y_i, Y_2, \dots Y_n$), se escribe:

$$f(Y_1, Y_2, \dots, Y_n | \beta_1 + \beta_2 X_i, \sigma^2)$$

$$= f(Y_1|\beta_1 + \beta_2 X_i, \sigma^2) f(Y_2|\beta_1 + \beta_2 X_i, \sigma^2) \dots f(Y_n|\beta_1 + \beta_2 X_i, \sigma^2)$$

Sustituyendo Y_i y su valor esperado en la distribución normal

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} e\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\}$$

se construye la función de verosimilitud FV,

$$f(Y_1, Y_2, \dots, Y_n|\beta_1 + \beta_2 X_i, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\}$$

$$= \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} e\left\{-\frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\}$$

donde los beta y la varianza residual son desconocidos

$$FV(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} e\left\{-\frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\}$$

Debido a la existencia del término exponencial, la optimización es mejor si se expresa en logaritmos naturales:

$$L = \ln(U_i) = -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$$

$$= -n\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}$$

Derivando con respecto a β_1 e igualando a cero:

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)(-1) = 0$$

Se obtiene: $\widehat{\beta_1} = \bar{Y} - \widehat{\beta_2}\bar{X}$; $\bar{Y} = \widehat{\beta_1} + \widehat{\beta_2}\bar{X}$ igual que con MCO. Derivando con respecto a β_2 e igualando a cero:

$$\frac{\partial L}{\partial \beta_2} = \frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)(-X_i) = 0$$

Se obtiene el mismo estimador de MCO:

$$\widehat{\beta_2} = \frac{\sum X_i Y_i - n\bar{X}\bar{Y}}{\sum X_i^2 - n\bar{X}^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

3.1.Varianza de los residuos

Derivando con respecto a σ^2 e igualando a cero:

$$\frac{\partial L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum (Y_i - \beta_1 - \beta_2 X_i)^2 = 0$$

Despejando σ^2 y sustituyendo $\widehat{\beta_1}$ y $\widehat{\beta_2}$:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum (Y_i - \widehat{\beta_1} - \widehat{\beta_2} X_i)^2$$

$$= \frac{1}{n}\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

$$\widehat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n}$$

El estimador de la varianza residual hallado por MV no varía a pesar del cambiante número de variables independientes que tenga el modelo, una o k, haciéndolo sesgado, mientras que el del MCO si cambia con el número de variables lo que lo hace insesgado: para una es $\widehat{\sigma}^2 = \sum \hat{u}_i^2/(n-2)$; con dos variables es $\widehat{\sigma}^2 = \sum \hat{u}_i^2/(n-3)$; para (k-1) es $\widehat{\sigma}^2 = \sum \hat{u}_i^2/(n-k)$

3.2. Medición de la magnitud del sesgo

Asuma un modelo con una sola variable independiente. Obtenga el valor esperado de las dos varianzas y réstelas para obtener el tamaño del sesgo. El valor esperado de la varianza del MV:

$$E\left[\frac{\sum \hat{u}_i^2}{n}\right] = \left[\frac{\sum \sigma^2}{n}\right] = \left[\frac{n\,\sigma^2}{n}\right] = \sigma^2$$

El valor esperado de la varianza del MCO:

$$E\left[\frac{\sum \hat{u}_i^2}{n-2}\right] = \left[\frac{\sum \sigma^2}{n-2}\right] = \left[\frac{n\,\sigma^2}{n-2}\right] = \sigma^2\left[\frac{n}{n-2}\right]$$

Restando, el sesgo es:

$$\sigma^2\left[\frac{n}{n-2}\right] - \sigma^2 = \sigma^2\left[\frac{n}{n-2} - 1\right] = \left[\frac{2}{n-2}\right]\sigma^2$$

Sin embargo, a medida que la muestra tienda a infinito, el término en paréntesis (2/(n-2)) o sesgo tiende a cero y las varianzas de los dos modelos tienden a ser iguales: las dos varianzas son asintóticamente insesgadas y consistentes.

4. MV para regresión lineal múltiple

Dado $Y = X\beta + U$ con $U \sim N(0, \sigma^2 I)$, la función de densidad normal multivariada es:

$$f(Y) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(U'U)}$$

$$f(Y) = \frac{1}{\sqrt{2\pi\sigma^2}} . e\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma\sqrt{2\pi}} . e\left[-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right]$$

Aplicando la multiplicatoria:

$$f(Y) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} . e\left[-\frac{1}{2\sigma^2}(X-\mu)^2\right]$$

es decir, multiplicando esta expresión "n" veces, se obtiene:

$$f(Y) = \frac{1}{(2\pi\sigma^2)^{n/2}} . e\left[-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)\right] = \frac{1}{(2\pi\sigma^2)^{n/2}} . e\left[-\frac{1}{2\sigma^2}U'U\right]$$

La función de log-likelihood es:

$$L = \ln f(U) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)$$

Derivando L con respecto a β:

$$\frac{\partial L}{\partial \beta} = -\frac{1}{\sigma^2}(-X'Y + X'X\beta)$$

Igualando a cero y despejando para β, se obtiene la fórmula del beta óptimo:

$$\boldsymbol{\hat{\beta} = (X'X)^{-1}X'Y}$$

La cual es igual al Beta del MCO y es insesgado.

Derivando L con respecto a σ^2 e igualando a cero:

$$\frac{\partial L}{\partial \sigma^2} = \frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y-X\beta)'(Y-X\beta)$$

Se obtiene:

$$\boldsymbol{\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/n = (U'U)/n}$$

Resultado diferente del de MCO. Aplicando el valor esperado:

$$E(\hat{\sigma}^2) = \sigma^2(n-k)/n$$

Este indica que la varianza de los residuos del VM es sesgada como se explicó en el aparte anterior.

5. La Matriz de Información de Fisher

Dada una $f(X/\beta)$, la matriz de información $I(\beta)$ es igual a la varianza de la función de resultado (score function) o varianza de la derivada de la función de verosimilitud. Si $^{f}(X/\beta)$ tiene estructura exponencial, entonces I = E(U›U). Imponiendo algunas condiciones se tiene:

$$I(\beta) = -E\left[\left(\frac{\partial U}{\partial \beta}\right)\right]$$

Entonces, para

$$I(\beta) = E\left[\left(\frac{\partial l}{\partial \beta}\right)\left(\frac{\partial l}{\partial \beta}\right)'\right]$$

Llegando a una derivada de segundo orden:

$$I(\beta) = E\left[\left(\frac{\partial l}{\partial \beta}\right)\left(\frac{\partial l}{\partial \beta}\right)'\right] = -E\left[\frac{\partial^2 l}{\partial \beta\, \partial \beta'}\right]$$

donde β es un vector (kx1) y $(\frac{\partial l}{\partial \beta})$ es un vector columna con k derivadas parciales

$$\frac{\partial l}{\partial \beta} = \begin{bmatrix} \partial l/\,\partial \beta_1 \\ \partial l/\,\partial \beta_2 \\ \vdots \\ \partial l/\,\partial \beta_k \end{bmatrix}$$

A su vez, cada una de las derivadas parciales puede ser derivada con respecto a los otros elementos del vector con tal de construir una Hessiana (k x k).

$$\frac{\partial[\partial l/\,\partial \beta\,]}{\partial \beta_1} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_1^2} & \frac{\partial^2 l}{\partial \beta_1\, \partial \beta_2} & \cdots & \frac{\partial^2 l}{\partial \beta_1\, \partial \beta_k} \end{bmatrix}$$

$$\frac{\partial^2 l}{\partial \beta\, \partial \beta'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta_1^2} & \frac{\partial^2 l}{\partial \beta_1\, \partial \beta_2} & \cdots & \frac{\partial^2 l}{\partial \beta_1\, \partial \beta_k} \\ \frac{\partial^2 l}{\partial \beta_2\, \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_2^2} & \cdots & \frac{\partial^2 l}{\partial \beta_2\, \partial \beta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \beta_k\, \partial \beta_1} & \frac{\partial^2 l}{\partial \beta_k\, \partial \beta_2} & \cdots & \frac{\partial^2 l}{\partial \beta_k^2} \end{bmatrix}$$

Retomando la función de log-likelihood

$$L = \ln f(U) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(Y - X\beta)'(Y - X\beta)$$

se obtienen las derivadas parciales de segundo orden:

$$\frac{\partial^2 L}{\partial \beta\, \partial \beta'} = \frac{X'X}{\sigma^2} \quad \rightarrow \quad -E\left(\frac{\partial^2 l}{\partial \beta\, \partial \beta'}\right) = \frac{X'X}{\sigma^2}$$

$$\frac{\partial^2 L}{\partial\beta\,\partial\sigma^2} = \frac{X'U}{\sigma^4} \quad \rightarrow \quad -E\left(\frac{\partial^2 l}{\partial\beta\,\partial\sigma^2}\right) = 0$$

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{U'U}{\sigma^6} \quad \rightarrow \quad -E\left(\frac{\partial^2 L}{\partial(\sigma^2)^2}\right) = \frac{n}{2\sigma^4}$$

$$E(U'U) = n\sigma^2$$

la matriz de información es igual a:

$$I(\beta) = I\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = \begin{bmatrix}\frac{\partial^2 L}{\partial\beta\,\partial\beta'} & \frac{\partial^2 L}{\partial\beta\,\partial\sigma^2}\\ \frac{\partial^2 L}{\partial\beta\,\partial\sigma^2} & \frac{\partial^2 L}{\partial(\sigma^2)^2}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2}(X'X) & 0\\ 0 & \frac{n}{2\sigma^4}\end{bmatrix}$$

y su inversa:

$$I^{-1}\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix} = \begin{bmatrix}\sigma^2(X'X)^{-1} & 0\\ 0 & \frac{2\sigma^4}{n}\end{bmatrix}$$

La diagonal cero del a matriz inversa, indica que el $\hat{\beta}$ y $\hat{\sigma}^2$ y están distribuidos independientemente entre ellos.

$$\begin{bmatrix}\hat{\beta}_{MV}\\ \hat{\sigma}^2_{MV}\end{bmatrix} \overset{a}{\rightarrow} N\left[\begin{pmatrix}\beta\\ \sigma^2\end{pmatrix}, \begin{pmatrix}\sigma^2(X'X)^{-1} & 0\\ 0 & \frac{2\sigma^4}{n}\end{pmatrix}\right]$$

La matriz de información Fisher viene a ser:

$$-E\begin{bmatrix}\frac{\partial^2 L}{\partial\beta\,\partial\beta'} & \frac{\partial^2 L}{\partial\beta\,\partial\sigma^2}\\ \frac{\partial^2 L}{\partial\beta\,\partial\sigma^2} & \frac{\partial^2 L}{\partial(\sigma^2)^2}\end{bmatrix} = \begin{bmatrix}\frac{1}{\sigma^2}(X'X) & 0\\ 0 & \frac{n}{2\sigma^4}\end{bmatrix}$$

6. La desigualdad de Cramer-Rao[1,2]

La matriz de información mide cuanta "información" se tiene acerca del parámetro β. Si es un estimador insesgado de β, entonces se puede probar que la **$Var(\hat{\beta}) > [1/I(\beta)]$**

El término, $[1/I(\beta)]$ se llama la "Cota de Cramer-Rao" o valor mínimo de Cramer-Rao. Cuanto menor sea la varianza del estimador de β, más se sabe de β. Si hay más de un parámetro, lo anterior se puede generalizar diciendo que la diferencia entre las varianzas es semidefinida positiva

$$\mathbf{Var(\hat{\beta}) - I^{-1}(\beta) \geq 0}$$

7. Propiedades de los estimadores MV

Siguiendo los mismos pasos, en la regresión lineal, se puede probar que los estimadores del MV son insesgados, consistentes y eficientes asintóticamente.

1. Consistencia

$$\text{plim}\big(\hat{\beta}\big) = \beta$$

2. Normalidad asintótica con media β y con varianza igual a la inversa de I(β).

$$\hat{\beta} \sim N\big(\beta, I^{-1}(\beta)\big)$$

3. Eficiencia Asintótica

Dado que es el estimador MV del parámetro β, se deriva

$$\sqrt{n}\big(\hat{\beta} - \beta\big) \xrightarrow{d} N(0, \sigma^2)$$
$$\sqrt{n}\hat{\beta} \text{ es mayor que } \sigma^2$$
$$\sqrt{n}\big(\hat{\beta} - \beta\big) \xrightarrow{d} N(0, V)$$

Sintetizando, en la medida que la muestra crece:

$$\begin{bmatrix} \hat{\beta}_{MV} \\ \hat{\sigma}^2_{MV} \end{bmatrix} \xrightarrow{a} N\left[\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}\right]$$

8. MV para una distribución de Bernoulli

Para una muestra aleatoria X_1, X_2 ...X_n, donde solo puede tener dos reslultados: 0,1.

$$P\big[X_j = 1\big] = \rho$$
$$P\big[X_j = 0\big] = 1 - \rho$$

Siendo ρ, desconocido

Ejemplo: lanzar una moneda n veces, asumiendo que hay una probabilidad de 50% para cara y sello, entonces, X_j , para cara.; X_j, para sello. La pregunta es cómo estimar ρ. El valor esperado de ρ

$$E[\rho] = 0{,}5(1) + 0{,}5(0) = 0{,}5$$

La función de probabilidad:

$$f(X) = \rho^X(1-\rho)^{1-X} = \begin{cases} \rho, \text{si } X = 1 \\ 1-\rho, \text{si } X = 0 \end{cases}$$

La función de verosimilitud:

$$Ln(\rho) = f(X_1)f(X_2) \ldots f(X_n) = \prod_{i=1}^{n} f(X_i)$$

$$\ln(\rho) = \prod_{i=1}^{n} \rho^{X_i}(1-\rho)^{1-X_i}$$

Expandiendo la función:

$$\ln(\rho) = \rho^{\sum_{i=1}^{n} X_i}(1-\rho)^{\sum(1-X_i)} = \rho^{\sum X_i}(1-\rho)^{\sum 1 - \sum X_i} = \rho^{\sum X_i}(1-\rho)^{n-\sum X_i}$$
$$\ln(\rho) = \ln \rho^{\sum X_i} + \ln(1-\rho)^{n-\sum X_i}$$
$$\ln(\rho) = (\sum X_i)\ln \rho + (n - \sum X_i)\ln(1-\rho)$$

Estimando la derivada parcial con respecto a "ρ", igualar a cero y encontrar el estimador:

$$\frac{d\ln(\rho)}{d\hat{\rho}} = (\sum X_i)\frac{1}{\rho} + n\frac{1}{1-\rho}(-1) - (\sum X_i)\frac{1}{1-\rho}(-1) = 0$$

$$\frac{d\ln(\rho)}{d\hat{\rho}} = \frac{\sum X_i}{\rho} - \frac{n}{1-\rho} + \frac{\sum X_i}{1-\rho} = 0$$

pasando al otro lado,

$$\frac{\sum X_i}{\rho} = \frac{1}{1-\rho}[n - \sum X_i]$$

e invirtiendo,

$$\frac{(1-\rho)}{\rho} = \frac{n - \sum X_i}{\sum X_i}$$
$$\frac{1}{\rho} - 1 = \frac{n - \sum X_i}{\sum X_i}$$
$$\frac{1}{\rho} = \frac{n - \sum X_i}{\sum X_i} + 1 = \frac{n - \sum X_i + \sum X_i}{\sum X_i} = \frac{n}{\sum X_i}$$
$$\frac{1}{\rho} = \frac{n}{\sum X_i}$$
$$\boldsymbol{\hat{\rho} = \frac{\sum X_i}{n} = \bar{X}}$$

Estimador insesgado:

$$E[\hat{\rho}] = \frac{1}{n}\sum E(X_i) = \frac{1}{n}[\sum \rho(1) + \sum(1-\rho)0] = \frac{1}{n}[n\rho]$$
$$E[\hat{\rho}] = \frac{n\rho}{n} = \rho$$
$$\boldsymbol{E[\hat{\rho}] = \rho}$$

Varianza: $\mathrm{Var}(X) = \sum(X_i - \bar{X})^2 \rho s.$

$$f(X) = \rho^X(1-\rho)^{1-X} = \begin{Bmatrix} \rho \text{ si } X = 1 \\ (1-\rho) \text{ si } X = 0 \end{Bmatrix}$$

$$\mathrm{Var}(X_i) = E(X_i - \rho)^2 = E(X_i - E(X_i))^2 = E(X_i - \hat{\rho})^2 = E(X_i - \rho)^2$$

Para X_1=1, X_i =0, respectivamente.

$$\mathrm{Var}(X_i) = [(1-\rho)^2\rho + (0-\rho)^2(1-\rho)]$$
$$\mathrm{Var}(X_i) = [1 - 2\rho + \rho^2]\rho + \rho^2(1-\rho) = \rho - \rho^2$$
$$\boldsymbol{\mathrm{Var}(X_i) = \rho[1-\rho]}$$
$$\boldsymbol{\sigma(X) = \sqrt{\rho(1-\rho)}}$$

Aplicando el Teorema del Límite Central para una distribución normal estándar:

$$Zn = \frac{\overline{X}n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}[\overline{X}n - \mu]}{\sigma} \sim N[0,1]$$

aplicándolo a Bernouilli, hacienda $\hat{\rho} = \overline{X}$; $\mu = \rho$, entonces,

$$\frac{\sqrt{n}[\hat{\rho} - \rho]}{\sqrt{\rho o(1-\rho)}}$$

Lo que para muestras grandes, se obtiene una distribución normal estándar:

$$\frac{\sqrt{n}[\hat{\rho} - \rho]}{\sqrt{\rho(1-\rho)}} \sim N[0,1]$$

La prueba de hipótesis:

$$\begin{aligned} &\text{Ho: } \rho = 0{,}5 \\ &\text{Ha: } \rho \neq 0{,}5 \end{aligned}$$

sustituyendo, se obtiene el estadístico para la prueba normal estándar

$$\frac{\sqrt{n}[\hat{\rho} - 0{,}5]}{\sqrt{0{,}5 \times 0{,}5}} = \frac{\sqrt{n}[\hat{\rho} - 0{,}5]}{\sqrt{0{,}5^2}} = \frac{\sqrt{n}[\hat{\rho} - 0{,}5]}{0{,}5} = \mathbf{2\sqrt{n}[\hat{\rho} - 0,5]}$$

9. Pruebas para muestras grandes muestras

A partir del modelo de Máxima Verosimilitud, se explican tres pruebas:[3]

1. Razón de verosimilitud (log-likelihood ratio)
2. Wald
3. Multiplicador de Lagrange (Lagrange Multiplier, LM)

Dado el modelo de regresión lineal

$$Y = X\beta + U$$

donde Y es (n x 1) , X es (nx k) , β es (k x 1) y U es (n x 1) . Los residuos de la regresión poblacional se distribuyen según $U \sim N(0, \sigma^2 + \Sigma)$

Retomando la función de MV para la muestra es:

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln \sigma^{2n} |\Sigma| - \frac{1}{2}\frac{U'\Sigma^{-1}U}{\sigma^2}$$

Sustituyendo $U'\Sigma^{-1}U$ se obtiene:

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln \sigma^2 - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma^2}(Y - X\hat{\beta})'\Sigma^{-1}(Y - X\hat{\beta})$$

Derivando con respecto a los parámetros desconocidos nos da las conocidas condiciones de primer orden e igualando a cero:

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2} X'\Sigma^{-1}(Y - X\beta) = 0$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(Y - X\beta)'\Sigma^{-1}(Y - X\beta) = 0$$

Solucionando las condiciones de primer orden:

$$\hat{\beta}_{MV} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$$
$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'\Sigma^{-1}(Y - X\hat{\beta})}{n}$$

Los parámetros estimados tienen la distribución:

$$\begin{bmatrix} \hat{\beta}_{MV} \\ \hat{\sigma}^2_{MV} \end{bmatrix} \xrightarrow{a} N\left[\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'\Sigma^{-1}X)^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{pmatrix}\right]$$

9.1.Prueba de Razón de Verosimilitud (log-likelihood ratio, LR)[4]

Explicación conceptual

Esta es una prueba de hipótesis que ayuda a elegir el "mejor" modelo entre dos modelos anidados (nested models) o jerarquizados . "Modelos anidados" son aquellos que tienen las mismas variables, sin embargo, se puede construir otro modelo con un número menor de las mismas variables; por ejemplo, definir cuál de los siguientes modelos es el que mejor se ajusta a los datos reales: el modelo 1 tiene cuatro variables (altura, peso, edad, sexo); el modelo 2 tiene un menor número de variables (edad, sexo). El modelo 2 está anidado dentro del modelo uno porque solo tiene dos de las variables predictoras (edad, sexo); en otras palabras, el modelo pequeño no es más que el modelo grande, pero que, al restringir el valor de algunas variables a cero, se hace pequeño.

El LR compara los dos modelos; si el modelo 1 es mejor que el modelo 2, los parámetros adicionales del modelo más complejo se usan en los análisis posteriores. El modelo 2 por tener menos variables se llama restringido y su prueba (LR) porque restringe o asigna el valor de cero a las variables altura y peso, mientras, el modelo 1 se llama no restringido (LNR). Dado que el LR solo se usa para comparar modelos anidados, es necesario tener en cuenta, que el agregar variables adicionales siempre incrementará la probabilidad, sin embargo, esto es marginalmente decreciente, es decir, llegará un punto en el que agregar más variables no mejora significativamente el ajuste.

El mejor modelo es el que hace que los datos sean más probables, o maximiza la función de verosimilitud, fn (X t-1,…, Xn | β).

En la Ho se evalúa el modelo más pequeño; aceptar la Ho indica que el más pequeño es el mejor; rechazar la Ho significa que el modelo más grande es una mejora significativa sobre el más pequeño.

El LR comienza con una comparación de las puntuaciones de probabilidad de los dos modelos :

$$LR = 2 * (\ln LNR - \ln LR) \sim \chi^2(q)$$

el cual se suele presentar en forma de diferencia logarítmica,

$$LR = 2 * \ln(LNR/LR) \sim \chi^2(q)$$

y cambiando los signos

$$LR = -2 * (\ln LR - \ln LNR) \sim \chi^2(q)$$

El LR tiene aproximadamente una distribución de chi-cuadrado. Para probar si la diferencia entre los dos modelos es estadísticamente significativa, se debe considerar los grados de libertad que son iguales al número de variables adicionales en el modelo más grande.

Ejemplo : Considere los modelos Grande y Pequeño. El Grande tiene cuatro variables más que el Pequeño. Las hipótesis son:

$$H_0: \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$$

$$H_a: \beta_2 \neq 0; \beta_3 \neq 0;\ \beta_4 \neq 0;\ \beta_5 \neq 0;$$

Los resultados de la función de MV para cada modelo son:

Pequeño -LnL = 1650.04, Grande-LnL = 1645.73, entonces,

LR = 2 (1650.04 -1645.73) = 9.42

grados de libertad = 4 :Grande tiene 4 parámetros adicionales al Pequeño; igual a decir, los cuatro son iguales a cero o cuatro restricciones, q = 4.

En la tabla $\chi^2+(4)$ para nivel del 95%, el valor crítico (P = 0.05) = 9.49

Dado que LR (9.42) < valor crítico (9.49), se acepta Ho, es decir, el modelo Grande no se ajusta a los datos significativamente mejor que el Pequeño, por lo que se infiere que los coeficientes de las cuatro variables no son estadísticamente significativos, es decir, iguales a cero.

Explicación matemática

Retomando la función de verosimilitud: cuando no hay restricciones en el espacio de parámetros, se sabe que:

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\ln|\Sigma| - \frac{1}{2\sigma^2}(Y - X\beta)'\Sigma^{-1}(Y - X\beta)$$

Usando la notación del capítulo anterior con coeficientes restringidos:

$$H_0: R\beta_R = r$$

donde R es (q x k), donde (q < l); es el β que cumple la restricción. Para incorporar las restricciones en la función de verosimilitud se construye

$$\ln L_R = \ln L + 2\lambda'(R\beta - r)$$

donde puede haber cualquier número de restricciones sobre el vector de coeficientes. La estimación con restricciones genera estimadores "restringidos"$(\hat{\beta}_R, \hat{\sigma}_R^2)$, que al ser sustituidos en la función L, producen la LnL_R $(\hat{\beta}_R, \hat{\sigma}_R^2)$. El término del numerador, LnL_R, corresponde a la LnL restringida y el denominador a la LnL no restringida. Estos a su vez se pueden expresar como funciones de los residuos restringidos $(\hat{U}_R^2{}'\hat{U}_R^2)$ y no restringidos $(\hat{U}^2{}'\hat{U}^2)$ respectivamente.

Obteniendo las derivadas parciales con respecto a los parámetros desconocidos e igualándolas a cero, se llega a:

$$\hat{\beta}_R = \hat{\beta} - (X'\Sigma^{-1}X)^{-1}R'[R(X'\Sigma^{-1}X)^{-1}R']^{-1}(R\hat{\beta} - r)$$

$$\hat{\sigma}_R^2 = \frac{(Y - X\hat{\beta}_R)'\Sigma^{-1}(Y - X\hat{\beta}_R)}{n}$$

Defínase λ como la razón de verosimilitud.

$$\lambda = \frac{LnL_R\left(\hat{\beta}_R, \hat{\sigma}_R^2\right)}{LnL\left(\hat{\beta}, \hat{\sigma}^2\right)}$$

El valor de λ nunca podrá se mayor que 1 porque el valor de los restringidos es menor que el de los no restringidos, entonces, $LnL\left(\hat{\beta}, \hat{\sigma}^2\right) < LnL_R\left(\hat{\beta}_R, \hat{\sigma}_R^2\right)]$. Al ser λ muy pequeña, se aumenta la probabilidad de rechazar Ho.

La razón de verosimilitud para muestras grandes se define como:

$$LR = -2\ln\lambda = 2[\ln L\left(\hat{\beta}, \hat{\sigma}^2\right) - LnL_R\left(\hat{\beta}_R, \hat{\sigma}_R^2\right)] \overset{a}{\sim} \chi^2(q)$$

la cual converge asintóticamente a una distribución chi-cuadrada con q grados de libertad.

Otro planteamiento para derivar los estimadores es:

$$LnL_R\left(\hat{\beta}_R, \hat{\sigma}_R^2\right) = LnL - U'(R\beta - r)$$

donde U es (qx1), vector de multiplicadores de Lagrange. El estimado $\hat{\beta}_R$ cumple la restricción $(R\beta_R = r)$. Los residuos estimados de la regresión son:

$$\hat{U}_R = Y - X\hat{\beta}_R$$

la varianza de los residuos: $\hat{\sigma}_R^2 = \hat{U}_R^2{}'\hat{U}_R^2 / n$, Formas equivalentes de escribir LR son:

$$LR = n(LnL\hat{U}_R{}'\hat{U}_R - LnL\hat{U}{}'\hat{U})$$

$$= nLnL\left(1 + \frac{\hat{U}_R{}'\hat{U}_R - \hat{U}{}'\hat{U}}{\hat{U}{}'\hat{U}}\right)$$

$$nLnL\left(\frac{1}{1 - (\hat{U}_R{}'\hat{U}_R - \hat{U}{}'\hat{U})/\hat{U}{}'\hat{U}}\right) = nLnL\left(\frac{1}{1 - (SRCR - SRCNR)/SRCNR}\right) \overset{a}{\sim} \chi^2(q)$$

9.2. Prueba de Wald[5]

Aquí solo se halla $\hat{\beta}$ no restringido. El vector $(R\hat{\beta} - r)$ un valor cercano a cero, acepta Ho, de lo contrario se rechaza. Se demostró anteriormente que $\hat{\beta}$ tiene distribución normal asintótica con media vector $\hat{\beta}$ y varianza-covarianza $RI^{-1}R'$, la cual, $RI^{-1}8R' = \sigma^2 + (X'X)^{-1}$. Tomando la matriz información

$$(R\hat{\beta} - r)'[RI^{-1}(\beta)R']^{-1}(R\hat{\beta} - r)' \overset{a}{\sim} \chi^2(q)$$

donde q es el número de restricciones en R. Sustituyendo, se obtiene la prueba de Wald

$$W = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2} \overset{a}{\sim} \chi^2(q)$$

Wald también se puede expresar en término de residuos puesto que $\hat{\sigma}^2 = \hat{U}_i^2{}'\hat{U}_i^2/n$

$$\mathbf{W} = \frac{\mathbf{n}(\hat{U}_R^2{}'\hat{U}_R^2 - \hat{U}^2{}'\hat{U}^2)}{\hat{U}^2{}'\hat{U}^2} = \frac{\mathbf{n(SRCR - SRCNR)}}{\mathbf{SRCNR}} \overset{a}{\sim} \chi^2(\mathbf{q})$$

9.3.La Prueba del Multiplicador de Lagrange[6]

Regresando al Lagrange de la sección anterior, se tiene:

$$H = \frac{n}{-2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\ln|\Omega| - \frac{1}{2\sigma^2}(Y - X\beta)'\Omega^{-1}(Y - X\beta) + 2\lambda'(R\beta - r)$$

Las condiciones de primer orden son:

$$\frac{\partial H}{\partial \beta} = \frac{\partial \ln L}{\partial \beta} + 2R'\lambda = 0$$
$$\frac{\partial H}{\partial \sigma^2} = \frac{\partial \ln L}{\partial \sigma^2} = 0$$
$$\frac{\partial H}{\partial \lambda} = 2(R\beta - r) = 0$$

El significado de λ: la ganancia en probabilidad de aliviar las restricciones. Eso es,

$$\frac{\partial \ln L}{\partial \hat{\beta}_R} = \lambda$$

Solucionando para λ, se obtiene:

$$\hat{\lambda} = -[R(X'\Sigma^{-1}X)R']^{-1}(R\hat{\beta} - r)$$

la cual tiene distribución normal como:

$$\hat{\lambda} \xrightarrow{d} N([R(X'\Sigma^{-1}X)^{-1}R']^{-1}(r - R\beta), \sigma^2 R(X'\Sigma^{-1}X)R')$$

Por lo que tiende aproximadamente a la distribución chi-cuadrado:

$$\hat{\lambda}'[\hat{\sigma}_R^2 R(X'\Sigma^{-1}X)R']^{-1}\hat{\lambda} \sim^a \chi^2$$

La varianza del error se estima con base en el modelo restringido, o, de manera equivalente, se encuentra a partir de las condiciones de primer orden.
La manipulación adecuada nos daría

$$\mathbf{LM} = \frac{\hat{U}_R{}'\Sigma^{-1}\hat{U}_R - \hat{U}'\Sigma^{-1}\hat{U}}{\hat{\sigma}_R^2} = \frac{\mathbf{n(SRCR - SRCNR)}}{\mathbf{SRCR/n}} = \frac{\mathbf{n(SRCR - SRCNR)}}{\mathbf{SRCR}} \overset{a}{\sim} \chi^2(\mathbf{q})$$

Ejemplo:
La tasa de crecimiento de las manufacturas depende de los cambios en la tasa de crédito bancario y la variación real de la tasa de cambio del peso con respecto a 18 divisas internacionales, según el ITCR[7]. Esta es la regresión sin restricciones

$DMANUF_t = 0.0526 - 0.0008673*DCRED_{t-1} - 0.00122*DCRED_{t-5} + 0.1129*DITCR_{t-4}$

$R^2 = 0.2534$; SRCNR = 0.114986; n = 69; k = 4; gl = 65. Todos los coeficientes son significativos al 95%.
Regresión con restricciones: excluya las tasas de interés: los dos rezagos de DCRED, entonces se tiene dos restricciones: H^o: $\beta_2 + = 0, \beta_3 = 0; q = 2$

$DMANUF^t = 0.04915 + 0.16906*DITCR_{t-}$
$R^2 = 0.1147$; SRCR = 0.165217; n = 69; k = 2; gl = 67. Todos los coeficientes son significativos al 95%.

$$LR = nLnL\left(\frac{1}{1-\frac{SRCR - SRCNR}{SRCNR}}\right) = 69LnL\left(\frac{1}{1-\frac{0.165217 - 0.114986}{0.114986}}\right)$$
$$= 39.62 \overset{a}{\sim} \chi^2(2)$$

$$W = \frac{n(SRCR - SRCNR)}{SRCNR} = \frac{69(0.165217 - 0.114986)}{0.114986} = 30.14 \overset{a}{\sim} \chi^2(2)$$

$$LM = \frac{n(SRCR - SRCNR)}{SRCR} = \frac{69(0.165217 - 0.114986)}{0.165217} = 20.98 \sim \chi^2(2)$$

El valor crítico para con nivel de confianza del 95% es 5.99. Los tres estadísticos son mayores que 5.99, entonces se rechaza la Ho, es decir, se acepta Ha: los dos coeficientes de DCRED son significativos como lo indica la regresión sin restricciones.

9.4.Relación entre las pruebas8

En el ejemplo anterior, si supiéramos tanto W como $\hat{\sigma}^2$ las tres pruebas son numéricamente idénticas. En el ejemplo construido, hubo que estimar $\hat{\sigma}^2$. Los resultados muestran que, aunque los tres estadísticos de prueba tienen χ2 con q grados de libertad, diferirán numéricamente en la práctica. Cuando se estiman tanto W como $\hat{\sigma}^2$, las diferencias entre los valores observados de las estadísticas de prueba serán más marcadas. Entonces,

$$\mathbf{W \geq LR \geq LM}$$

Pero como todos son asintóticamente χ2 con los mismos grados de libertad, el punto crítico es siempre el mismo. Invariablemente, esto puede llevar a algún conflicto sobre si uno decide rechazar el nulo. También hay una relación gráfica que se evidencia en la Figura 1.

El multiplicador de Lagrange se basa en la pendiente de la función de verosimilitud en el estimador restringido. La prueba de razón de verosimilitud se basa en la diferencia de la función de verosimilitud evaluada en las estimaciones restringidas y no restringidas; es decir, el rango de probabilidad. La prueba de Wald se basa en la diferencia entre los estimadores restringidos y no restringidos; es decir, el dominio de la función de verosimilitud.

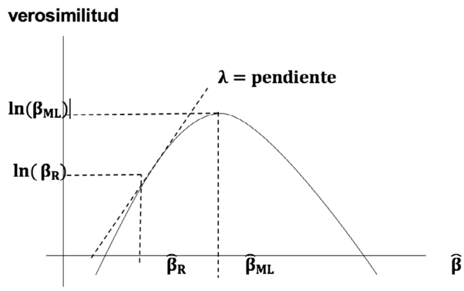

Figura 1. Comparación de pruebas MV
Fuente: Buck, Andrew J. Econometrics Lecture Notes. Temple University. http://www.ajbuckeconbikesail.net/notes/lectures.html

Las tres pruebas son asintóticamente iguales, pero, dan diferentes resultados en muestras finitas. Como se discutió anteriormente, las tres pruebas abordan la misma pregunta básica, ¿vale la pena restringir los parámetros a cero (es decir, omitir estas variables explicativas) para mejorar el ajuste del modelo?

Para realizar una prueba de razón de verosimilitud, se deben estimar los dos modelos que se desean comparar. La ventaja de las pruebas del ML y Wald es que se aproximan a la prueba LR, pero requieren que solo se calcule un modelo. Tanto la prueba del multiplicador de Wald como la de Lagrange son asintóticamente equivalentes a la prueba LR, es decir, a medida que el tamaño de la muestra se vuelve infinitamente grande, los valores de las estadísticas de la prueba del multiplicador de Wald y Lagrange se acercarán cada vez más a la estadística de prueba de la prueba LR. En muestras finitas, los tres tenderán a generar estadísticas de prueba algo diferentes, pero generalmente llegarán a la misma conclusión.

10. Conclusión

MV no es difícil de aprender, además, garantiza mejores estimaciones de los parámetros que los mínimos cuadrados ordinarios porque su estructura matemática se ajusta al comportamiento real de los datos. Es importante profundizar el modelo MV en los cursos de economía para mejorar su competitividad.

Notas

1. Cramér, Harald (1946). Mathematical Methods of Statistics. Princeton, NJ: Princeton Univ. Press. ISBN 0-691-08004-6. OCLC 185436716.
2. Rao, Calyampudi Radakrishna (1945). "Information and the accuracy attainable in the estimation of statistical parameters". Bulletin of the Calcutta Mathematical Society. 37: 81–89. MR 0015748.
3. Fox, J. (1997) Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications.
4. Johnston, J. and DiNardo, J. (1997) Econometric Methods Fourth Edition. New York, NY: The McGraw-Hill Companies, Inc.
5. Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. Transactions of the American Mathematical Society.

Vol. 54, No. 3 (Nov., 1943), pp. 426-482 (57 pages). Published By: American Mathematical Society. https://doi.org/10.2307/1990256, https://www.jstor.org/stable/1990256.

6. Engle. R. F. (1982a). "A General Approach to Lagrange Multiplier Model Diagnostics", Journal of Econometrics, 20:83-104.
7. ITCR: Índice de la tasa de cambio real, metodología del FMI. Fuente: www.banrep.gov.co.
8. Berndt, E. R. and N. E. Savin (1977), "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45:1263-1278. Ch. 13: Wald, Likelihood Ratio. and Lagrange Multiplier Tests 825.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Referencias bibliográficas

1. Amemiya, T. (1985). Advanced Econometrics. Oxford: Basil Blackwell. ISBN 0-631-13345-3.
2. Berndt, E. R. and Savin, N. E. (1977), "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model", Econometrica, 45 (5). 1263-1277. Ch. 13: Wald, Likelihood Ratio. and Lagrange Multiplier Tests 825. https://doi.org/10.2307/1914072
3. Buck, A. J. (2021). Econometrics Lecture Notes. Temple University. http://www.ajbuckeconbikesail.net/notes/lectures.html
4. Chow. G.C. (1983). Econometric Methods, McGraw-Hill, New York.
5. Cramér, H. (1946). Mathematical Methods of Statistics. Princeton, NJ: Princeton Univ. Press. ISBN 0-691-08004-6. OCLC 185436716. https://archive.org/details/in.ernet.dli.2015.223699
6. Engle. R. F. (1982a). "A General Approach to Lagrange Multiplier Model Diagnostics", Journal of Econometrics,20(1) 83-104. https://doi.org/10.1016/0304-4076(82)90104-X
7. Fox, J. (1997) Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications. https://psycnet.apa.org/record/1997-08857-000
8. Greene, W. (2012) Econometric Analysis, Pearson. http://www.mysmu.edu/faculty/zlyang/ECON6002_21-Web/William_H_Greene-Econometric_Analysis-7th.pdf
9. Gujarati, D. y Porter D. (2010) Econometría Básica. Ed. McGraw-Hill. https://fvela.files.wordpress.com/2012/10/econometria-damodar-n-gujarati-5ta-ed.pdf
10. Hamilton, J. (1994) Time Series Analysis, Princeton University Press.
11. Johnston, J. (1984) Econometric Methods, McGraw-Hill, 3rd ed.
12. Johnston, J. and DiNardo, J. (1997) Econometric Methods Fourth Edition. New York, NY: The McGraw-Hill Companies, Inc. https://economics.ut.ac.ir/documents/3030266/14100645/econometric%20methods-johnston.pdf
13. Judge, G. G., Carter Hill, R., Griffiths, W. E., Lutkepohl, H. y Lee, Tsoung-Chao (1984). Theory and Practice of Econometrics, John Wiley & Sons, Nueva York. https://pdfcoffee.com/george-g-judge-william-e-griffiths-r-carter-hill-helmut-ltkepohl-tsoung-chao-lee-the-theory-and-practice-of-econometrics-wiley-series-in-probability-and-statistics-wiley-1985-pdf-free.html
14. Kennedy, P. (2008) A Guide to Econometrics, Blackwell Publishing. http://masonlec.org/site/rte_uploads/files/Econometrics%20Book%20-%20Intro,%20Ch%201%20and%202.pdf
15. Lütkepohl, H. (1991). Introduction to Multiple Time Series Analysis, Springer Verlag, New York. https://www.springer.com/gp/book/9783540401728
16. Maddala, G.S. (1992) Econometria. Ed McGraw-Hill. https://www.academia.edu/15259731/INTRODUCCI%C3%93N_A_LA_ECONOMETR%C3%8DA_LICENCIATURA_EN_ADMINISTRACI%C3%93N_Y_DIRECCI%C3%93N_DE_EMPRESAS_2O_CURSO_FABIOLA_PORTILLO_FEBRERO

17. Novales, A. (1993). Econometría. 2da. edición. Ed. McGraw-Hill. https://es.scribd.com/doc/252571609/Econometria-Alfonso-Novales-2da-Edicion

18. Pindyck, R. y Rubinfeld D. (2002). Modelos econométricos y pronóstico económico. Ed. McGraw-Hill.

19. Pulido, A. (1987).Modelos econométricos. 2da. edición. Ed. Pirámide. ISBN 10: 8436802144ISBN 13: 9788436802146

20. Rao, C. R. (1948) "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Application to Problems of Estimation", Proceedings of the Cambridge Philosophical Society, 44150-57. https://doi.org/10.1017/S0305004100023987

21. Rao, C. R. (1945). "Information and the accuracy attainable in the estimation of statistical parameters". Bulletin of the Calcutta Mathematical Society. 37: 81–89. MR 0015748.

22. Wald, A. (1943). Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. Transactions of the American Mathematical Society. Vol. 54, No. 3 (Nov., 1943), pp. 426-482 (57 pages). Published By: American Mathematical Society. https://doi.org/10.2307/1990256

23. Wooldridge, J. M. (2010). Introducción a la Econometría. Un enfoque moderno. Editorial Thomson-Learning. https://herioscarlanda.files.wordpress.com/2018/10/wooldridge-2009-introduccic3b3n-a-la-econometrc3ada-un-enfoque-moderno.pdf