ECONOMETRÍA
APLICADA UTILIZANDO R.

PAPIME PE302513 LIBRO ELECTRÓNICO Y COMPLEMENTOS DIDÁCTICOS EN MEDIOS COMPUTACIONALES, PARA EL FORTALECIMIENTO DE LA ENSEÑANZA DE LA ECONOMETRÍA

Capítulo 5

Normalidad

Lucía A. Ruiz Galindo

Objetivo

Estudiar la importancia e implicaciones del supuesto de normalidad en el modelo de regresión lineal y de manera específica en la inferencia estadística de sus parámetros, y presentar en R, aplicaciones de la prueba de Jarque-Bera (Jarque-Bera 1980, 1987), utilizada para detectar si los términos estocásticos en el modelo siguen o no una distribución normal.

Introducción

Una vez estimado el modelo y habiendo aprobado la evaluación económica se llevan a cabo pruebas de especificación que se distinguen dos partes: 1) determinística: que es la combinación lineal de los parámetros y 2) estocástica: la que se asocia al término aleatorio.

Las pruebas de especificación correcta consisten en estudiar si la información empírica incorporada al modelo, la que se utiliza para la estimación proporciona evidencia a favor o en contra de los supuestos tanto determinística del modelo como e la aleatoria o estocástica.

Es decir, 1) la parte aleatoria: estudia si las variable independientes son las únicas que explican a la variable dependiente, si hay permanencia estructural en los parámetros y si la forma funcional en el que se han introducido la variables es correcta o no lo es y la parte estocástica 2) la parte estocástica: analizan los supuestos que Gauss-Markov, que establecen que los errores aleatorios tienen media cero, son independiente de las variables explicativas, homoscedásticos y no autocorrelacionados, y también se verifica que se satisfaga el supuesto de normalidad, todo ello se lleva a cabo usando pruebas de hipótesis estadísticas.

Modelo General de Regresión Lineal

Especificación:

$$y_t = \beta_1 + \beta_2 x_{t2} + \ldots + \beta_K x_{tK} + \varepsilon_t$$

donde $\beta_1, \ldots, \beta_K$ son los parámetros del modelo, y_t es la variable dependiente, las x_{tk}'s, $k = 2, \ldots, K$, son las variables independientes, ε_t es el término o error estocástico, t, $t = 1, \ldots, T$, es un índice que indica el número de la observación y T es el total de observaciones.

Supuestos:

- Linealidad en los parámetros.
- Las K-1 variables independientes son las únicas que explican a la dependiente.
- El número de observaciones T, es mucho mayor que el de parámetros K.
- Las variables explicativas son linealmente independientes de manera que ninguna es combinación lineal de otra o de otras y por tanto el rango de X es K.
- Los parámetros no cambian en la muestra, es decir, hay permanencia estructural

Modelo General de Regresión Lineal

Supuestos Gauss-Markov:

- $E(\varepsilon_t) = 0, \ \forall t = 1, \ \ldots, T$
- $\{\varepsilon_1, \varepsilon_2, \ \ldots, \varepsilon_T\}$ y $\{x_{1k}, x_{2k}, \ \ldots, x_{Tk}\}$ son independientes $\forall k = 2, \ \ldots, K.$
- $V(\varepsilon_t) = \sigma^2, \ \forall t = 1, \ \ldots, T.$
- $Cov(\varepsilon_t, \varepsilon_s) = 0, \ \forall t, s = 1, \ \ldots, T, \ t \neq s.$
- ε_t se distribuye Normal, $\forall t = 1, \ \ldots, T.$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Estimación Puntual: MCO y MV

MV: $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$ constituyen una muestra aleatoria

Función de verosimilitud y log-verosimilitud

$$\mathrm{l}(\beta_1, \ldots, \beta_K;\ \varepsilon_1, \ldots, \varepsilon_T) = \left(\frac{1}{2\sigma^2}\right)^{\frac{T}{2}} - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

$$\mathrm{l}(\beta_1, \ldots, \beta_K;\ \varepsilon_1, \ldots, \varepsilon_T) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\varepsilon_t^2$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Distribuciones de los parámetros

$$\hat{\beta}_k \sim N\left(\beta_k, \sigma^2 \, (X'X)_{ii}^{-1}\right)$$

$$(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}$$

- Intervalos de confianza para β_k.

$$\sigma^2 \text{ conocida,} \quad \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 \, (X'X)_{ii}^{-1}}} \sim N(0,1)$$

$$\sigma^2 \text{ desconocida,} \quad \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 \, (X'X)_{kk}^{-1}}} \sim tS_{T-K}.$$

$$P\left(\hat{\beta}_k - \tau_{\alpha/2}\sqrt{\hat{\sigma}^2 \, (X'X)_{ii}^{-1}} \le \beta_k \le \hat{\beta}_k + \tau_{\alpha/2}\sqrt{\hat{\sigma}^2 \, (X'X)_{ii}^{-1}}\right) = 1 - \alpha$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para β_k (σ^2 desconocida).

$$\mathrm{H_0:\ \beta_k = b_k} \quad \text{vs} \quad \mathrm{H_1:\ \beta_k \neq b_k}$$

Estadístico de prueba bajo $\mathrm{H_0}$:

$$\tau = \frac{\hat{\beta}_k - b_k}{\sqrt{\hat{\sigma}^2\,(X'X)^{-1}_{kk}}} \sim tS_{T-K}$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para combinaciones lineales de β_k´s (σ^2 desconocida).

$$H_0:\ R\beta = r \quad \text{vs} \quad H_0:\ R\beta \neq r$$

Estadístico de prueba bajo H_0:

$$f = \frac{1}{m\hat{\sigma}^2}\left(R\hat{\beta} - r\right)'[R(X'X)^{-1}R']^{-1}\left(R\hat{\beta} - r\right) \sim F_{(m,T-K)}$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Prueba de significancia estadística

Individual

$$H_0 \colon \beta_k = 0 \quad \text{vs} \quad H_1 \colon \beta_k \neq 0$$

Conjunta

$$H_0 \colon R\beta = 0 \quad \text{vs} \quad H_0 \colon R\beta \neq 0$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Intervalo de confianza para σ^2

$$(T-K)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{T-K}$$

$$P\left((T-K)\frac{\hat{\sigma}^2}{\chi_{\alpha/2}} \leq \sigma^2 \leq (T-K)\frac{\hat{\sigma}^2}{\chi_{1-\alpha/2}}\right) = 1-\alpha$$

Modelo General de Regresión Lineal

Importancia Normalidad en la Inferencia Estadística

- Prueba de hipótesis para σ^2

$$H_0:\ \sigma^2 = s \quad \text{vs} \quad H_1: \sigma^2 \neq s,$$

Estadístico de prueba bajo H_0:

$$(T-K)\frac{\hat{\sigma}^2}{s^2} \sim \chi^2_{T-K}$$

Modelo General de Regresión Lineal

Prueba de Normalidad de Jarque-Bera (JB)

- Prueba estadística

$$H_0:\ s = 0, c = 3 \quad \text{vs} \quad H_1:\ s \neq 0 \ \frac{y}{o} c \neq 3$$

- Estadístico de prueba

$$JB = T\left[\frac{\widehat{cs}^2}{6} + \frac{(\widehat{cc} - 3)^2}{24}\right] \sim \chi^2_{(2)}$$