PRESENTACIÓN

Este documento titulado *Pruebas de Diagnóstico* y elaborado por la Dra. Lucía A. Ruiz Galindo, es un reporte de investigación del proyecto **Análisis Multivariado de Series de Tiempo**, aprobado por el Consejo Divisional de Ciencias y Humanidades y registrado con el número 607.

El objetivo de este trabajo es estudiar las Pruebas de Diagnóstico más comúnmente utilizadas en la evaluación econométrica de un modelo. Dentro de ellas se estudian las debidas a Durbin y a Watson, y la Breusch y Godfrey para probar no autocorrelación de los errores; las de White, Koenker-Basset y Breush y Pagan, para analizar su heteroscedasticidad; la de Jarque y Bera para saber si se tienen errores normales; y la de Ramsey para analizar su correcta especificación. Todas ellas y otras más se deben de llevar a cabo para saber si la información incorporada en el modelo da evidencia a favor en contra de cada uno de los supuestos que se hacen sobre su término estocástico o bien, sobre los correspondientes a la parte determinista del mismo.

Dr. Abelardo Mariña Flores
Jefe del Departamento de Economía

PRUEBAS DE DIAGNÓSTICO

por

LUCÍA ATZIMBA RUIZ GALINDO

Reporte de Investigación

UNIVERSIDAD AUTÓNOMA METROPOLITANA - AZCAPOTZALCO
División de Ciencias Sociales y Humanidades
Departamento de Economía

Diciembre, 2016

Pruebas de Diagnóstico

Las pruebas que se presentan en las siguientes Secciones se basan en el modelo general de regresión lineal

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_K x_{tK} + \varepsilon_t. \tag{1}$$

y cuando no es así, en la Sección correspondiente se formula el modelo adecuado para llevar a cabo la prueba que se va a presentar. Además, cuando en las pruebas se hace uso de regresiones auxiliares el término estocástico que en ellas se incorpora es υ_t, y se supone satisface los supuestos de Gauss-Markov.

1. Pruebas de autocorrelación

1.1. Durbin-Watson

Durbin y Watson (D-W)(Durbin y Watson, 1950) proponen probar autocorrelación de primer orden, es decir,

$$\varepsilon_t = \theta_0 + \theta_1 \varepsilon_{t-1} + \upsilon_t \tag{2}$$

donde υ_t es el término estocástico que satisface los supuestos G-M. Observe que si cuando $\theta_1 = 0$, no existe autocorrelación de primer orden entre los errores, de otra manera sí la hay. Las hipótesis a probar son

$$H_0 : \text{No autocorrelación} \quad \text{vs} \quad H_1 : \text{Autocorrelación}.$$

o equivalentemente

$$H_0 : \theta_1 = 0 \quad \text{vs} \quad H_1 : \theta_1 \neq 0.$$

El estimador de θ_1 en 2 esta dado por

$$\hat{\theta}_1 = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2}.$$

Sin embargo, el estadístico que formulan *D-W* es

$$\begin{aligned} DW &= \frac{\sum_{t=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} \\ &= \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t^2 - 2\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1} + \sum_{t=2}^{T} \hat{\varepsilon}_{t-1}^2}{\sum_{t=1}^{T} \hat{\varepsilon}_t^2} = 2(1 - \hat{\theta}_1), \end{aligned}$$

donde $\hat{\theta}_1$ es el estimador de θ_1 en 2

$$\hat{\varepsilon}_t = \theta_1 \hat{\varepsilon}_{t-1} + \upsilon_t$$

[1]

Observe en la regresión auxiliar que cuando $\hat{\theta}_1 = 0$, no existe autocorrelación, en cuyo caso $DW = 2$. Sin embargo, cuando hay evidencia a favor de que los errores están autocorrelacionados ($\hat{\theta}_1 \neq 0$), la autocorrelación puede ser positiva, $0 < \hat{\theta}_1 < 1$, o negativa $-1 < \hat{\theta}_1 < 0$, en el primer caso $0 < DW < 2$, y en el segundo $2 < DW < 4$. ¿Qué sucede entonces cuando el $2 < DW < 4$? La prueba no es concluyente.

D-W establecieron un límite superior (d_S) y uno inferior (d_I) porque existen valores críticos que no pueden ser tabulados dentro de la distribución que ellos crearon, por lo que establecen que

- Si $DW < d_I$, se rechaza la hipótesis nula de no autocorrelación.
- Si $DW > d_S$, no se rechaza la hipótesis nula.
- Si $d_I < DW < d_S$, la prueba no es concluyente.

[1]Aquí y en todo lo que sigue, υ_t denota el error estocástico en todas las regresiones auxiliares y por tanto, tiene los mismos supuestos al correspondiente modelo general de regresión lineal presentado con anterioridad. Esta especificación de la regresión auxiliar corresponde a un modelo autorregresivo de primer orden, $AR(1)$, el cual para que sea estacionario debe de satisfacer $\theta_1 < 1$.

1.2. h-Durbin

A diferencia de la prueba de autocorrelación de D-W, la h-Durbin (Durbin, 1970) considera que el modelo de regresión inicial incorpora el primer rezago de la variable dependiente, es decir,

$$y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + ... + \beta_K x_{tK} + \alpha y_{t-1} + \varepsilon_t$$

y al igual que la D-W, es para probar autocorrelación de primer orden, por ello utiliza la misma regresión auxiliar

$$\hat{\varepsilon}_t = \theta \hat{\varepsilon}_{t-1} + \upsilon_t.$$

Las hipótesis a probar son las que se plantearon en la prueba D-W y el estadístico de la prueba es

$$hD = \hat{\theta}\sqrt{\frac{T}{1 - T[(ee(\hat{\alpha}))^2]}} \underset{t\to\infty}{\overset{H_0}{\sim}} N(0,1)$$

1.3. Breusch-Godfrey (Breusch y Godfrey, 1978)

La prueba se utiliza para analizar la autocorrelación de orden $q \geq 1$, las hipótesis a probar son

$$H_0 : \text{No autocorrelación} \quad \text{vs} \quad H_1 : \text{Autocorrelación}$$

para lo cual se considera la regresión auxiliar

$$\hat{\varepsilon}_t = \beta_1 + \beta_2 x_{t2} + ... + \beta_K x_{tK} + \rho_1 \hat{\varepsilon}_{t-1} + ... + \rho_q \hat{\varepsilon}_{t-q} + \upsilon_t,$$

de manera que si alguna $\rho_i \neq 0, i = 1, ..., q$, hay autocorrelación de orden $j = max\{1, 2, ..., q\}, \rho_j \neq 0$, por lo que las hipótesis a probar son

$$H_0 : \rho_1 = \rho_2 = ... = \rho_q = 0$$

vs

$$H_1 : \text{Existe al menos una } i \text{ tal que } \rho_i \neq 0,\ i = 1, ..., q.$$

Cuando el tamaño de la muestra es grande entonces se obtiene el siguiente estadístico de prueba

$$(T-q)R^2 \underset{T\to\infty}{\overset{H_0}{\sim}} \chi^2_{(q)}.$$

2. Pruebas de Heteroscedasticidad

- White con términos cruzados,
- White sin términos cruzados,
- Koenker-Basset y
- Breusch-Pagan.

Las pruebas de heteroscedasticidad mencionadas se basan en el modelo de regresión general formulado en 1 y en todas ellas las hipótesis a probar son

$$H_0 : \text{Homoscedasticidad} \quad \text{vs} \quad H_1 : \text{Heteroscedasticidad}.$$

2.1. White con términos cruzados

White (White, 1980) propone una prueba de heteroscedasticidad en la que considera la regresión auxiliar

$$\hat{\varepsilon}_t^2 \;=\; \beta_1 + \sum_{k=2}^{K} \beta_k x_{tk} + \sum_{k=2}^{K} \alpha_k x_{tk}^2 + \sum_{k=2}^{K} x_{tk} \sum_{j=k+1}^{K} \gamma_{kj} x_{tj} + \upsilon_t \tag{3}$$

y en términos de los parámetros en (1,3) la prueba es

$$H_0 : \beta_k = 0,\ \ \alpha_k = 0, \gamma_{kj} = 0,\ \forall\ \ k = 2, ..., K \text{ y } j = k+1, ..., K$$

vs

$$H_1 : \text{Existe al menos una } k \text{ tal que } \beta_k \neq 0,\ \ \alpha_k \neq 0 \text{ y/o una } \gamma_{kj} \neq 0.$$

Observe que al aplicar esperanza a 3, del lado izquierdo de la igualdad se obtiene la varianza del residual y del lado derecho, bajo H_0, se obtiene β_1 y por tanto, los errores son homoscedásticos.

2.2. White sin términos cruzados

Como su nombre lo indica esta prueba considera la regresión auxiliar 3 sin términos cruzados, esto es

$$\hat{\varepsilon}_t^2 = \beta_1 + \sum_{k=2}^{K} \beta_k x_{tk} + \sum_{k=2}^{K} \alpha_k x_{tk}^2 + \upsilon_t$$

y con base en ella, las hipótesis a probar son

$$H_0 : \beta_k = 0,\ \ \alpha_k = 0\ \forall\ \ k = 2, ..., K$$

vs

$$H_1 : \text{Existe al menos una } k \text{ tal que } \beta_k \neq 0 \text{ y/o } \alpha_k \neq 0.$$

2.3. Koenker-Basset

Un inconveniente de la prueba de White con términos cruzados es el gran número de variables independientes en la regresión auxiliar 3, ya que implica el uso de muchos grados de libertad, aún cuando la regresión original tenga un número moderado de variables independientes, por ello, Koenker-Basset (Koenker y Basset, 1978) proponen la regresión auxiliar

$$\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 \hat{y}_t + \alpha_2 \hat{y}_t^2 + ... + \alpha_n \hat{y}_t^n + \upsilon_t,$$

y las hipótesis a probar son

$$H_0 : \alpha_j = 0 \ \forall j = 1, ..., n \ \text{ vs } \ H_1 : \alpha_j \neq 0 \ \text{ para al menos una} j.$$

2.4. Breusch-Pagan

Breusch y Pagan (Breusch y Pagan, 1979) consideran σ_t^2 como la varianza del error y se describe como, $\sigma_t^2 = f(\alpha_1 + \alpha_2 z_{t2} + \alpha_3 z_{t3} + ... + \alpha_n z_{tn})$, es decir, σ^2 es algún tipo de función de las variables z no estocásticas, alguna de las x o todas ellas pueden servir como z, así que las hipótesis a probar son

$$H_0 : \alpha_2 = \alpha_3 = ... = \alpha_n = 0$$

vs

$$H_1 : \text{Existe al menos una } k \text{ tal que } \alpha_k \neq 0, \ k = 2, ..., n.$$

Y la regresión auxiliar es

$$\hat{\varepsilon}_t^2 / \tilde{\sigma}^2 = \alpha_1 + \alpha_2 z_{t2} + \alpha_3 z_{t3} + ... + \alpha_n z_{tn} + \upsilon_t,$$

donde $\tilde{\sigma}^2$ es estimador de máxima verosimilitud de σ^2.

Por lo tanto, las hipótesis a probar son

$$H_0 : \alpha_2 = \alpha_3 = ... = \alpha_n = 0$$

vs

$$H_1 : \text{Existe al menos una } k \text{ tal que } \alpha_k \neq 0,\ \ k = 2, ..., n.$$

El siguiente estadístico de prueba es el que se utiliza para verificar la existencia de heteroscedasticidad:

$$\Theta = \frac{1}{2}(SRC) \sim \chi^2_{(T-1)}.$$

3. Prueba de normalidad

La prueba de normalidad de Jarque-Bera (Jarque y Bera, 1987) es una prueba asintótica o de grandes muestras,[2] donde las hipótesis a probar son

$$H_0 : \text{Normalidad} \quad \text{vs} \quad H_1 : \text{No normalidad}$$

o bien,

$$H_0 : CS = 0 \ \ CC = 3 \quad \text{vs} \quad H_1 : CS \neq 0 \ \ \text{y/o} \ \ CC \neq 3,$$

donde CS y CC son los estimadores.

Y utiliza el siguiente estadístico de prueba

$$JB = T\left[\frac{\widehat{CS}^2}{6} + \frac{(\widehat{CC}-3)^2}{24}\right] \overset{H_0}{\sim} \chi^2_{(2)},$$

[2] Si x_i es una secuencia infinita de variables aleatorias independientes, tal que $E(x_i) = \mu_i < \infty$ y $V(x_i) = \sigma^2$ tienen un valor finito para $i = 1, 2, ..., t$.

esto es, $\widehat{CS} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t^3}{\left(\sqrt{\hat{\sigma}^2}\right)^3}$ y $\widehat{CS} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t^4}{\left(\sqrt{\hat{\sigma}^2}\right)^4}$.

4. Ramsey RESET

La prueba RESET (Ramsey, 1969) se usa para analizar si el modelo está bien especificado, partiendo del modelo de regresión lineal general, las hipótesis a probar son

$$H_0 : \text{Forma funcional correcta} \quad \text{vs} \quad H_1 : \text{Forma funcional incorrecta}$$

y la regresión auxiliar es

$$y_t = \beta_1 + \sum_{k=2}^{K} \beta_k x_{tk} + \sum_{j=2}^{n} \alpha_j \hat{y}_t^j + \upsilon_t.$$

Con ello las hipótesis a probar son de manera que las hipótesis anteriores se pueden escribir como

$$H_0 : \alpha_j = 0 \ \ \forall \ j = 2, ..., n$$

$$\text{vs}$$

$$H_1 : \text{Existe al menos una } j \text{ tal que } \alpha_j \neq 0, \text{ para al menos una} j = 2, ..., n.$$

Referencia

Breusch, T. S. (1978). “Testing for Autocorrelation in Dynamic Linear Models”, *Australian Economic Papers*, vol. 17, N° 31, pp. 334-355.

Breusch, T. S. y A. R. Pagan (1979). “A Simple Test for Heteroscedasticity and Random Coefficient Variation”, *Econometrica*, vol. 47, N°5, pp. 1287-1294.

Durbin, J. y G. S. Watson (1950). “Testing for Serial Correlation in Least Squares Regression”, *Biometrika*, vol. 37, N° 3-4, pp. 409-428.

Durbin, J. (1970). "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors Are Lagged Dependent Varia-bles", *Biometrika*, vol. 38, N° 3, pp. 410-421.

Godfrey, L. G. (1978). "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables", *Econometrica*, vol. 46, N°6, pp. 1293-1301.

Gujarati, D. N. (2004). *Econometría.* México D.F. 4^a ed. McGraw Hill.

Jarque, C. M. y A. K. Bera (1987). "A Test for Normality of Observations and Regression Residuals", *International Statical Review*, vol. 55, N° 2, pp. 163-172.

Johnston, J. y J. Dinardo (1997). *Econometric Methods.* United States of America. 4^a ed. McGraw Hill.

Koenker, R. y G. Basset Jr. (1978). "Regression Quantiles", *Econometrica*, vol. 46, N°1, pp. 33-50.

Ramsey, J. B. (1969). "Tests for Specification Error in Classical Linear Least Squares Analysis", *Journal of the Royal Statical Society, Series B (Methodological)*, vol. 31, N° 2, pp. 350-371.

White H. (1980). "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica*, vol. 48, N° 4, pp. 817-838.

Wooldridge, J. M. (2011). *Introducción a la econometría. Un enfoque moderno.* México, D.F. 4^a ed. Thomson Learning.