Métodos Estadísticos para Economía y Gestión IN 540 Clase 7

Perturbaciones no Esféricas

17 de junio de 2010

La autocorrelación es un problema que surge cuando rompemos el supuesto 5 de no autocorrelación en los errores. Ello implica que:

$$Cov(\epsilon_i, \epsilon_j) \neq 0 \qquad i \neq j$$

La autocorrelación en el término de error se da en los datos se serie de tiempo, donde es un problema bastante común.

Luego, nuestra matriz de varianzas y covarianzas del error ya no será una matriz diagonal (como en el caso de varianzas esféricas y no esférica pero sólo con heterocedasticidad) ya que el término de error se encuentra correlacionado consigo mismo a través del tiempo.

La forma que toma la matriz cuando sólo tenemos autocorrelación pero los errores son homocedásticos:

$$E[\epsilon\epsilon'] = \sigma^2\Omega = \begin{pmatrix} \sigma^2 & \sigma_{1,2} & \sigma_{1,3} & \cdots & \sigma_{1,T} \\ \sigma_{2,1} & \sigma^2 & \sigma_{2,3} & \cdots & \sigma_{2,T} \\ \sigma_{3,1} & \sigma_{3,2} & \sigma^2 & \cdots & \sigma_{3,T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{T,1} & \sigma_{T,2} & \sigma_{T,3} & \cdots & \sigma^2 \end{pmatrix}$$

donde $\sigma_{p,q} = cov(\epsilon_p, \epsilon_q)$.
Nuestro modelo ahora será:

$$\begin{aligned} Y_t &= X_t\beta + \epsilon_t \\ \epsilon_t &= \rho\epsilon_{t-1} + \eta_t \end{aligned}$$

al proceso que sigue ϵ_t se le denomina AR(1) (autorregresivo de orden 1).

Matriz de Varianzas y Covarianzas cuando ϵ_t es un AR(1)

En este caso el término de error tiene la forma señalada anteriormente:

$$\epsilon_t = \rho\epsilon_{t-1} + \eta_t$$

1. $V(\epsilon_t) = V(\rho\epsilon_{t-1} + \eta_t) = \rho^2 V(\epsilon_{t-1}) + \sigma^2_\eta$, de esta forma $V(\epsilon_t) = \frac{\sigma^2_\eta}{1-\rho^2} = \sigma^2$.
2. Como $E(\epsilon_t) = 0$, $Cov(\epsilon_t, \epsilon_{t-1}) = E(\epsilon_t\epsilon_{t-1})$. Calculemos esta última esperanza:

$$\begin{aligned}
\epsilon_t\epsilon_{t-1} &= \epsilon_{t-1} \cdot (\rho\epsilon_{t-1} + \eta_t) \\
&= \rho\epsilon^2_{t-1} + \epsilon_{t-1}\eta_t \\
E(\epsilon_t\epsilon_{t-1}) &= \rho E(\epsilon^2_{t-1}) + E(\epsilon_{t-1}\eta_t) \\
E(\epsilon_t\epsilon_{t-1}) &= \rho\sigma^2
\end{aligned}$$

3. Siguiendo la misma lógica anterior, $E(\epsilon_t\epsilon_{t-2})$ se calcula de la siguiente forma:

$$\begin{aligned}
\epsilon_t\epsilon_{t-2} &= \epsilon_{t-2}\cdot(\rho\epsilon_{t-1}+\eta_t) \\
&= \rho\epsilon_{t-2}\epsilon_{t-1}+\epsilon_{t-2}\eta_t \\
E(\epsilon_t\epsilon_{t-2}) &= \rho E(\epsilon_{t-2}\epsilon_{t-1})+E(\epsilon_{t-2}\eta_t) \\
E(\epsilon_t\epsilon_{t-2}) &= \rho^2\sigma^2
\end{aligned}$$

4. Así se puede derivar la siguiente expresión genérica:

$$E(\epsilon_t\cdot\epsilon_{t-(T-1)}) = \rho^{T-1}\sigma^2$$

Matriz de Varianzas y Covarianzas cuando ϵ_t es un AR(1)

5. Entonces:

$$E(\epsilon\epsilon_t) = \sigma^2\Omega = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}$$

La autocorrelación en el término de error puede ser producida por varias causas:

- *Existencia de ciclos y tendencias*: Si la autocorrelación es positiva (ρ), un valor alto de ϵ_t que genera un valor de y_t por sobre su media condicional, tendrá una probabilidad elevada de ir seguido por un valor alto de ϵ_{t+1}, y por ello, de un valor de y_{t+1} por encima del promedio; lo mismo ocurría para y_t debajo del promedio.
- *Variables omitidas*: Omisión tanto de variables relevantes, de no linealidades y de relaciones dinámicas (rezagos de la variable dependiente) serán incorporadas al término de error, causando posible autocorrelación.

Como vimos de las secciones anteriores, bajo autocorrelación MCO sigue siendo insesgado, pero pierde eficiencia, por lo cual ya no es MELI. El estimador de mínima varianza en este contexto es MCG, y en caso de desconocerse la forma de la autocorrelación se debe utilizar MCF.

Sin embargo y siguiendo el espíritu de la corrección de White, Newey y West (1987) propusieron una corrección para la matriz de varianzas y covarianzas de MCO.

Recordemos que en este contexto se cumple que:

$$V(\hat{\beta}_{MCO}|X) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

mientras que el estimador de Newey-West corresponde a:

$$V(\hat{\beta}_{MCO}|X) = n\sigma^2(X'X)^{-1}S(X'X)^{-1}$$

donde el estimador consistente de S es:

$$S = \frac{1}{n}\sum_{j=0}^{L}\sum_{t=j+1}^{n} w_j\hat{\epsilon}_j\hat{\epsilon}_{t-j}[x_t x'_{t-j} + x_{t-j}x'_t]$$

con:

$$w_j = 1 - \frac{j}{L+1}$$

L corresponde al orden máximo de autocorrelación del término de error (que no siempre es fácil determinar).

1. Test de Durbin-Watson (d): Lejos el test más utilizado para detectar autocorrelación de los residuos es el test propuesto en 1951 por Durbin y G.S Watson.

El test está diseñado para detectar autocorrelación en los residuos de la forma $\epsilon_t = \rho\epsilon_{t-1} + \eta_t$, donde η_t es ruido blanco. La nula es no autocorrelación de los residuos $(H_0 : \rho = 0 \quad H_1 : \rho \neq 0)$ y el test se define como:

$$d = \frac{\sum_{t=2}^{n}(\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^{n}\hat{\epsilon}_t^2}$$

Si $\rho > 0$, los valores de $\hat{\epsilon}_t$ probablemente serán muy cercanos, por lo cual el denominador será muy pequeño en comparación al residuo mismo. Ello implica que d será pequeño.

Si $\rho < 0$, entonces el denominador probablemente será grande, más grande que el residuos en si mismo. Ello implica que d será grande.

Se puede demostrar que para muestra grandes d converge a:

$$d \simeq 2(1 - \hat{\rho})$$

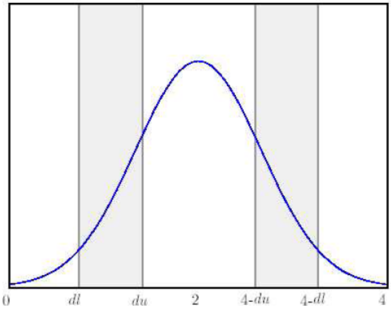
con:

$$\hat{\rho} = \frac{\sum_{t=2}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-1}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2}$$

donde ρ puede ser obtenido de la siguiente regresión:

$$\hat{\epsilon}_t = \rho \hat{\epsilon}_{t-1} + \eta_t$$

Respecto a los valores críticos del test, la distribución en muestras finitas depende del supuesto de normalidad de los errores y de la matriz X, por lo cual Durbin y Watson derivaron las tablas de valores de críticos para facilitar la aplicación del test.



Sin embargo, dichos valores poseen rangos indeterminados, en los cuales no podemos tomar una decisión respecto a la nula.

Por ejemplo, el test rechaza la nula de autocorrelación en favor de la alternativa de correlación positiva si $DW < dl$ y lo rechaza ante la alternativa de correlación negativa de los errores si $DW > 4 - dl$. El test posee dos zonas grises que se presentan en los intervalos (dl, du) y $(4 - du, 4 - dl)$, en las cuales no podemos decir nada respecto de la nula. Finalmente, si DW cae dentro del intervalo $(du, 4 - du)$ no se rechaza la nula de no autocorrelación.

El test posee dos grandes omisiones.

1. Sólo sirve para detectar autocorrelación de orden 1 en los errores y segundo, no puede ser aplicado si se incluyen regresores de la variable dependiente en el modelo (porque se construye bajo el supuesto de regresores determinísticos).
2. Además, se debe tener presente que el test está construido bajo normalidad de los errores y que existen las zonas grises o indeterminadas de las que hablábamos con anterioridad.

2. Test de h-Durbin (h): Una variación del test DW puede ser aplicada cuando existen variables rezagadas de la variable dependiente en nuestro modelo. Esta variación se conoce como test de h-Durbin. El estadígrafo es:

$$h = \left(1 - \frac{DW}{2}\right)\sqrt{\frac{n}{1 - n\hat{\sigma}^2_{\hat{\alpha}}}} \sim^{a} N(0,1)$$

donde $\hat{\sigma}^2_{\hat{\alpha}}$ a la varianza del parámetro asociado al primer rezago de la variable dependiente incluido en el modelo.

Algunas notas respecto al test.

- No importa cuantos rezagos de Y se hallan incluido en el modelo: sólo nos interesa la varianza del primero de ellos.
- El test no es aplicable cuando $n\hat{\sigma}^2_{\hat{\alpha}} > 1$ y tercero, las propiedades del test sólo son conocidas asintóticamente, por lo cual debe ser implementado con cuidado en muestras pequeñas.

3. Test de Breusch y Godfrey: Este test es una alternativa para testear autocorrelaciones de ordenes superiores a 1 y se basa en el test LM.

La nula, al igual que en todos los test de autocorrelación es que los residuos no se encuentran correlacionados.

Consideremos para distintos valores de k, el siguiente conjunto de estadísticos:

$$r_k = \frac{\sum_{t=1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-k}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2}$$

note que si $k = 1$, entonces estamos en una caso parecido al estadístico DW.

Los pasos para realizar el test son:

1. Estimar el modelo por MCO y obtener los residuos $\hat{\epsilon}$. El modelo puede incluir rezagos de la variable dependiente.
2. Estimar una regresión auxiliar de $\hat{\epsilon}_t$ sobre p rezagos: $\hat{\epsilon}_t,\ldots,\hat{\epsilon}_{t-p}$ incluyendo las variables exógenas (X) del modelo original. Note que deberá excluir p observaciones.
3. Calcular el R^2 de la regresión auxiliar.
4. Construir el estadígrafo $nR^2 \sim \chi^2_p$

Si no existe autocorrelación, entonces los residuos MCO no deberían ser explicados por sus retardos, por lo cual el R^2 de la regresión auxiliar debería ser cercano a cero, lo cual nos llevaría a un bajo valor del estadígrafo y a un no rechazo de la nula.

4. Test de Box-Pierce-Ljung (Q-Stat): Este test se basa en el cuadrado de las primeras p autocorrelaciones de los residuos MCO. El estadígrafo se define como:

$$Q = n \sum_{j=1}^{p} r_j^2$$

donde:

$$r_j = \frac{\sum_{t=j+1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-j}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2}$$

La distribución del estadígrafo bajo la nula de no autocorrelación es χ^2 con grados de libertad igual a p menos el número de rezagos del error incluidos en la especificación autorregresiva del error.

De ello se deduce que el test permite detectar autocorrelación de ordenes superiores a 1.

Como vimos anteriormente la matriz Ω en presencia de autocorrelación es:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \dots & 1 \end{pmatrix}$$

Se puede demostrar que la matriz P en este caso es:

$$P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

Entonces utilizando esta matriz P podemos transformar el modelo y aplicar Mínimos Cuadrados Generalizados. Al premultiplicar X e Y por la matriz P tendremos que la primera observación se transforma de la siguiente forma:

$$\sqrt{1-\rho^2}y_1 = (\sqrt{1-\rho^2})x_1'\beta + (\sqrt{1-\rho^2})\epsilon_1$$

Y para el resto de las $(T-1)$ observaciones la transformación es la siguiente:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + \underbrace{\epsilon_t - \rho\epsilon_{t-1}}_{\eta_t} \qquad (\dagger)$$

El que la primera observación de la muestra tenga un trato especial, es porque para ella no existe una observación anterior, y por lo tanto, es imposible aplicar la transformación anterior.

Estimación MCF: El Método de Cochrane Orcutt:
La matriz P que transforma nuestro modelo en un libre de autocorrelación en el error, es tal que cada observación de las variables dependientes, explicativas y término de error, se debe transformar de acuerdo a (†). Si es que nuestro modelo es el siguiente:

$$\begin{aligned} y_t &= x_t\beta + \epsilon_t \\ \epsilon_t &= \rho\epsilon_{t-1} + \eta_t \end{aligned}$$

El modelo transformado es de la siguiente forma:

$$\begin{aligned} \underbrace{y_t - \rho y_{t-1}}_{y_t^*} &= \underbrace{(x_t - \rho x_{t-1})}_{x_t^*}\beta + \underbrace{\epsilon_t - \rho\epsilon_{t-1}}_{\eta_t} \\ \Rightarrow y_t^* &= x_t^*\beta + \eta_t \end{aligned}$$

El Método de Cochrane-Orcutt es un procedimiento iterativo para obtener la estimación de β y ρ:

1. Estimar por MCO la regresión de interés, ignorando la presencia (conocida) de autocorrelación de primer orden en el término de error.
2. Utilizar los residuos MCO para estimar el parámetro ρ. Esto puede hacerse mediante una regresión de $\hat{\epsilon}_t$ contra $\hat{\epsilon}_{t-1}$, o a partir del estadístico DW de la estimación anterior.
3. Utilizar este parámetro $\hat{\rho}$ para transformar las variables, y obtener y_t^* y x_t^*.
4. Estimar por MCO un modelo con las variables transformadas, para obtener un nuevo vector de coeficientes β.
5. Utilizar esta nueva estimación para computar otro vector de residuos, y utilizar estos residuos para obtener una nuevaestimación de ρ.
6. Repetir este procedimiento hasta que los β convergan.