

Econometría II. 3^a LADE
Tema 4. Autocorrelación.

1 Introducción

En el tema precedente hemos estudiado las propiedades de los estimadores mínimo cuadrático ordinarios en el caso de que la matriz de varianzas y covarianzas del vector de perturbaciones sea no escalar. Esto puede deberse tanto a que los valores de la diagonal principal de esta matriz sean distintos entre sí o a que los valores fuera de la diagonal principal sean distintos de 0. En el primer caso tenemos problemas de heteroscedasticidad, mientras que en el segundo tenemos problemas de autocorrelación. Dado que las causas que pueden producir cada uno de estos dos problemas son distintas entre sí, parece adecuado estudiarlos de forma separada. En este capítulo nos centraremos en el caso de problemas de autocorrelación, mientras que en el siguiente estudiaremos los problemas de heteroscedasticidad.

El esquema que vamos a seguir en estos dos capítulos es similar. En primer lugar debemos estudiar los efectos que se derivan de la presencia de, bien, autocorrelación o, bien, heteroscedasticidad sobre la estimación mco. En segundo lugar, debemos determinar métodos que nos permitan detectar estos problemas. Por último, hay que conocer procedimientos que nos permitan solucionar estos problemas.

2 Efectos sobre la estimación mco

Dado un modelo lineal general definido como:

$$y = X\beta + u$$

con $E(u) = 0$, consideramos que se incumple el supuesto de no autocorrelación siempre que $Cov(u_i, u_j) \neq 0$, para al menos una pareja de valores i, j , donde $i \neq j$. Como consecuencia de esto, la matriz de varianzas y covarianzas del vector de perturbaciones u no es escalar, por cuanto no todos los valores fuera de la diagonal principal son distintos de 0. En el capítulo anterior hemos estudiado cuáles son los efectos que tiene la quiebra del supuesto de no escalaridad de la matriz $Var(u)$, por lo que no es necesario repetirlos aquí.

Los resultados del capítulo anterior se cimientan sobre la base de que $E(u) = 0$, por lo que la no escalaridad de la matriz de varianzas y covarianzas está causada por las perversas propiedades del vector de perturbaciones u , al tiempo que la especificación del modelo coincide con la del proceso generador de los datos. Sin embargo, esta no es la única explicación posible del fenómeno de autocorrelación. Existe otra, radicalmente distinta de la anterior, que pone el acento en una inadecuada especificación del modelo. Para comprender las diferencias entre ambos puntos de vista, vamos a analizar el siguiente caso. Supongamos que el proceso generador de los datos viene dado por la siguiente relación:

$$PGD : \quad y_t = \beta_1 + \beta_1 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + v_t$$

donde, y_t son las ventas de un producto determinado, x_{2t} es la renta de la región donde se comercializa ese bien, x_{3t} es el precio de este bien, x_{4t} es un índice de precios relacionado con los bienes que compiten con el que estamos estudiando y v_t es una perturbación aleatoria que sigue una distribución $niid(0, \sigma_v^2)$.

Supongamos ahora que el investigador encargado del análisis de la variable plantea una especificación que no coincide con el proceso generador de los datos, por cuanto omite la variable x_{4t} , por ejemplo. El modelo empírico, por tanto, queda definido en los siguientes términos:

$$ME : \quad y_t = \beta_1 + \beta_1 x_{2t} + \beta_3 x_{3t} + e_t \quad (1)$$

donde e_t es la perturbación del modelo empírico. Es evidente, que dado que la especificación empírica no coincide con el proceso generador de los datos, la perturbación del modelo empírico contiene no solo un componente aleatorio, sino que además incluye el componente determinístico omitido. Así, resulta que $e_t = \beta_4 x_{4t} + v_t$. Esta variable aleatoria tiene las siguiente propiedades:

a) La esperanza de la perturbación del modelo empírico es ahora:

$$E(e_t) = E(\beta_4 x_{4t} + v_t) = \beta_4 E(x_{4t}) + E(v_t) = \beta_4 E(x_{4t})$$

Por tanto, este primer resultado nos indica que la esperanza de la perturbación del modelo empírico puede ser distinta de 0, lo que nos está indicando la presencia de un error en la especificación del modelo. En el caso de que la variable x_4 sea de tipo determinístico, tal y como mantiene

uno de los supuestos básicos del modelo lineal general, entonces el resultado anterior debería tener en cuenta que $E(x_{4t}) = x_{4t}$. En cualquier caso, al ser, en general, $E(e_t) \neq 0$, la conclusión a la que llegamos es que la estimación mco del modelo (1) está sesgada.

b) La varianza de la perturbación del modelo empírico es:

$$\begin{aligned} Var(e_t) &= Var(\beta_4 x_{4t} + v_t) = Var(\beta_4 x_{4t}) + Var(v_t) + 2 Cov(\beta_4 x_{4t}, v_t) \\ &= \beta_4^2 Var(x_{4t}) + \sigma_v^2 \end{aligned}$$

donde hemos incluido el supuesto de no relación entre las variables explicativas y la perturbación del proceso generador de los datos, lo que implica que $Cov(\beta_4 x_{4t}, v_t) = 0$.

c) La covarianza de orden i-esimo de la perturbación del modelo empírico es:

$$\begin{aligned} Cov(e_t, e_{t-i}) &= E\{\{\beta_4 [x_{4t} - E(x_{4t})] + v_t\} \{\beta_4 [x_{4t-i} - E(x_{4t-i})] + v_{t-i}\}\} \\ &= \beta_4^2 Cov(x_{4t}, x_{4t-i}) + Cov(v_t, v_{t-i}) + \beta_4 Cov(x_{4t}, v_{t-i}) \\ &\quad + \beta_4 Cov(x_{4t-i}, v_t) \end{aligned}$$

En virtud de este resultado vemos que la covarianza entre e_t y e_{t-i} depende de la covarianza de orden i-ésimo de la variable x_4 . Si resulta que $Cov(x_{4t}, x_{4t-i}) = 0$, tenemos que $Cov(e_t, e_{t-i}) = 0$. Por tanto, si x_4 es un regresor determinista o es estocástico pero incorrelado, la omisión de una variable relevante no causa autocorrelación. Entonces, si el estudio utiliza datos de corte transversal, es factible admitir la no correlación entre los valores de las variables. No ocurre lo mismo cuando disponemos de datos de series temporales, ya que las variables macroeconómicas que se utilizan habitualmente en estos estudios exhiben un mayor o menor grado de correlación. Por ejemplo, ¿seríamos capaces de admitir que los valores actuales del PIB no están relacionados con los valores del periodo anterior?

Si asumimos que las variables explicativas están correlacionadas entre sí tiene dos implicaciones. En primer lugar, el supuesto de determinidad de los regresores no es útil en este escenario. Ya sabemos que estas variables de series temporales las podemos representar a partir de procesos ARMA(p,q), por lo que admitimos su carácter estocástico. Además, en segundo lugar, si ocurre que $Cov(x_{4t}, x_{4t-i}) \neq 0$ para algún $i \neq 0$, resulta que $Cov(e_t, e_{t-i}) \neq 0$.

Por tanto, desde este punto de vista, un error en la especificación del modelo empírico puede causar problemas de autocorrelación. Desde este punto de vista, la presencia de autocorrelación debe interpretarse como un indicativo de que la especificación anterior es incorrecta, ya que no es compatible con el supuesto de no correlación de las perturbaciones.

Debemos señalar que la omisión de variables relevantes no es el único tipo de mala especificación que conlleva que las perturbaciones del modelo estén correlacionadas entre sí. Aunque no van a ser estudiados aquí, también la existencia de parámetros que cambian a lo largo del tiempo o la incorrecta selección de la forma funcional, entre otras razones, pueden suponer la quiebra del supuesto de no correlación

Dado que ya conocemos los daños que produce sobre la estimación mco la presencia de problemas de autocorrelación, el siguiente paso es el de diseñar métodos que nos permitan determinar cuando existe evidencia en contra del supuesto de no autocorrelación. En la siguiente sección, se presentan diversos estadísticos. El siguiente paso es saber como se solucionan los problemas de autocorrelación, aspecto que estudiaremos en la tercera sección del tema.

3 Contrastes de autocorrelación

En esta sección vamos a proporcionar diversos métodos que nos permitan detectar la presencia de autocorrelación en los residuos del modelo estimado. El primer problema al que nos debemos enfrentar es determinar el tipo de autocorrelación frente al que nos encontramos. Es decir, debemos caracterizar la forma que adopta la correlación serial de las perturbaciones del modelo. Una primera aproximación, muy utilizada en la práctica, es suponer que la perturbación del modelo siguen un proceso autorregresivo de primer orden. Sin embargo, es posible que la relación entre las perturbaciones sea de un orden distinto, por lo que se hace necesario considerar casos alternativos. En lo que sigue vamos a estudiar diversos estadísticos que permiten contrastar la existencia distintos tipos de autocorrelación.

3.1 AR(1)

En esta sección vamos a considerar el caso más común, aquel en el que existe correlación de primer orden entre las perturbaciones del modelo. El planteamiento que vamos a seguir parte de suponer que queremos estudiar

la presencia de autocorrelación en el siguiente modelo:

$$y = X\beta + u$$

donde $\hat{\beta}$ representa el estimador mco y $\hat{u} = y - X\hat{\beta}$. A lo largo de esta sección, la hipótesis que vamos a formular que las perturbaciones del modelo anterior pueden definirse de acuerdo al siguiente modelo:

$$u_t = \rho u_{t-1} + e_t \quad t = 2, 3, \dots, T$$

donde e_t es un ruido blanco. Si en el modelo anterior $\rho = 0$, entonces u_t y u_{t-1} no están correlacionados, por lo que deberíamos concluir que no existe autocorrelación de primer orden. Por el contrario, si $\rho \neq 0$ debemos concluir que la relación entre estas dos variables existe y, por tanto, las perturbaciones del modelo presentan autocorrelación. Para contrastar la hipótesis nula $H_o : \rho = 0$ disponemos de diversos estadísticos. De entre ellos, vamos a estudiar aquí a dos, el estadístico de Durbin-Watson y la h de Durbin.

3.1.1 Durbin-Watson

El estadístico comúnmente utilizado para contrastar esta hipótesis $H_o : \rho = 0$ es el propuesto en Durbin y Watson (1950). Este estadístico se define de la siguiente manera.

$$DW = \frac{\sum_{i=2}^T (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{i=1}^T \hat{u}_t^2}$$

donde $\hat{u}_t$ es el residuo mínimo cuadrático ordinario del periodo t , con $t = 1, 2, \dots, T$. Para entender la intuición que tiene el uso de estadístico, debemos tener en cuenta lo siguiente. Con una serie de sencillas operaciones, el estadístico DW se puede expresar en los siguiente términos:

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

Si tenemos en cuenta que $\sum_{i=1}^T \hat{u}_t^2 = \sum_{i=2}^T \hat{u}_t^2 + \hat{u}_1^2$ y que $\sum_{i=1}^T \hat{u}_t^2 = \sum_{i=2}^T \hat{u}_{t-1}^2 + \hat{u}_T^2$ entonces es directo comprobar que los dos primeros sumandos de la expresión anterior se aproximan a la unidad, conforme aumenta el número de observaciones. Por tanto, se cumple que:

$$DW \simeq 1 + 1 + 2 \frac{\sum_{i=2}^T \hat{u}_t \hat{u}_{t-1}}{\sum_{i=1}^T \hat{u}_t^2} = 2 + 2\hat{\rho} = 2(1 - \hat{\rho})$$

donde $\hat{\rho}$ es un estimador consistente del coeficiente de correlación de primer orden de los residuos mco, coeficiente que toma valores en el intervalo $(-1, 1)$. Entonces, si no existe autocorrelación entre u_t y u_{t-1} , es de esperar que el valor del estimador del coeficiente de autocorrelación de primer orden se aproxime hacia 0 y, en consecuencia, el valor del estadístico Durbin-Watson tome un valor en un entorno cercano a 2. Por el contrario, la existencia de autocorrelación positiva entre u_t y u_{t-1} supone que $\hat{\rho}$ tome valores positivos, lo que hará que el estadístico de Durbin-Watson se aleje del valor 2 y se aproxime hacia 0 a medida que el grado de correlación sea mayor. De forma similar, si existe correlación negativa entre u_t y u_{t-1} , el estadístico de Durbin-Watson se aleja de 0, aproximándose hacia 4 conforme el valor del coeficiente de autocorrelación se acerque hacia -1 .

Como vemos, el estadístico de Durbin-Watson toma valores dentro del intervalo $(0, 4)$. Si planteamos como hipótesis nula del contraste la no existencia de autocorrelación, entonces $H_o : \rho = 0$. Esta hipótesis nula se aceptará siempre y cuando el valor muestral del estadístico sea

estadísticamente igual a 2. En principio, esto sería relativamente sencillo, por cuanto tan solo deberíamos definir un intervalo de confianza para este estadístico y comprobar si el valor 2 se encuentra dentro de este intervalo. El problema que aparece es que el estadístico de Durbin-Watson no sigue ninguna distribución conocida. Este extremo se puede comprobar sencillamente, sin más que redefinir el estadístico en los siguientes términos:

$$DW = \frac{\hat{u}' A \hat{u}}{\hat{u}' \hat{u}}$$

donde $\hat{u}$ es el vector de residuos mínimo cuadrático ordinarios y la matriz A adopta la siguiente forma:

$$A = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\ & & & \cdots & & & & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

Si, adicionalmente, tenemos en cuenta que $\hat{u} = M u$, el estadístico Durbin-Watson se puede expresar como cociente de dos formas cuadráticas del vector de residuos mínimo cuadrático ordinario:

$$DW = \frac{\hat{u}' A \hat{u}}{\hat{u}' \hat{u}} = \frac{u' M A M u}{u' M u} = \frac{u' L u}{u' M u}$$

donde $L = M A M$. La expresión anterior no es sino un cociente de dos formas cuadráticas del vector u por lo que podríamos pensar que el estadístico DW seguiría una distribución del tipo F de Snedecor. Sin embargo, esto no es posible, ya que el numerador de la expresión anterior no sigue una distribución χ^2 . Además, aunque la siguiera, nada nos garantiza la independencia entre el numerador y el denominador del anterior cociente. A pesar de todo esto, dado que conocemos la distribución del estadístico Durbin-Watson podríamos intentar tabularla para distintos valores críticos, pero esta opción no está libre de problemas. Para verlo, podemos expresar el estadístico Durbin-Watson de la siguiente manera:

$$DW = \frac{\sum_{i=1}^{T-k} \theta_i \nu_i^2}{\sum_{i=1}^{T-k} \nu_i^2}$$

donde θ_i son las raíces características de la matriz MA que son distintas de 0. En la expresión anterior comprobamos que el estadístico de Durbin-Watson se puede interpretar como un cociente de sumas de cuadrados de $T - k$ variables aleatorias que siguen una distribución normal. En el numerador, esta suma está ponderada por los coeficientes θ_i ($i = 1, 2, \dots, T - k$). Entonces, la función de probabilidad se puede derivar de forma numérica para un conjunto de θ_i dados. Desafortunadamente, estos parámetros dependen de los regresores incluidos en el modelo estimado, por lo que no se puede obtener la función de distribución con carácter general. Una posible solución sería calcular los valores críticos particulares para cada una de las regresiones empleadas. En la actualidad, este método es relativamente poco costoso dado los avances de la informática. Sin embargo, en los años 50 y 60, esta solución requería de un coste de tiempo ciertamente elevado lo que la hizo inviable. La solución parcial que se adoptó es determinar dos estadísticos d_L y d_U ($d_L < d_U$) que no dependan de la formación de la matriz X y que sean de tal forma que, considerando el caso de autocorrelación positiva, $DW < d_L$ suponga que la hipótesis nula no es cierta, mientras que $DW > d_U$ implique el cumplimiento de dicha hipótesis. De forma simétrica, si consideramos la presencia de autocorrelación negativa, $DW > 4 - d_U$ implica la aceptación de la hipótesis nula, mientras que $DW < 4 - d_L$ supondrá el rechazo de dicha hipótesis.

Esto implica que si el valor de DW se encuentra en un entorno cercano a 2 aceptaremos la hipótesis nula y, por tanto, consideramos que no existe autocorrelación entre los residuos del modelo. Por el contrario, valores del estadístico alejados de 2 conllevan el rechazo de la hipótesis nula bien sea a favor de la alternativa de autocorrelación positiva o bien a favor de la alternativa de autocorrelación negativa. Además, existen dos zonas en las que el contraste es inconclusivo, es decir, que puede suponer tanto la aceptación como el rechazo de la hipótesis. Si el valor de DW pertenece a cualquiera de estas dos zonas, no podemos decir nada acerca de la presencia de autocorrelación en el modelo.

Como se puede entender, esto es una deficiencia importante del contraste

Durbin-Watson. No es la única. Otro de los problemas que presenta su uso es que los valores d_L y d_U están calculados bajo el supuesto de que la matriz X es de tipo determinista. Si esto no se cumple, los valores anteriores no se pueden usar. Este problema se agrava cuando uno de los regresores es un retardo de la variable endógena. Nerlove y Wallis demuestran que en este tipo de circunstancias el estadístico DW está sesgado hacia 2, independientemente de las características de la perturbación aleatoria del modelo. Por tanto, aunque el modelo presente problemas de autocorrelación, el estadístico DW será incapaz de detectar su presencia.

3.1.2 h-Durbin

Una posible manera de solventar este tipo de problemas es utilizar el estadístico h-Durbin. Supongamos que estimamos un modelo que contiene como regresores una serie de retardos de la variable endógena. En concreto, y adoptando una visión general del problema, supongamos que estimamos estudiar el siguiente modelo:

$$y_t = \beta_1 + \sum_{i=2}^k \beta_i x_{it} + \sum_{i=1}^p \gamma_i y_{t-i} + e_t$$

El estadístico h-Durbin se define de la siguiente manera:

$$h = \hat{\rho} \sqrt{\frac{T}{1 - T \widehat{Var}(\hat{\gamma}_1)}} \stackrel{as}{\sim} N(0, 1)$$

donde $\hat{\rho}$ es estimador mínimo cuadrático ordinario del coeficiente de autocorrelación de primer orden, T es el tamaño muestral utilizado y $\widehat{Var}(\hat{\gamma}_1)$ es la varianza estimada del estimador del parámetro asociado al primer retardo de la variable endógena. Bajo la hipótesis nula de no autocorrelación ($H_0 : \rho = 0$) el estadístico h se distribuye asintóticamente como una $N(0, 1)$. Dado que la potencia del estadístico no es elevada, es habitual utilizar el contraste con una sola de probabilidad. Así, cuando resulte que si observamos que $\hat{\rho}$, deberemos comparar el valor del estadístico h con 1.65, valor crítico de una distribución normal para un nivel de significación del 5%. Si $|h| > 1.65$, entonces rechazaremos la hipótesis nula en favor de la alternativa de autocorrelación de primer orden. En otro caso, aceptaremos la hipótesis nula.

3.2 Wallis

Un contraste que está muy relacionado con los anteriores es el contraste de Wallis. Imaginemos que nuestra base de datos contiene información trimestral de las variables que vamos a utilizar en nuestra especificación. Al usar datos de frecuencia inferior al año podemos observar la presencia de un componente estacional en las variables. Si dicho componente no está correctamente reflejado en la especificación del modelo, es posible que la perturbación del modelo siga el siguiente modelo:

$$u_t = \rho_4 u_{t-4} + e_t \quad t = 5, 6, \dots, T$$

Entonces, resulta que el coeficiente de correlación de primer orden es 0, lo que hace que el estadístico DW tiende a aceptar la hipótesis nula, cuando en realidad existe autocorrelación entre las perturbaciones del modelo. Para solucionar este problema, Wallis (1972) propuso un nuevo estadístico que no es sino una pequeña modificación del contraste de Durbin-Watson. Este estadístico se define como sigue:

$$W = \frac{\sum_{i=5}^T (\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{i=1}^T \hat{u}_t^2}$$

De forma similar a lo que ocurre con el estadístico de Durbin-Watson, la hipótesis nula del contraste es ahora $H_0 : \rho_4 = 0$. Bajo esta hipótesis nula, la distribución del estadístico de Wallis es qualitativamente similar a la del Durbin-Watson, compartiendo sus problemas para determinar con exactitud sus valores críticos. De nuevo, Wallis acota esta distribución, determinando unos valores mínimos y máximos y una zona de indeterminación.

3.3 Breusch-Godfrey

Los estadísticos anteriores son válidos para contrastar la existencia de un tipo específico de autocorrelación. Si queremos contrastar la presencia de autocorrelación, pero desconocemos el tipo concreto de autocorrelación que pueden seguir las perturbaciones o, simplemente, no queremos formular un único tipo de autocorrelación podemos utilizar el método propuesto casi al mismo tiempo en Breusch (1978) y Godfrey (1978), quienes posteriormente

realizaron el trabajo conjunto Breusch y Godfrey (1979). Su planteamiento es más amplio que los estadísticos pertenecientes a la familia de Durbin-Watson, permitiendo que las perturbaciones puedan ser generadas por modelos más generales. El estadístico es un multiplicador de Lagrange que se calcula de la siguiente manera.

El modelo que queremos estudiar es el siguiente:

$$y = X\beta + u$$

donde la matriz X puede contener retardos de la variable endógena. A la hora de formular la hipótesis nula del contraste resulta importante determinar el proceso que puede seguir el vector de perturbaciones aleatorias. En general, vamos a considerar que la perturbación del periodo t puede seguir un esquema autorregresivo de orden p

$$\begin{aligned} u_t &= \varphi_1 u_{t-1} + \dots + \varphi_p u_{t-p} + \varepsilon_t \\ \varepsilon_t &\sim i.i.d.N(0, \sigma^2). \end{aligned}$$

Bajo la hipótesis nula, $H_0 : \varphi_1 = \varphi_2 = \dots = \varphi_p = 0$, las perturbaciones están incorrelacionadas entre sí.

Si ignoramos las p primeras observaciones, el modelo se puede escribir de la siguiente manera:

$$y_t = x'_t \beta + \sum_{j=1}^p \varphi_j (y_{t-j} - x'_{t-j} \beta) + \varepsilon_t.$$

Como vemos, este modelo es no lineal en β y en φ , lo que lo podemos expresar de la siguiente forma:

$$y_t = f_t(\beta, \varphi) + \varepsilon_t.$$

Ahora, si tomamos una expansión de Taylor de primer orden alrededor de $\varphi = 0$ evaluada en $\tilde{\beta}$, tenemos que:

$$y_t \approx f_t(\tilde{\beta}, 0) + \sum_{j=1}^{k+p} \left(\frac{\partial f_t(\beta, \varphi)}{\partial \theta_j} \right)_{|\tilde{\beta}, 0} (\theta_j - \tilde{\theta}_j) + \varepsilon_t$$

donde $\theta' = (\beta', \varphi')$. Si denominamos,

$$\begin{aligned} \left(\frac{\partial f_t}{\partial \beta_j} \right)_{|\tilde{\beta}, 0} &= x_{tj} \quad (j = 1, \dots, k) \\ \left(\frac{\partial f}{\partial \varphi_j} \right)_{|\tilde{\beta}, 0} &= y_{t-j} - x'_{t-j} \tilde{\beta} = \tilde{u}_{t-j} \quad (j = 1, \dots, p). \end{aligned}$$

Entonces,

$$\underbrace{y_t - f_t(\tilde{\beta}, 0)}_{\tilde{u}_t} = \sum_{j=1}^k x_{tj} (\beta_j - \tilde{\beta}_j) + \sum_{j=1}^p \tilde{u}_{t-j} \varphi_j + \varepsilon_t .$$

En forma matricial, se puede expresar de la siguiente manera:

$$\begin{aligned} \underset{(T-p) \times 1}{y - f} &= \underset{(T-p) \times (k+p)}{\begin{bmatrix} X \tilde{u} \end{bmatrix}} \underset{p}{\begin{bmatrix} \beta - \tilde{\beta} \\ \dots \\ \varphi \end{bmatrix}}^k + \varepsilon \\ &= F\theta + \varepsilon. \end{aligned} \tag{2}$$

con

$$\underset{(T-p) \times p}{\tilde{u}} = [\tilde{u}_{-1}, \dots, \tilde{u}_{-p}]$$

Para un proceso AR(p), la función log-verosímil se puede expresar así:

$$\mathcal{L} = const - \left(\frac{T-p}{2} \right) \ln(\sigma^2) - \left(\frac{1}{2\sigma^2} \right) S(\varphi, \beta)$$

y, concentrando esta función sobre σ^2 :

$$\mathcal{L}' = const' - \left(\frac{T-p}{2} \right) \ln[S(\varphi, \beta)]$$

donde

$$S(\varphi, \beta) = \varepsilon' \varepsilon \approx [(y - f) - F\theta]' [(y - f) - F\theta] .$$

A partir de este resultado, tenemos que:

$$\begin{aligned} \frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \theta} &= \left(\frac{1}{2\tilde{\sigma}^2} \right) (2F'(y - f)) \\ &= \frac{1}{\tilde{\sigma}^2} F'(y - f) . \end{aligned}$$

La matriz de varianzas y covarianzas de $\tilde{\theta}$ viene dada por

$$V(\tilde{\theta}) = \tilde{\sigma}^2 (F'F)^{-1} .$$

En consecuencia, si formamos el estadístico de multiplicadores de Lagrange, resulta que:

$$\begin{aligned}
LM &= \left(\frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \theta} \right)' V(\tilde{\theta}) \left(\frac{\partial \mathcal{L}(\tilde{\theta})}{\partial \theta} \right) \\
&= \frac{1}{\tilde{\sigma}^2} (y - f)' F \left[\tilde{\sigma}^2 (F' F)^{-1} \right] F' (y - f) \frac{1}{\tilde{\sigma}^2} \\
&= \frac{1}{\tilde{\sigma}^2} \left[(y - f)' F (F' F)^{-1} F' (y - f) \right] \\
&\rightarrow {}^d \chi_p^2 \text{ under } H_0.
\end{aligned}$$

Debemos tener en cuenta que $\tilde{\sigma}^2$ es la estimación del parámetro de dispersión bajo la hipótesis nula, por lo que coincide con $\frac{SR}{T-p}$. entonces,

$$\tilde{\sigma}^2 = (y - f)'(y - f)/(T - p) = TSS / (T - p)$$

donde TSS es la suma total resultante de la regresión (2). Además,

$$\begin{aligned}
&\left[(y - f)' F (F' F)^{-1} F' (y - f) \right] \\
&= (y - f)' P_F (y - f) \\
&= ESS
\end{aligned}$$

es la suma explicada de esta misma regresión (2). Por tanto,

$$\begin{aligned}
LM &= [ESS / TSS] (T - p) \\
&= (T - p) R^2
\end{aligned}$$

donde R^2 es el coeficiente de determinación calculado en (2).

Aunque aparentemente complicado, el cálculo de este estadístico es bastante sencillo en la práctica. Para su obtención, debemos seguir los siguientes pasos:

1. Estimar el modelo original por MCO y obtener el vector de residuos
2. Estimar por MCO la siguiente regresión auxiliar:

$$\hat{u} = X \eta + \hat{u}_p \delta + v$$

donde

$$\hat{u}_p = (\hat{u}_{-1}, \hat{u}_{-2}, \dots, \hat{u}_{-p})$$

siendo $\hat{u}_{-j}$ el vector que contiene la información del j-ésimo retardo de los residuos.

3. Obtener el coeficiente de determinación R^2 de la regresión auxiliar. Entonces, el estadístico de Breusch-Godfrey es igual a:

$$LM = (T - p) R^2 \rightarrow^d \chi_p^2 \text{ under } H_0.$$

Bajo la hipótesis nula, es de esperar que este coeficiente de determinación se aproxime hacia 0, ya que la matriz X es ortogonal al vector de residuos y, si la hipótesis nula es cierta, la correlación entre u y u_p se aproxima a 0. Por el contrario, si existe autocorrelación, el coeficiente de determinación tenderá a tomar valores distintos de cero, más próximos a la unidad conforme el grado de autocorrelación sea mayor.

Por último, destacar que este estadístico es válido tanto si la perturbación sigue un proceso AR(p) o un MA(p), tal y como se demuestra en Godfrey (*Econometrica*, 78).

3.4 Análisis de los residuos

Un método alternativo a los que hemos visto con anterioridad es el análisis de los residuos de la estimación. El razonamiento intuitivo es el siguiente. Si el modelo estimado no presenta problemas de autocorrelación, los coeficientes de la función de autocorrelación de los residuos deben aproximarse hacia 0. Por el contrario, si existe autocorrelación, estos coeficientes serán distintos de cero.

No obstante, la mera inspección visual de la función de autocorrelación de los residuos no es suficiente para determinar la existencia de autocorrelación con un mínimo de garantías. Para ello, podemos hacer uso de los contrastes de significación global de los p primeros coeficientes de la función de autocorrelación. Este contraste se define de la siguiente manera:

$$Q = T \sum_{j=1}^m r_j^2 \rightarrow^d \chi_m^2$$

donde r_j es el coeficiente de correlación lineal muestral entre $\hat{u}_t$ y $\hat{u}_{t-j}$. Este coeficiente, se define:

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

Bajo la hipótesis nula, $H_o : \rho_1 = \rho_1 = \dots = \rho_M = 0$, donde p es el j -ésimo coeficiente de la función de autocorrelación de las perturbaciones, el estadístico de Box-Pierce se distribuye asintóticamente bajo una distribución χ_m^2 . Cuando la matriz de regresores contiene exclusivamente variables exógenas, el estadístico de Box-Pierce es equivalente al de Breusch-Godfrey.

El estadístico de Box-Pierce tiene problemas de falta de potencia. Para solventar este problema, Ljung y Box (1979) proponen un refinamiento del estadístico anterior:

$$LB = T(T+2) \sum_{j=1}^m \frac{r_j^2}{T-j} \rightarrow^d \chi_m^2$$

Asintóticamente este contraste es equivalente al de Box-Pierce, distribuyéndose también de acuerdo a una χ_m^2 .

El problema que tienen estos estadísticos es la elección del valor de p . En general, se suele tomar un valor m elevado para no omitir ningún tipo de autocorrelación. Por ejemplo, con datos anuales $m = 4$ puede ser apropiado, mientras que si los datos son trimestrales o mensuales, el uso de $m = 8$ y $m = 24$ resulta recomendable.

Finalmente, debemos destacar que los contrastes de Breusch-Godfrey y el de Box-Pierce presentan grandes similitudes. A tal punto que, bajo la hipótesis de que las variables explicativas no están correlacionadas con las perturbaciones, ambos estadísticos coinciden asintóticamente. Esto no quita para que sus propiedades en muestras finitas no coincidan.

4 Soluciones

Una vez detectado el problema debemos corregirlo de forma que los estimadores presenten propiedades óptimas. El problema al que nos enfrentamos es que, en primer lugar, debemos identificar cuál es el origen del problema de autocorrelación. Si admitimos que la especificación del modelo es correcta, pero el vector de perturbaciones aleatorias está correlacionado temporalmente, entonces la solución ha de venir vía uso de métodos como mínimos cuadrados generalizados o máxima verosimilitud. Por contra, si consideramos que la presencia de autocorrelación no es sino un reflejo de

una mala especificación, el método anterior no solucionará el problema. Al contrario, puede llevarnos a más confusión. Por contra, una nueva especificación del modelo, que corrija los problemas que dan origen a la aparición de los problemas, sí que es efectiva en este escenario. En lo que sigue discutimos cómo se pueden solucionar estos problemas.

4.1 Estimation with autocorrelation

El modelo considerado es

$$y = X\beta + u$$

donde asumimos que $Var(u) = \Omega$. Para estimar de forma eficiente este modelo hemos visto en el capítulo anterior que es necesario tener en cuenta la forma que adopta la matriz Ω . Esta matriz cambia con cada tipo de autocorrelación considerado, por lo que no se hace conveniente analizar de forma separada cada tipo de autocorrelación. Es lo que vamos a realizar en las siguientes secciones.

4.1.1 Caso AR(1)

Este es el caso que se considera mayoritariamente en las aplicaciones. Supongamos que:

$$\begin{aligned} u_t &= \rho u_{t-1} + e_t \\ e_t &\sim i.i.d. N(0, \sigma^2). \end{aligned}$$

Lo que vamos a realizar a continuación es determinar la forma que adopta la matriz Ω . Así, Multiplying by u_{t-h}

$$u_t u_{t-h} = \rho u_{t-1} u_{t-h} + e_t u_{t-h}$$

and taking expectations:

$$E[u_t u_{t-h}] = \rho E[u_{t-1} u_{t-h}]$$

or

$$\gamma_h = \rho \gamma_{h-1} \quad h = 1, 2, \dots \quad (3)$$

where $\gamma_h = E[u_t u_{t-h}]$ is the autocovariance at lag h . Now,

$$E[u_t - \rho u_{t-1}]^2 = E[e_t^2] = \sigma^2$$

implies

$$E[u_t^2 + \rho^2 u_{t-1}^2 - 2\rho u_t u_{t-1}] = \sigma^2$$

or

$$(1 + \rho^2) \gamma_0 - 2\rho \gamma_1 = \sigma^2. \quad (4)$$

Solving (3) and (4),

$$\gamma_h = \frac{\sigma^2 \rho^h}{1 - \rho^2} \quad h = 0, 1, 2, \dots$$

In matrix form (by symmetry) :

$$\Omega_{T \times T} = \frac{\sigma^2}{1 - \rho^2} \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \\ \vdots & & \ddots & \\ \rho^{T-1} & & & 1 \end{bmatrix}.$$

It can be shown that

$$\Omega^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} 1 & -\rho & & & \\ -\rho & 1 + \rho^2 & -\rho & & 0 \\ & -\rho & 1 + \rho^2 & -\rho & \\ & & \ddots & & \\ 0 & & & -\rho & 1 + \rho^2 & -\rho \\ & & & & -\rho & 1 + \rho^2 \end{bmatrix}$$

and the matrix Q such that $Q'Q = \Omega^{-1}$ is

$$Q_{T \times T} = \begin{bmatrix} (1 - \rho^2)^{1/2} & & & & \\ & -\rho & 1 & & 0 \\ & & -\rho & 1 & \\ & & & \ddots & \\ & & & & -\rho & 1 \\ 0 & & & & & -\rho & 1 \end{bmatrix}.$$

We only have one parameter to estimate to apply feasible GLS : ρ .

4.1.2 Feasible GLS.

Cochrane-Orcutt procedure. A natural estimator of ρ is:

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

where

$$\hat{u}_t = y_t - x_t' \hat{\beta}_{OLS}.$$

The estimate $\hat{\rho}$ is consistent and asymptotically normal, provided $\hat{\beta}_{OLS}$ is consistent and $|\rho| < 1$. But, as we saw, this is the case. So, feasible GLS is obtained by replacing $\hat{\rho}$ in Q and transforming the model

$$\begin{aligned} \hat{Q}y &= \hat{Q}X\beta + \hat{Q}u \\ \hat{\beta}_{FGLS} &= \left(X' \hat{\Omega}^{-1} X \right)^{-1} X' \hat{\Omega}^{-1} y. \end{aligned}$$

A simpler procedure is possible if the initial observation is dropped, i.e., use Q^*

$$Q^*_{(T-1) \times T} = \begin{bmatrix} -\rho & 1 & & & \\ & -\rho & 1 & & \\ & & \ddots & \ddots & \\ & & & -\rho & 1 \end{bmatrix}.$$

This is called the *Cochrane-Orcutt* procedure.

Durbin's regression procedure. Durbin's regression procedure uses another method to obtain a consistent estimate of ρ . Consider transforming the model as follows using quasi-differences:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})' \beta + (u_t - \rho u_{t-1})$$

or

$$y_t = \rho y_{t-1} + x_t' \beta - \rho x_{t-1}' \beta + e_t. \quad (5)$$

1. Run OLS on (5) unrestricted and use the coefficient on the lagged dependent variable (*LDV*) as a consistent estimate $\hat{\rho}$.
2. Transform the original equation by $\hat{Q}$ or $\hat{Q}^*$ and run OLS on the transformed variables.

4.1.3 Properties of feasible GLS.

1. In general, they are biased (unless we have symmetric errors).
2. If $\hat{\rho}$ is consistent, so is $\hat{\beta}_{FG}$.
3. If no *LDV* in the regression $\hat{\beta}_{FG}$ is asymptotically equivalent to $\hat{\beta}_{GLS}$ and statistical tests based on it are asymptotically valid.
4. If no *LDV* and the error process is incorrectly specified, $\hat{\beta}_{FG}$ is consistent but no longer asymptotically equivalent to $\hat{\beta}_{GLS}$. It is not asymptotically efficient and the tests are not asymptotically valid.
5. If there is a *LDV* in the regression, $\hat{\rho}$ is not consistent, $\hat{\beta}_{FG}$ is not consistent and not asymptotically efficient and the tests are not asymptotically valid. We need to go to MLE.

4.1.4 Maximum likelihood procedure.

We recall the transformation $e = Qu \sim N(0, \sigma^2 I)$ given by

$$\begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2} & & 0 \\ -\rho & 1 & \\ & \ddots & \\ 0 & & -\rho & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_T \end{bmatrix}.$$

The Jacobian of the transformation is

$$\left| \frac{\partial e_i}{\partial u_j} \right| = |Q| = (1 - \rho^2)^{1/2}$$

since the determinant of a triangular matrix is the product of the diagonal elements. We set up the likelihood function in terms of the errors e and transform in terms of u and thus y . Let $\theta = (\beta, \rho)$ be the full vector of parameters. To start

$$L(\theta, e) \stackrel{\text{prop}}{\propto} (\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} (e'e) \right].$$

Transforming in terms of the u 's:

$$L(\theta, u) \propto (\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} \left(\underbrace{u'Q'Qu}_{\Omega^{-1}} \right) \right] \overset{\text{Jacobian}}{\downarrow} (1 - \rho^2)^{1/2}$$

and, in terms of the y 's:

$$L(\theta, y) \propto (\sigma^2)^{-T/2} \exp \left[-\frac{1}{2\sigma^2} (y - X\beta)' \Omega^{-1} (y - X\beta) \right] (1 - \rho^2)^{1/2}.$$

Taking logarithms:

$$\ell n L(\theta, y) = \text{const} - \frac{T}{2} \ell n(\sigma^2) + \frac{1}{2} \ell n(1 - \rho^2) - \frac{1}{2\sigma^2} (y^* - X^*\beta)' (y^* - X^*\beta)$$

where y^* and X^* are transformed with Q . The next step is to concentrate out σ^2 :

$$\frac{\partial \ell n L}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{T} (y^* - X^*\beta)' (y^* - X^*\beta) = \frac{S(\rho, \beta)}{T}.$$

We substitute back to get the concentrated log likelihood function:

$$\ell n L(\rho, \beta; y) = \ell n L(\hat{\sigma}^2, \rho, \beta; y) = \text{const} - \left(\frac{T}{2} \right) \ell n [S(\rho, \beta)] + \frac{1}{2} \ell n(1 - \rho^2)$$

where

$$\begin{aligned} S(\rho, \beta) &= (y^* - X^*\beta)' (y^* - X^*\beta) \\ &= [Q(y - X\beta)]' [Q(y - X\beta)] \\ &= (1 - \rho^2) (y_1 - x'_1\beta)^2 + \sum_{t=2}^T (y_t - \rho y_{t-1} - x'_t\beta + \rho x'_{t-1}\beta)^2 \\ &= (1 - \rho^2) (y_1 - x'_1\beta)^2 + \sum_{t=2}^T [(y_t - \rho y_{t-1}) - (x'_t - \rho x'_{t-1})\beta]^2. \end{aligned}$$

This is a nonlinear problem, although conditional on ρ , maximizing the log likelihood function with respect to β is equivalent to running GLS. The toughest part is to solve for ρ first. A procedure was suggested by Beach and Mackinnon (1978). Hence, the iterative procedure is:

1. For a given β solve for ρ (nonlinear);
2. For this ρ apply GLS, get another $\hat{\beta}$;
3. Iterate until convergence.

4.1.5 Other procedures.

Hildreth-Lu scanning technique (HILU).

1. Choose some value $\hat{\rho}^{(1)}$ and run the GLS regression

$$\hat{\beta}_{FG}^{(1)} = \left(X' \hat{Q}^* \hat{Q}^* X \right)^{-1} X' \hat{Q}^* \hat{Q}^* y$$

using the matrix Q^* for the transformation, i.e. without the first observation;

2. Calculate $S \left(\hat{\rho}^{(1)}, \hat{\beta}_{FG}^{(1)} \right)$ and the log-likelihood

$$\ln L_{(1)}^* \left(\hat{\rho}^{(1)}, \hat{\beta}_{FG}^{(1)} \right) = \text{const} - \frac{(T-1)}{\alpha} \ln \left[S \left(\hat{\rho}^{(1)}, \hat{\beta}_{FG}^{(1)} \right) \right];$$

3. Scan over many possible values of $\rho \in (-1, 1)$ and obtain a set of log-likelihood values $\ln L_{(1)}^*, \ln L_{(2)}^*, \dots$;
4. Choose $\hat{\rho}$ and the corresponding $\hat{\beta}_{FG}$ such that $\ln L^*$ is a maximum.

Scanning technique with the exact likelihood function (MIGRID).

This is the same procedure as HILU but the likelihood function is exact and the GLS regression is

$$\hat{\beta}_{FG} = \left(X' \hat{Q}' \hat{Q} X \right)^{-1} X' \hat{Q}' \hat{Q} y$$

This uses the transformation matrix $\hat{Q}$ and, hence, includes the first observation. This is computationally inefficient [especially compared to the maximum likelihood technique of Beach-Mackinnon] but useful if we want to check for multiple optima and global maximum.

Cochrane-Orcutt iterative procedure This is a sequential procedure based on the approximate likelihood function (without the 1st observation).

1. Conditional of $\hat{\rho}$ max $\ln L^*$ with respect to $\beta \Leftrightarrow$ OLS of $\hat{Q}^* y$ on $\hat{Q}^* X$.
2. Conditional on $\hat{\beta}$, max $\ln L^*$ with respect to $\rho \Leftrightarrow$ OLS of $\hat{u}_t$ on $\hat{u}_{t-1}$,
3. Keep iterating until convergence.

Nonlinear least squares. Recall Durbin's procedure, where we had the transformation

$$y_t = \rho y_{t-1} + x_t' \beta - x_{t-1}' (\rho \beta) + \varepsilon_t, \quad t = 2, \dots, T.$$

Note that for $t = 2, \dots, T$, we loose the 1st observation.

This a nonlinear regression problem with white noise errors of the form

$$Y_t = f_t(\rho, \beta) + \varepsilon_t.$$

So, we can use a method such as the Gauss-Newton procedure to minimize the residuals sum of squares. This has the same asymptotic properties as MLE.

4.2 Extensions to more general $AR(\rho)$ processes

The model considered is now:

$$\begin{aligned} y &= X\beta + u \\ u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + e_t. \end{aligned}$$

where e_t is *i.i.d.* $N(0, \sigma^2)$. Here,

$$Q_{T \times T} = \begin{bmatrix} R & 0 \\ & P \end{bmatrix}$$

where R is lower triangular (quite complex) and P is $p \times p$

$$P_{(T-p) \times T} = \begin{bmatrix} -\rho_p & \dots & -\rho_1 & 1 & & \\ & \ddots & & & & \\ & & & & & \\ & & & -\rho_p & \dots & -\rho_1 & 1 \end{bmatrix}.$$

For Cochrane-Orcutt (approximate MLE), we loose the first p observations. The iterative procedure uses P as the transformation matrix for GLS on β and OLS on a p^{th} order autoregressive process on $\hat{u}_t$.

4.2.1 Exact MLE

The log-likelihood function is

$$\ln L(\rho, \beta; y) = \text{const} + \frac{1}{2} \ln(R'R) - \frac{T}{2} \ln[u_*' R' R u_* + (y - X\beta)' P' P (y - X\beta)]$$

where $u_* = (u_1, \dots, u_p)$; the first p residuals. We can iterate in a similar fashion as for the $AR(1)$ case.

4.3 Estimation with moving-average processes.

The model is

$$\begin{aligned} y &= X\beta + u \\ u_t &= e_t + \theta e_{t-1} \quad e_t \sim i.i.d.N(0, \sigma^2). \end{aligned}$$

Para obtener la matriz de varianzas y covarianzas del vector de perturbaciones, debemos tener en cuenta que:

a)

$$E[u_t] = 0.$$

b)

$$\begin{aligned} E[u_t^2] &= E \left[e_t^2 + 2\theta \overset{0}{\uparrow} e_t e_{t-1} + \theta^2 e_{t-1}^2 \right] \\ &= (1 + \theta^2) \sigma^2. \end{aligned}$$

c)

$$\begin{aligned} E[u_t u_{t-1}] &= E[(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2})] \\ &= E \left[\begin{array}{cccc} e_t e_{t-1} & + & \theta e_t e_{t-2} & + & \theta e_{t-1}^2 & + & \theta^2 e_{t-1} e_{t-2} \\ \downarrow & & \downarrow & & & & \downarrow \\ 0 & & 0 & & & & 0 \end{array} \right] \\ &= \theta \sigma^2. \end{aligned}$$

d)

$$E[u_t u_{t-s}] = 0 \quad \forall s \geq 2.$$

So,

$$\Omega = \sigma^2 \begin{bmatrix} 1 + \theta^2 & \theta & & & & \\ \theta & (1 + \theta^2) & \theta & & & \\ & \theta & (1 + \theta^2) & & & \\ & & \ddots & \ddots & \ddots & \\ & & & \theta & (1 + \theta^2) & \theta \\ & & & & \theta & (1 + \theta^2) \end{bmatrix}$$

which is called a band symmetric Toeplitz matrix. The problem with feasible GLS is that Ω^{-1} is a real mess and no simple nonsingular matrix Q such that $Q'Q = \Omega^{-1}$ is known. Therefore, there is no simple two-step FGLS procedure for MA(q) processes. We go directly to MLE (either exact or approximate).

4.3.1 Approximate maximum likelihood for MA(1) error.

This is approximate MLE in the sense that we take the initial error $e_0 = 0$. That is, we condition on $e_0 = 0$ (imposing the unconditional expectation for the initial error). Then, we can write the model as

$$\begin{aligned} e_t &= (y_t - x_t' \beta) - \theta e_{t-1} \quad t = 1, \dots, T \\ e_0 &= 0. \end{aligned} \tag{6}$$

Maximizing the likelihood function here is equivalent to minimizing the sum of squared residuals. We can minimize this using the Gauss-Newton algorithm. Recall the iterative scheme

$$\psi^* = \tilde{\psi} + \left(\sum_{t=1}^T z_t z_t' \right)^{-1} \sum_{t=1}^T z_t e_t$$

where

$$z_t = -\frac{\partial e_t}{\partial \psi}.$$

Here, $\psi' = (\beta', \theta)$, and

$$\begin{aligned} \frac{\partial e_t}{\partial \beta} &= -x_t - \theta \frac{\partial e_{t-1}}{\partial \beta} \quad t = 1, \dots, T \\ \frac{\partial e_0}{\partial \beta} &= 0 \end{aligned} \tag{7}$$

and

$$\begin{aligned} \frac{\partial e_t}{\partial \theta} &= -\theta \frac{\partial e_t}{\partial \theta} - e_{t-1} \quad t = 1, \dots, T \\ \frac{\partial e_0}{\partial \theta} &= 0. \end{aligned}$$

Note that these can be calculated by recursion. Now, the information matrix is block diagonal since

$$\text{p lim } T^{-1} \sum_{t=1}^T \frac{\partial e_t}{\partial \beta} \frac{\partial e_t}{\partial \theta} = 0.$$

[We can show this by a recursion argument using the independence of x_t and e_{t-1} .] This block diagonality suggests an iterative procedure involving two separate regressions.

$$e_t \text{ on } \frac{\partial e_t}{\partial \beta} \quad (8)$$

$$e_t \text{ on } \frac{\partial e_t}{\partial \theta} \quad (9)$$

Equation (9) does not yield the MLE of θ conditional on β but yields an approximate GLS estimator.

4.3.2 Why is this approximate GLS?

Conditional on θ , minimizing the sum of squared residuals is obtained when

$$\frac{\partial S}{\partial \beta} = 2 \sum_{t=1}^T \frac{\partial e_t}{\partial \beta} e_t = 0 \quad (10)$$

where

$$S(\cdot) = \sum_{t=1}^T e_t^2.$$

Now, do a recursive substitution of (6) and (7)

$$e_t = \sum_{j=0}^{t-1} (-\theta)^j (y_{t-j} - x'_{t-j} \beta)$$

and

$$\frac{\partial e_t}{\partial \beta} = - \sum_{j=0}^{t-1} (-\theta)^j x_{t-j}.$$

Substitute in (10)

$$\sum_{t=1}^T x_t^* (y_t^* - x_t^{*'} \beta) = 0 \quad (11)$$

where

$$\begin{aligned} y_t^* &= \sum_{j=0}^{t-1} (-\theta)^j y_{t-j} \\ x_t^* &= \sum_{j=0}^{t-1} (-\theta)^j x_{t-j} = - \frac{\partial e_t}{\partial \beta}. \end{aligned}$$

So, the GLS estimator is obtained by regressing the transformed y_t^* on x_t^* ; i.e., (11) yields

$$\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} y^*.$$

This is identical to the estimate obtained by regressing e_t on $\frac{\partial e_t}{\partial \beta}$ and subtracting the result from the current estimate of β .

Otros textos recomendados

1. Greene, W.H., 1999, Análisis Econométrico, Ed. Prentice Hall Iberia, Madrid. Cap. 13.
2. Johnston, J., 1987, Métodos de Econometría, Ed. Vicens Vives, Barcelona. Cap. 8.
3. Judge, G.G., R. C. Hill, W. E. Griffiths, H. Lütkepohl y T.C. Lee, 1988, Introduction to the Theory and Practice of Econometrics, Ed. John Wiley & Sons. Cap. 9.
4. Kmenta, J., 1985, Elementos de Econometría, Ed. Vicens Vives, Barcelona. Cap. 8.
5. Maddala, G.S. (1992) Introduction to Econometrics. Ed: MacMillan. Cap. 6.
6. Novales, A. (1993). Econometría. Ed. McGraw-Hill. Cap.
7. Pindyck, R.S. y D. L. Rubinfeld, 2000, Econometría: Modelos y Pronósticos, Ed. Mc Graw-Hill. Cap. 6

Referencias

References

- [1] Box, G. E. P. y D. A. Pierce (1970). "Distribution of residual autocorrelations in autoregressive integrated moving average time series models", Journal of the American Statistical Association, 65, 1509-1526.
- [2] Breusch, T. S. (1978). "Testing for autocorrelation in dynamic linear models", Australian Economic Papers, 17, 334-355.
- [3] Breusch, T.S. y L.G. Godfrey (1981). "A review of recent work on testing for autocorrelation in dynamic economic models", en Macroeconomic Analysis: Essays in Macroeconomics and Economics, De: D. A. Currie, R. Nobay y D. Peel, Croom Helm, Londres
- [4] Durbin, J. (1970). "Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables", Econometrica, 38, 410-421.

- [5] Durbin, J. y G. S. Watson (1950). "Testing for serial correlation in least squares regression I", *Biometrika*, 37, 409-428.
- [6] Durbin, J. y G. S. Watson (1951). "Testing for serial correlation in least squares regression II", *Biometrika*, 38, 159-178.
- [7] Godfrey, L. G. (1976). "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables". *Econometrica*, 46, 1303-1310.
- [8] Godfrey, L. (1978). "Testing against general autoregressive and moving average models when the regressors include dependent variables". *Econometrica*, 46, 1293-1301.
- [9] Nerlove, M. y K. F. Wallis (1966). "Use of the Durbin-Watson statistic in inappropriate situations". *Econometrica*, 34, 235-238.
- [10] Wallis, K.F. (1972). "Testing for fourth quarter autocorrelation in Quarterly Regression Equations", *Econometrica*, 40, 617-636.