CAPÍTULO 6

CONTRASTES DE ESFERICIDAD

6.1. INTRODUCCIÓN

Cuando hemos introducido las hipótesis del Modelo Lineal General en el Capítulo 1 hemos comenzado agrupando los determinantes del comportamiento de la variable dependiente en dos partes: sistemática y aleatoria.

En la parte sistemática se recogían aquellos factores para los que era posible identificar una pauta de influencia estable y se trataba de representar explícitamente esta pauta. La parte aleatoria daba cuenta de la influencia de todos aquellos factores que, ni individualmente, ni en su conjunto, tienen una influencia identificable sobre la variable dependiente. A toda ésta influencia sin pauta definida la llamamos perturbación aleatoria. Da cuenta de un efecto conjunto de un grupo grande de factores pero no tiene una pauta definida ni, en principio, relevante.

Pero para estar seguros de que la perturbación aleatoria no “camufla” nada relevante debe garantizarse el que dicha perturbación cumpla unas propiedades estocásticas determinadas. La primera condición es que su esperanza matemática sea cero. Esta propiedad es fundamental ya que de no ser cero indicaría que el efecto conjunto de los factores representados por la perturbación aleatoria tiende a ser positivo o negativo según sea el signo de expectativa; y si es positivo o negativo entonces dicho efecto ya parecería ser relevante. Podría identificarse una pauta a tener en cuenta a la hora de explicar el comportamiento de la variable dependiente.

Otra característica que debe cumplir la perturbación aleatoria es la de no estar autocorrelada. De estarlo, el comportamiento en un periodo dependería de lo acontecido en periodos anteriores y eso ya no sería algo irrelevante. Por lo tanto, debemos garantizar que no existe autocorrelación.

La tercera característica que se menciona es que la perturbación debe ser homoscedástica en el sentido de que la dispersión en torno a la media en todos los periodos debe ser la misma. También en este caso la falta de homogeneidad en la varianza nos apuntaría a que algo, relevante, estaba siendo olvidado.

Por último, se suele indicar que la perturbación debe seguir una distribución Normal. Las características de esta distribución de probabilidad, simétrica en torno a la media, un determinado nivel de apuntamiento, parecen los mejores indicadores para caracterizar lo que hemos llamado “efecto conjunto no relevante”.

Una perturbación aleatoria que cumple todas estas propiedades decimos que es esférica. Es decir, decimos que la perturbación aleatoria es esférica cuando:

- Es estacionaria con una media igual a cero.

- No está autocorrelacionada.

- Es homoscedástica.

- Sigue una distribución Normal.

Este capítulo está dedicado a examinar algunos de los procedimientos propuestos para concluir si la evidencia empírica utilizada conduce a aceptar o rechazar el carácter esférico de la perturbación aleatoria. No se pretende cubrir todos los procedimientos propuestos en la literatura; sólo se va a prestar atención a los relacionados con el contraste de los Multiplicadores de Lagrange. Un tratamiento de otros métodos puede encontrarse en Fomby et al. (1984).

En la Sección 6.2 se hará una presentación intuitiva de los procedimientos para contrastar la estacionariedad, la no autocorrelación, la homoscedasticidad y la normalidad. Un resultado que se extrae del análisis de todos estos procedimientos es que de lo que se trata es contrastar que un subconjunto de todos los parámetros cumplen una serie de restricciones. Ésta es la razón por la que la sección siguiente, la 6.3, está dedicada a estudiar la forma que toma el contraste de los Multiplicadores de Lagrange cuando las restricciones sólo afectan a un subgrupo del total de parámetros.

En la última sección, la 6.4, derivamos el estadístico asociado con los Multiplicadores de Lagrange para contrastar las hipótesis nulas de no autocorrelación, homoscedasticidad y normalidad.

6.2 CONTRASTES DE ESFERICIDAD.

En esta sección vamos a distinguir cuatro grandes apartados para dar cuenta de los procedimientos propuestos para contrastar cuatro aspectos asociados con la esfericidad.

6.2.1. Estacionariedad

Como ya se ha indicado en el análisis univariante, la estacionariedad se refiere a que la distribución de probabilidad de las perturbaciones no varía a lo largo de los diferentes períodos.

Como puede verse en Aznar y Trívez (1993), puede hablarse de la estacionariedad global y local. La primera, presta atención a todas las observaciones muestrales y, en la segunda, sólo se presta atención a algunas de estas observaciones muestrales en particular.

Nosotros, en esta Sección, vamos a centrarnos en los instrumentos y problemas asociados con la detección y el diagnóstico del segundo tipo de no estacionariedad. Los procedimientos a utilizar para el primer tipo ya han sido comentados al hablar del análisis univariante y de la cointegración.

Sobre cómo detectar observaciones aisladas que denoten falta de estacionariedad, una primera aproximación útil puede consistir en tomar individualmente cada uno de los residuos MCO del modelo y aplicar contrastes asociados con los conceptos de "residuo estudentizado interno" y "residuo estudentizado externo" tal como puede verse en Aznar y Trívez (1993).

A continuación, vamos a demostrar que este análisis individualizado puede ser complementado especialmente con técnicas multivariantes.

Comenzaremos presentando cuatro casos diferentes para ilustrar el tipo de problemas que se presentan y apuntar las soluciones que parecen más razonables.

CASO 1. Regresión Normal Sin Atípicos

Observación	X	y	$\hat{u}$
1	6	4	-0,56
2	9	6	0,14
3	14	8	-0,027
4	18	11	1,23
5	25	12	-0,79

$$\hat{y}_t = 1{,}96 + 0{,}43X_t$$
$$\qquad\quad (0{,}06)$$

$R^2 = 0{,}94$

En la derecha, aparecen las estimaciones MCO de los parámetros, el valor tomado por el estadístico R^2 y, entre paréntesis, el valor estimado de la desviación típica del estimador MCO del parámetro.

Suponemos que estas estimaciones de los parámetros representan la verdadera estructura de la relación entre x e y.

CASO 2. Regresión con Atípico en la variable explicativa

Observación	X	Y	$\hat{u}$	$\hat{u}^*$	$\hat{u}^{**}$
1	6	4	-2,14	-0,044	-0,217
2	9	6	-0,60	0,28	0,251
3	14	8	0,62	-0,50	-0,302
4	18	11	3,01	0,26	0,655
5	50(22)	12	0,89		-0,387

Regresión con las 5 observaciones:

$$\hat{y}_t = 5{,}22 + 0{,}15\ X_t \qquad R^2 = 0{,}66$$

Regresión con sólo las 4 primeras observaciones:

$$\hat{y}_t^* = 0{,}69 + \underset{(0{,}049)}{0{,}55}\ X_t \qquad R^2 = 0{,}97$$

Regresión con la serie corregida:

$$\hat{y}_t^{**} = 1{,}15 + \underset{(0{,}039)}{0{,}51}\ X_t \qquad R^2 = 0{,}97$$

La observación atípica afecta a la variable explicativa en la observación 5. Se presentan tres regresiones diferentes con sus correspondientes residuos. La primera es la que se obtiene utilizando todas las observaciones, la segunda, la que se obtiene eliminando la observación atípica y, en la tercera, se utiliza la serie corregida obtenida sustituyendo el valor atípico de la variable explicativa por el valor que resulta de sumar al valor de la observación 4 el promedio de los tres incrementos anteriores.

Se observa que utilizando las cinco observaciones originales se distorsiona la estructura estimada pasando la pendiente estimada de 0,43 a 0,15 y el término constante de 1,96 a 5,22; empeora notablemente el ajuste y los residuos envían un mensaje erróneo indicando que algo atípico ocurre en la observación 4 cuando, en realidad, el problema está en la observación 5.

Si la regresión se lleva a cabo eliminando la observación atípica o corrigiéndola, la estructura se estima razonablemente bien, el ajuste se mantiene alto y los residuos no envían ningún mensaje erróneo.

<u>CASO 3</u>. Regresión con Atípico en la variable dependiente

Observación	X	Y	$\hat{u}$	$\hat{u}^*$	$\hat{u}^{**}$
1	6	4	2,17	-0,044	-0,305
2	9	6	0,89	0,28	0,21
3	14	8	-2,56	-0,50	-0,26
4	18	11	-3,93	0,26	0,759
5	25	26(13,3)	3,42		-0,402

Regresión con las 5 observaciones:

$$\hat{y}_t = -4{,}73 + 1{,}09\, X_t \qquad R^2 = 0{,}82$$

Regresión con solo las 4 primeras observaciones:

$$\hat{y}_t = 0{,}69 + \underset{(0{,}05)}{0{,}55}\, X_t \qquad R^2 = 0{,}98$$

Regresión con la serie corregida:

$$\hat{y}^*_t = 1{,}33 + \underset{(0{,}037)}{0{,}49}\, X_t \qquad R^2 = 0{,}97$$

La evidencia de estos datos es similar a la comentada para el caso anterior: la regresión con las cinco observaciones distorsiona la estructura entre las dos variables, empeora el ajuste y la serie de residuos envía un mensaje erróneo. La situación se arregla bien utilizando sólo las cuatro primeras observaciones o la serie corregida.

<u>CASO 4</u>. Regresión con Atípicos en ambas variables en el mismo periodo

Observación	X	y	$\hat{u}$
1	6	4	-0,37
2	9	6	0,14
3	14	8	-0,32
4	18	11	0,69
5	50	26	-0,13

Regresión con las 5 observaciones:

$$\hat{y}_t = 1{,}40 + \underset{(0{,}014)}{0{,}49}\, X_t \qquad R^2 = 0{,}99$$

Esta regresión mantiene la estructura estimada entre las dos variables, mejora el ajuste y la serie de residuos no envía ningún mensaje equívoco. En este caso no hay ninguna razón ni para eliminar observaciones ni para corregir la serie.

Las conclusiones que pueden derivarse de la evidencia presentada en los cuatro casos analizados son las siguientes:

1. La presencia de determinados atípicos puede distorsionar gravemente la estimación de la estructura que establece la relación entre ambas variables. Esta distorsión será mayor cuanto menor sea la evidencia disponible.

2. La práctica seguida en muchos trabajos consistente en hacer la regresión con los datos originales, analizar la serie de residuos y, según sean éstos, reestimar y chequear el modelo puede llevar a resultados bastante insatisfactorios.

3. La forma más razonable de tratar las observaciones atípicas consiste en llevar a cabo una determinación de este tipo de observaciones para cada una de las variables con el procedimiento comentado en el Apartado 4.2. Una vez que se han detectado, en aquellos casos en que sólo hay atípicos en la variable dependiente o sólo en la variable explicativa o en ambas pero no coincidentes, entonces la mejor solución es eliminar esas observaciones o estimar con las series corregidas. Aquellas observaciones en las que haya coincidencia de atípicos en las dos variables pueden mantenerse, en algunos casos, siempre que la correspondencia se ajuste a la estructura de la relación.

Hemos comentado que la solución al problema que se deriva de la presencia de atípicos va o bien en la línea de eliminar las observaciones catalogadas como atípicos o bien utilizando series corregidas para llevar a cabo la estimación.

Considerar dos variables, X_t e y_t, medidas en desviaciones respecto a sus medias.

Suponer los siguientes tres modelos:

$$y_t = \beta_1 X_t + u_{1t}$$

$$y_t^* = \beta_2 X_t^* + u_{2t}$$

$$y_t = \beta_3 X_t + \delta d_{1t} + u_{3t}$$

d_{1t} es la variable que toma el valor 1 en la observación t_0 y 0 en el resto. Las variables X_t^* e y_t^* se definen así:

$$y_t^* = y_t - d_{1t}\hat{\eta}_1$$

$$X_t^* = X_t - d_{1t}\hat{\eta}_2$$

en donde:

$$\hat{\eta}_1 = \frac{\Sigma d_{1t} y_t}{\Sigma d_{1t}^2} = y_{t0} \;;\; \hat{\eta}_2 = \frac{\Sigma d_{1t} X_t}{\Sigma d_{1t}^2} = X_{t0}$$

Podemos definir tres estimadores y analizar las relaciones entre ellos:

$$\hat{\beta}_1 = \frac{\Sigma X_t y_t}{\Sigma X_t^2}$$

$$\hat{\beta}_2 = \frac{\Sigma y_t^* X_t^*}{\Sigma X_t^{*2}} = \frac{\Sigma(y_t - y_{t_0} d_{1t})(X_t - X_{t_0} d_{1t})}{\Sigma(X_t - X_{t_0} d_{1t})^2} = \frac{\Sigma X_t y_t - X_{t0} y_{t0}}{\Sigma X_t^2 - X_{t0}^2}$$

Se ve que el estimador $\hat{\beta}_2$ es el mismo que $\hat{\beta}_1$ eliminando la observación t_0.

A partir del último modelo podemos definir:

$$\begin{bmatrix} \hat{\beta}_3 \\ \hat{\delta} \end{bmatrix} = \frac{1}{|\;|} \begin{bmatrix} \sum d_{1t}^2 & -\sum d_{1t} X_t \\ -\sum d_{1t} X_t & \sum X_t^2 \end{bmatrix} \begin{bmatrix} \sum X_t y_t \\ \sum d_{1t} y_t \end{bmatrix}$$

en donde:

$$|\;| = (\sum d_{1t}^2)(\sum X_t^2) - (\sum d_{1t} X_t)^2$$

El estimador $\hat{\beta}_3$ toma la forma siguiente:

$$\hat{\beta}_3 = \frac{(\sum d_{1t}^2)(\sum X_t y_t) - (\sum d_{1t} X_t)(\sum d_{1t} y_t)}{(\sum d_{1t}^2)(\sum X_t^2) - (\sum d_{1t} X_t)^2} = \hat{\beta}_2$$

teniendo en cuenta que:

$$\sum d_{1t}^2 = 1 \;,\; \sum d_{1t} X_t = X_{t_0} \;,\; \sum d_{1t} y_t = y_{t_0} \;,\; \sum\left(d_{1t} X_t^2\right) = X_{t_0}^2$$

Hay que destacar que la equivalencia encontrada se alcanza suponiendo que las dos variables se limpian en el mismo periodo. Si cada variable se limpia en periodos diferentes la equivalencia ya no se da.

Un último aspecto al que vamos a prestar atención en este apartado es el carácter multivariante que, en general, tiene la parte sistemática de la mayoría de los modelos econométricos. Las implicaciones de este carácter multivariante son que puede ocurrir que se haya detectado una observación como atípica para una de las variables individuales pero que no sea atípica para el conjunto de la parte sistemática y, al contrario, que no se detecte, para una observación, nada atípico para las variables individuales, y que dicha observación sea atípica para la parte sistemática en su conjunto.

Teniendo en cuenta la posibilidad de estos resultados, al lado del procedimiento, ya descrito, para detectar atípicos en series individuales, es necesario pensar en técnicas multivariantes que sirvan para detectar observaciones atípicas, para un grupo de variables, entre las cuales se define una relación lineal.

A continuación se describe un procedimiento que persigue este objetivo basado en los mismos principios sobre los que se apoyaba el método univariante: hacer una primera eliminación de las observaciones más distintas con base en una medida multivariante, calcular con el resto de las observaciones un centro de gravedad y una norma de distancia a ese centro de gravedad, derivar una región de confianza con dicha norma de distancia en torno a ese centro de gravedad y determinar cuáles de las observaciones eliminadas al comienzo pueden calificarse como atípicas.

Los dos conceptos básicos que nos van a servir para elaborar el procedimiento que se propone son: la matriz "hat" y la distancia de Mahalanobis.

La matriz "hat" se define como:

$$H = X(X'X)^{-1}X'$$

Esta matriz cobra sentido cuando se define el vector de valores estimados, en el modelo lineal general:

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

El elemento genérico, h_{ts}, viene dado por:

$$h_{ts} = x_t'(X'X)^{-1}x_s$$

en donde x_t y x_s son, respectivamente, la fila t-ésima y la columna s-ésima de X. Este elemento nos informa de la aportación de la observación s-ésima del vector y en el cálculo de la observación t-ésima del vector de valores ajustados. La suma de todos estos elementos cuando t = s viene dada por:

$$\mathrm{tr}\, X(X'X)^{-1}X' = \mathrm{tr}\, (X'X)^{-1}X'X = k$$

en donde k es el número de variables.

Fácilmente se ve que la matriz H es simétrica e idempotente; es decir, H = H' y HH = H. Por lo tanto, se tiene que:

$$\sum_{t=1}^{T} h_{tt} = k \tag{6.1}$$

Por ser simétrica e idempotente se tiene que:

$$h_{tt} = (HH)_{tt} = \sum_{s=1}^{T} h_{ts}h_{st} = \sum_{s=1}^{T} h_{ts}h_{ts} = \sum_{s=1}^{T} h_{ts}^{2} \tag{6.2}$$

Es inmediato también el siguiente resultado:

$$\mathrm{Cov}\,(\hat{y}) = \sigma^2 H$$

$$\mathrm{Cov}(\hat{u}) = \sigma^2 (I - H)$$

Siguiendo con la interpretación dada al elemento genérico podemos decir que h_{tt} mide el efecto de la observación t-ésima sobre el valor ajustado en el mismo periodo. Teniendo en cuenta (6.1), el valor medio de los h_{tt} viene dado por k/T.

A partir de (6.2) podemos escribir:

$$h_{tt} = \sum h_{ts}^2 = h_{tt}^2 + \sum_{s\neq t}^{T} h_{ts}^2 \qquad \text{para } s \neq t \tag{6.3}$$

de donde se obtiene que:

$$0 \leq h_{tt} \leq 1$$

También es claro a partir de (6.3) que si $h_{tt} = 0$ entonces todos h_{ts} son iguales a cero; si $h_{tt} = 1$ todos h_{ts} son cero para todo valor de s excepto para s = t. Cuando $h_{tt} = 1$, el valor ajustado de la variable coincide con el valor observado siendo el residuo igual a cero, lo mismo que su varianza.

El segundo concepto al que nos hemos referido es el de la distancia de Mahalanobis. Sea $x_t' = (X_{1t}, X_{2t}, ..., X_{kt})$ el vector de la observación t-ésima de las k variables (si una de ellas es la constante hay que eliminarla previamente y continuar el análisis con las (k-1) restantes). El vector de medias y matriz de varianzas y covarianzas se definen, respectivamente, como:

$$\overline{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$$

$$V = \frac{1}{T-1} \sum \left(X_t - \overline{X}\right)\left(X_t - \overline{X}\right)'$$

La distancia de Mahalanobis del punto correspondiente a la observación t-ésima al punto correspondiente al vector de las medias, $\overline{X}$, viene dada por:

$$MD_t^2 = \left(X_t - \overline{X}\right)' V^{-1}\left(X_t - \overline{X}\right) \qquad (6.4)$$

Un resultado importante es el que pone en relación la distancia de Mahalanobis y el elemento correspondiente de la matriz "hat". La relación viene dada por:

$$MD_t^2 = (T\text{-}1)\ h_{tt}$$

Esta expresión lo que nos dice es que los dos conceptos nos informan de lo mismo. Teniendo en cuenta (6.4) se ve que:

$$0 \leq MD_t^2 \leq (T\text{-}1)$$

Para elaborar un procedimiento para detectar atípicos en un marco multivariante a partir de ambos conceptos, vamos a comenzar considerando el caso en que todas las variables individuales son estacionarias. Las etapas del proceso que se propone son las siguientes:

1) Se obtienen para las T observaciones los correspondientes elementos de la diagonal principal, h_{tt}, de la matriz H de las k variables explicativas.

2) Se eliminan el 10% de las observaciones a las que corresponda un mayor valor de h_{tt}.

3) Con el resto de las observaciones calcular el vector de medias y la matriz de varianzas y covarianzas, según las expresiones vistas anteriormente.

4) Utilizando la expresión (6.4) calcular la distancia de cada una de las observaciones excluidas en 2) respecto del vector de medias obtenido en 3).

5) Concluir con que una de las observaciones eliminadas en 2) es atípica cuando su correspondiente distancia calculada en 4) supera una determinada cota especificada a priori. Esta cota puede fijarse, por ejemplo, haciéndola igual al punto crítico correspondiente a una variable χ^2 con (k-1) grados de libertad para un nivel de significación muy pequeño como el 1% o menor.

6.2.2. Contraste de autocorrelación

Los procedimientos para contrastar autocorrelación, en general, parten de un modelo dado y asumen una estructura autorregresiva concreta. La hipótesis nula de no autocorrelación que se contrasta equivale al contraste de la hipótesis de que ciertos coeficientes toman el valor cero.

Considerando el esquema más simple, el modelo del que partimos puede escribirse como:

$$y_t = \beta X_t + u_t \tag{6.5}$$

con $$u_t = \rho u_{t-1} + v_t \tag{6.6}$$

en donde X_t es estocástica independiente de v_t; β y ρ son parámetros con $|\rho| < 1$. La hipótesis nula de no autocorrelación se concreta en:

$$H_0: \rho = 0. \tag{6.7}$$

Sustituyendo (6.6) en (6.5) se obtiene:

$$y_t = \beta X_t + \rho u_{t-1} + v_t \tag{6.8}$$

Lo lógico es estimar ϕ a partir de (6.8) y contrastar la hipótesis nula (6.7) utilizando el t-ratio. Por ser u_{t-1} una entidad no observable se sustituye por el residuo MCO definido a partir de (6.5):

$$\hat{u}_t = y_t - \hat{\beta} X_t$$

con $$\hat{\beta} = \frac{\sum X_t y_t}{\sum X_t^2}$$

Hecha esta sustitución, el estimador MCO de ρ en (6.8) utilizando el Resultado 1.38 viene dado por:

$$\hat{\rho} = (\hat{u}'_{-1} M \hat{u}'_{-1})^{-1} \hat{u}'_{-1} My \tag{6.9}$$

con $$M = I_T - X(X'X)^{-1}X'$$

Asintóticamente y bajo H_0 la distribución de probabilidad de $\hat{\phi}$ viene dada por:

$$\hat{\phi} \sim N(0, \sigma^2(\hat{u}'_{-1} M \hat{u}_{-1})^{-1})$$

El t-ratio toma la forma siguiente:

$$t = \frac{\hat{\rho}}{\hat{\sigma}_{\hat{\phi}}}$$

Asumiendo un nivel de significación ε_0, la región crítica del contraste es la siguiente:

$$\frac{\hat{\rho}}{\hat{\sigma}_{\hat{\phi}}} > t_{\varepsilon_0}$$

en donde t_{ε_0} es el punto crítico correspondiente a la distribución t-Student.

Por ser X_t independiente de v_t entonces, asintóticamente, se tiene que:

$$\frac{\hat{u}'_{-1}M\hat{u}_{-1}}{T} = \frac{\hat{u}'_{-1}\hat{u}_{-1}}{T} - \frac{\hat{u}'_{-1}X}{T}\frac{(X'X)^{-1}}{T}\frac{X'\hat{u}_{-1}}{T} = \frac{\hat{u}'_{-1}\hat{u}_{-1}}{T} \qquad (6.10)$$

Teniendo en cuenta este resultado y que bajo H_0 se cumple My = $\hat{u}$, el estimador escrito en (6.9) puede escribirse como:

$$\hat{\rho} = (\hat{u}'_{-1}\hat{u}_{-1})^{-1}\hat{u}'_{-1}\hat{u}$$

que sería el estimador MCO de ρ en la regresión:

$$\hat{u}_t = \rho\hat{u}_{t-1} + v_t \qquad (6.11)$$

Suponer ahora que en lugar del modelo escrito en (6.5) consideramos el siguiente modelo:

$$y_t = \lambda y_{t-1} + \beta X_t + u_t$$

La peculiaridad de este modelo es que uno de los regresores, y_{t-1}, ya no es independiente de $\hat{u}_{t-1}$, y el resultado (6.10) ya no se cumple. En este caso ya no puede utilizarse (6.11) para estimar ϕ sino el siguiente modelo:

$$\hat{u}_t = \lambda y_{t-1} + \beta X_t + \rho\hat{u}_{t-1} + v_t \qquad (6.12)$$

Utilizando este modelo llegaríamos a un estimador como el escrito en (6.9) pero incluyendo en la definición de M las dos variables, y_{t-1} y X_t.

La generalización al caso en el que en el modelo hay k regresores y la estructura del proceso autorregresivo es p es inmediata. Escribamos el modelo como:

$$y = X\beta + u \tag{6.13}$$

con $$u_t = \rho_1 u_{t-1} + \ldots + \rho_p u_{t-p} + v_t$$

en donde ahora X es una matriz $T \times k$ que puede incluir o no observaciones de valores retardados de y_t. β es un vector de k parámetros.

El modelo auxiliar generalización de (6.12), es:

$$\begin{aligned} \hat{u} &= X\beta + \hat{u}_{-1}\rho_1 + \ldots + \hat{u}_{-p}\rho_p + v = \\ &= X\beta + \hat{U}_{-p}\rho + v \end{aligned}$$

en donde ahora ϕ es un vector de p elementos. El estimador MCO de este vector viene dado por:

$$\hat{\phi} = (\hat{U}'_{-p} M \hat{U}_{-p})^{-1} \hat{U}'_{-p} M \hat{u}$$

La distribución asintótica de este estimador puede escribirse:

$$\hat{\rho} \sim N[0, \sigma^2 (\hat{U}'_{-p} M \hat{U}_{-p})^{-1}]$$

y el estadístico del contraste será:

$$LM = \hat{\rho}'(V\hat{a}r\,\hat{\phi})^{-1}\,\hat{\rho} = \frac{1}{\sigma^2}\hat{u}'\hat{U}_{-p}[\hat{U}'_{-p}M\hat{U}_{-p}]^{-1}\hat{U}'_{-p}\hat{u} \tag{6.14}$$

El parámetro σ^2 es desconocido y se sustituye por su estimador MCO definido a partir de (6.13). El estadístico escrito en (6.14) bajo H_0 y asintóticamente sigue una distribución χ^2 con p grados de libertad. Elegido un nivel de significación, ε_0, la región crítica del contraste viene dada por:

$$LM > \chi^2_{\varepsilon_0}(p)$$

Como se verá en la Sección 6.4, este es, precisamente, el contraste de los Multiplicadores de Lagrange.

6.2.3. Heteroscedasticidad

El término heteroscedasticidad denota una situación en la que la varianza de las perturbaciones no es constante a lo largo de las diferentes observaciones.

Como hemos comentado para la Autocorrelación, para contrastar la heteroscedasticidad hay que partir de un modelo y de un patrón concreto para la forma que adopta la heterogeneidad de la varianza. Existen tantos procedimientos de contraste como supuestos de partida se adopten. Nosotros, en esta sección, adoptaremos el enfoque de Breush-Pagan (1980) por ser uno de los que supone el marco más general. El modelo es el escrito en (6.13) y el patrón de heteroscedasticidad que adoptaremos es:

$$\sigma_t^2 = h(z_t' \alpha) \qquad t = 1, 2, ..., T$$

en donde h(.) es una función general independiente de t, $z_t' = (z_{1t}, ..., z_{pt})$ es un vector $p \times 1$ de la observación t-ésima de un conjunto de p variables exógenas que pueden o no estar relacionadas con los regresores y $\alpha' = (\alpha_1, \alpha_2, ..., \alpha_p)$ es un vector $p \times 1$ de parámetros. El primer elemento de z_t es la constante, es decir, $z_{1t} = 1$. $\forall$t. La hipótesis nula de homoscedasticidad toma la forma: $\alpha_2 = \alpha_3 = ... = \alpha_p = 0$.

Considerar ahora la regresión:

$$g_t = z_t' \alpha + v_t$$

en donde: $$g_t = \frac{\hat{u}_t^2}{\hat{\sigma}^2}$$

con $$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T}$$, siendo $\hat{u}$ el vector de residuos MCO.

Sea $\bar{g} = \frac{\sum_{t=1}^{T} g_t}{T}$ y $\hat{g}_t = z_t'\hat{\alpha}$. Además, vamos a definir el estadístico RSS = $\sum_{t=1}^{T} (\hat{g}_t - \bar{g})^2$. Breush y Pagan demuestran que, bajo la hipótesis nula, asintóticamente se tiene que:

$$\frac{1}{2} \text{ RSS} \sim \chi^2(p - 1)$$

Elegido un nivel de significación, ε_0, la región crítica del contraste viene dada por:

$$\frac{1}{2} \text{ RSS} > \chi^2_{\varepsilon_0}(p - 1)$$

Como se verá en la Sección 6.4 este contraste coincide con el de los Multiplicadores de Lagrange.

Como ya hemos indicado, este contraste es uno entre los muchos que han sido propuestos en la literatura. Para otros contrastes ver las referencias citadas en la Sección 6.1.

6.2.4. Contraste de normalidad

El hecho de que una perturbación aleatoria siga una distribución normal en torno a cero es muy deseable porque garantiza la simetría entre aportaciones positivas y negativas compensándose en promedio. Los dos aspectos más destacables de una distribución normal son la garantía de esta simetría y la existencia de una curvatura con pendientes suaves.

Precisamente son los estadísticos que dan cuenta de estos dos aspectos los que sirven de base para el procedimiento que se usa para contrastar la hipótesis de normalidad.

Para el vector de residuos MCO del modelo, se definen los estadísticos de asimetría (g_1) y de curtosis, o apuntamiento (g_2), como sigue:

$$g_1 = \frac{m_3}{m_2^{3/2}} \qquad g_2 = \frac{m_4}{m_2^2} - 3$$

en donde m_j es el momento central de orden j definido como:

$$m_j = \frac{\sum_{t=1}^{T} (\hat{u}_t - \bar{\hat{u}})^j}{T}$$

Bajo la hipótesis nula de que el vector de perturbaciones sigue una distribución normal, g_1 y g_2 se distribuyen, respectivamente, como $N(0, \frac{6}{T})$ y $N(0, \frac{24}{T})$. Supuesto un nivel de significación, se puede establecer un intervalo en torno a cero y decidir que no hay normalidad cuando uno de los dos estadísticos o los dos tome un valor que este fuera del intervalo respectivo.

Pero también puede definirse un estadístico que resuma la información de los dos de la forma siguiente:

$$LM_N = \frac{Tg_1^2}{6} + \frac{Tg_2^2}{24}$$

Bajo la hipótesis nula, este estadístico, asintóticamente, sigue una distribución χ^2 con dos grados de libertad.

Supuesto un nivel de significación la región crítica del contraste viene dada por:

$$LM_N > \chi^2_{\varepsilon_0}(2)$$

En la Sección 6.4 se verá que este contraste coincide con el de los Multiplicadores de Lagrange derivado por Bera y Jarque (1982).

6.3 EL CONTRASTE DE LOS MULTIPLICADORES DE LAGRANGE PARA UN SUBCONJUNTO DE PARÁMETROS.

Hemos visto en la sección anterior que la mayor parte de los contrastes examinados eran casos particulares del contraste de los Multiplicadores de Lagrange.

Esta sección la vamos a dedicar a examinar los rasgos esenciales del contraste de los Multiplicadores de Lagrange. No se pretende llevar a cabo un tratamiento exhaustivo sino completar algunos aspectos del contraste ya vistos en el Capítulo 3. Un tratamiento más completo puede verse en Engle (1984), Godfrey (1988) y Gourieroux y Monfort (1989).

Sean $y_1 ....y_t$ los T elementos de una muestra. Denotaremos por $L(\theta)$ la función de verosimilitud en donde θ es un vector de k parámetros. El gradiente y la matriz hessiana vienen dados por:

$$d(\theta) = \frac{\partial l(\theta)}{\partial \theta}$$

$$D(\theta)= \frac{\partial^2 l(\theta)}{\partial \theta \partial \theta'}$$

en donde $l(\theta)$ es el logaritmo de la función de verosimilitud.

La matriz de información para las T observaciones muestrales viene dada por:

$$I(\theta) = -E[D(\theta)]$$

Suponemos que la hipótesis nula es una hipótesis conjunta que toma la forma de r ($r<k$) restricciones lineales de los elementos de θ. Entonces, siendo R una matriz de constantes conocidas de orden rxk, con rango igual a r, y q un vector de constantes conocidas de orden rx1, la hipótesis nula puede escribirse como:

$$H_0 : R\theta - q = 0$$

Adoptaremos los supuestos habituales en la literatura; ver, por ejemplo, Godfrey (1988).

El estimador máximo-verosímil (MV) es aquél que cumple que:

$$l(\tilde{\theta}) = \sup_{\theta \in \Theta} l(\theta)$$

A partir de los supuestos comentados se demuestra que el estimador MV es consistente con la siguiente distribución asintótica:

$$\sqrt{T}\,(\tilde{\theta} - \theta) \underset{asi}{\sim} N[0, \lim T\, I(\theta)^{-1}]$$

El estimador MV restringido viene dado por la maximización de $l(\theta)$ sujeto a las restricciones comentadas.

La función lagrangiana es:

$$L = l(\theta) + \lambda' (R\theta\text{-}q)$$

Las condiciones de primer orden nos permiten escribir:

$$\frac{\partial L}{\partial \theta} = 0 \Rightarrow d(\tilde{\theta}_R) + R'\tilde{\lambda} = 0 \tag{6.15}$$

$$\frac{\partial L}{\partial \lambda} = 0 \Rightarrow R\,\tilde{\theta}_R - q = 0 \tag{6.16}$$

Hay que destacar que si las restricciones sólo afectan a los r últimos elementos de θ entonces (6.15) puede escribirse como:

$$d_1(\tilde{\theta}_R) = 0$$

$$d_2(\tilde{\theta}_R) + R'\tilde{\lambda} = 0$$

en donde:

$$d_1(\theta) = \frac{\partial l(\theta)}{\partial \theta_1} \quad \text{y} \quad d_2(\theta) = \frac{\partial l(\theta)}{\partial \theta_2}$$

De forma inmediata puede demostrarse que (Ver Godfrey (1988)):

$$\frac{1}{\sqrt{T}} \tilde{\lambda} \underset{asi}{\sim} N[0, \lim (RV(\theta)^{-1}R')^{-1}] \tag{6.17}$$

en donde: $V(\theta) = T^{-1} I(\theta)$

El contraste de los multiplicadores de Lagrange o contraste del gradiente está basado en cuan diferentes de cero son los elementos de λ o, equivalentemente, los elementos de $d(\theta)$. Si puede aceptarse que todos los elementos de λ son cero entonces eso se puede interpretar diciendo que las restricciones bajo H_0 no están forzando a los datos y que hay evidencia favorable para aceptar las restricciones.

Teniendo en cuenta (6.17) podemos escribir que bajo H_0:

$$LM = \frac{1}{\sqrt{T}} \tilde{\lambda}' RV(\theta)^{-1}R' \frac{\tilde{\lambda}}{\sqrt{T}} = \tilde{\lambda}' RI(\tilde{\theta}_R)^{-1}R' \tilde{\lambda} \tag{6.18}$$

sigue, asintóticamente, una distribución chi-cuadrado central con r grados de libertad. Elegido un nivel de significación, ε, la hipótesis nula se rechaza si

$$LM > \chi^2_\varepsilon(r)$$

Teniendo en cuenta (6.15), una forma equivalente de (6.18) es:

$$LM = d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R) \tag{6.19}$$

Veamos ahora la forma que toma el contraste cuando las restricciones sólo afectan a un subvector de θ.

Particionemos el vector θ de la siguiente manera:

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{6.20}$$

donde θ_1 es un vector de s elementos y θ_2 es un vector de r elementos, siendo $s + r = k$. La hipótesis nula que queremos contrastar es:

$$H_0 : \theta_2 = 0$$

Esta hipótesis nula es un caso particular con:

$$R = [0 \; I_r] \;\; y \;\; q = 0$$

en donde I_r es la matriz unidad de orden rxr.

Particionando de conformidad con (6.20) el resto de los vectores y matrices de interés, escribiremos:

$$\tilde{\theta} = \begin{bmatrix} \tilde{\theta}_1 \\ \tilde{\theta}_2 \end{bmatrix} ; \tilde{\theta}_R = \begin{bmatrix} \tilde{\theta}_{R1} \\ \tilde{\theta}_{R2} \end{bmatrix} = \begin{bmatrix} \tilde{\theta}_{R1} \\ 0 \end{bmatrix}$$

$$d(\theta) = \begin{bmatrix} \dfrac{\partial l(\theta)}{\partial \theta_1} \\ \dfrac{\partial l(\theta)}{\partial \theta_2} \end{bmatrix} = \begin{bmatrix} d_1(\theta) \\ d_2(\theta) \end{bmatrix}$$

$$I(\theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} ; I(\theta)^{-1} = \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix}$$

$$V(\theta) = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} ; V(\theta)^{-1} = \begin{bmatrix} V^{11} & V^{12} \\ V^{21} & V^{22} \end{bmatrix}$$

En este caso, la función lagrangiana viene dada por:

$$L = l(\theta) + \lambda'\theta_2$$

Las condiciones de primer orden son:

$$\frac{\partial L}{\partial \theta_1} = d_1(\theta) \Rightarrow d_1(\tilde{\theta}_R) = 0 \qquad (6.21)$$

$$\frac{\partial L}{\partial \theta_2} = d_2(\theta) + \lambda \Rightarrow d_2(\tilde{\theta}_R) + \tilde{\lambda} = 0 \qquad (6.22)$$

$$\frac{\partial L}{\partial \lambda} = \theta_2 \Rightarrow \tilde{\theta}_{R2} = 0 \qquad (6.23)$$

El contraste LM escrito en (6.18) toma ahora la forma siguiente:

$$LM = \tilde{\lambda}'\, I^{22}(\tilde{\theta}_R)\, \tilde{\lambda} \qquad (6.24)$$

en donde:

$$I^{22} = \left(I_{22} - I_{12}'\left(I_{11}\right)^{-1} I_{21}\right)^{-1}$$

Si la matriz de información es diagonal en bloques, es decir, si $I_{12}=I_{21}=0$, entonces (6.24) puede escribirse como:

$$LM = \tilde{\lambda}' I_{22}^{-1} \tilde{\lambda} \qquad (6.25)$$

Ejemplo 6.1

Considerar el siguiente MLG:

$$y = X\beta + Z\gamma + u = W\delta + u \qquad (6.26)$$

en donde y es Tx1, X y Z son, respectivamente, Txk y Txp y u es un vector de Tx1 perturbaciones que cumple las hipótesis habituales de no autocorrelación, homoscedasticidad, normalidad e independencia de los regresores. Es decir, $u \sim N(0, \sigma^2 I_T)$.

La hipótesis nula que se va a contrastar son las p restricciones:

$$H_0 : \gamma = 0 \qquad (6.27)$$

El logaritmo de la función de verosimilitud del modelo escrito en (6.26) toma la forma siguiente:

$$l(\theta) = - \frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} u'u$$

en donde:

$$\theta' = (\beta', \gamma', \sigma^2)$$

es un vector de k + p + 1 elementos.

El gradiente viene dado por:

$$\frac{\partial l(\theta)}{\partial \beta} = \frac{X'(y-X\beta-Z\gamma)}{\sigma^2} \qquad (6.28)$$

$$\frac{\partial l(\theta)}{\partial \gamma} = \frac{Z'(y-X\beta-Z\gamma)}{\sigma^2} \qquad (6.29)$$

$$\frac{\partial l(\theta)}{\partial \sigma^2} = \frac{u'u-T\sigma^2}{2\sigma^4} \qquad (6.30)$$

y la matriz de segundas derivadas:

$$D(\theta) = \begin{bmatrix} -\frac{X'X}{\sigma^2} & -\frac{X'Z}{\sigma^2} & -\frac{X'u}{\sigma^4} \\ & -\frac{Z'Z}{\sigma^2} & -\frac{Z'u}{\sigma^4} \\ & & \frac{T}{2\sigma^4} - \frac{u'u}{\sigma^6} \end{bmatrix}$$

Por lo tanto, la matriz de información es:

$$I(\theta) = \begin{bmatrix} \frac{X'X}{\sigma^2} & \frac{X'Z}{\sigma^2} & 0 \\ \frac{Z'X}{\sigma^2} & \frac{Z'Z}{\sigma^2} & 0 \\ 0 & 0 & \frac{T}{2\sigma^4} \end{bmatrix} \tag{6.31}$$

Se ve como $I(\theta)$ es diagonal en bloques siendo éstos los que corresponden a (β', γ') y σ^2.

Teniendo en cuenta (6.22), y sustituyendo los estimadores restringidos en (6.29) se tiene que:

$$\tilde{\lambda} = - d_2(\tilde{\theta}_R) = \frac{Z'\tilde{u}_R}{\tilde{\sigma}_R^2} \tag{6.32}$$

en donde:

$$\tilde{\sigma}_R^2 = \frac{\tilde{u}'_R\tilde{u}_R}{T}$$

$$\tilde{u}_R = y - X\tilde{\beta}_R$$

$$\tilde{\beta}_R = (X'X)^{-1}X'y$$

Teniendo en cuenta la forma que adopta la matriz de información escrita en (6.31), la submatriz $I^{22}(\tilde{\theta}_R)$ es:

$$I(\tilde{\theta}_R)^{22} = \tilde{\sigma}_R^2 \,[Z'Z-Z'X(X'X)^{-1}X'Z]^{-1} = \tilde{\sigma}_R^2\,(Z'M_XZ)^{-1} \tag{6.33}$$

A partir de (6.32) y (6.33) se puede escribir:

$$LM = \frac{1}{\tilde{\sigma}_R^2}\tilde{u}'_R Z(Z'M_xZ)^{-1}Z'\tilde{u}_R \qquad (6.34)$$

También podemos escribir (6.34) como:

$$LM = \frac{1}{\tilde{\sigma}_R^2} y'M_xZ(Z'M_xZ)^{-1}(Z'M_xZ)(Z'M_xZ)^{-1}Z'M_xy =$$

$$= \frac{1}{\tilde{\sigma}_R^2}\tilde{\gamma}'(Z'M_xZ)\tilde{\gamma} \qquad (6.35)$$

En donde $\tilde{\gamma}$ es el estimador MV de γ obtenido a partir del modelo general escrito en (6.26).

Por otra parte, hay que tener en cuenta que:

$$R\tilde{\delta} - q = R\tilde{\delta} - R\tilde{\delta}_R = \tilde{\gamma} - 0 = \tilde{\gamma}$$

ya que, en este caso:

$$R = [0, I_p]$$

por la forma que adopta H_0 en (6.27).

También se tiene que:

$$\frac{1}{\tilde{\sigma}_R^2}(Z'M_xZ) = \frac{1}{\tilde{\sigma}_R^2} R(W'W)^{-1}R'$$

Por lo tanto, el estadístico escrito en (6.35) puede escribirse como:

$$LM = \frac{1}{\tilde{\sigma}_R^2}(R\tilde{\delta} - q)'[R(W'W)^{-1}R']^{-1}(R\tilde{\delta} - q) \qquad (6.36)$$

que es idéntico al contraste de la F para la hipótesis nula (6.27), con la única excepción del estimador de la varianza de las perturbaciones que aparece en el denominador de (6.36) ya que en el contraste de la F aparece el estimador MV obtenido a partir de (6.26) en lugar del restringido.

6.4 CONTRASTES DE ESFERICIDAD A PARTIR DEL CONTRASTE DE LOS MULTIPLICADORES DE LAGRANGE.

En esta sección vamos a derivar estadísticos para contrastar autocorrelación, heteroscedasticidad y normalidad en la perturbación aleatoria de un modelo basándonos en el Principio de los Multiplicadores de Lagrange.

6.4.1. CONTRASTE DE AUTOCORRELACIÓN

Considerar el MLG:

$$y = X\beta + u \tag{6.37}$$

con

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{6.38}$$

en donde ε_t es ruido blanco. La hipótesis nula a contrastar es la ausencia de autocorrelación que se concreta en:

$$H_0 : \rho=0 \tag{6.39}$$

En una segunda fase adoptaremos un modelo autorregresivo más general:

$$u_t = \rho_1 u_{t-1}+........+\rho_p u_{t-p}+\varepsilon_t \tag{6.40}$$

siendo la hipótesis nula a contrastar:

$$H_0 : \rho_1 = \rho_2 = ... = \rho_p = 0 \tag{6.41}$$

La observación t-ésima del modelo escrito en (6.37) toma la forma siguiente:

$$y_t = \sum_{i=1}^{k} \beta_i X_{it} + u_t \tag{6.42}$$

Retardando (6.42) un periodo, multiplicando por ρ y restando a y_t se obtiene:

$$y_t-\rho y_{t-1} = \sum_{\iota=1}^{\kappa} \beta_i (X_{it} - \rho X_{it-1}) + \varepsilon_t \tag{6.43}$$

Para facilitar las derivaciones utilizaremos:

$$\tilde{y}_t = y_t-\rho y_{t-1} \quad y \quad \tilde{X}_t = X_t-\rho X_{t-1}$$

Notar que bajo la hipótesis nula:

$$\tilde{y}_t = y_t \quad y \quad \tilde{X}_t = X_t$$

La función de verosimilitud para las T observaciones puede escribirse como:

$$L(y/y_0,y_{-1},......,y_{-k+1}, \beta, \rho, \sigma_\varepsilon^2) = (2\pi\sigma_\varepsilon^2)^{-T/2} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\Sigma(u_t - \rho u_{t-1})^2\right\}$$

y su logaritmo:

$$l = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}\sum(u_t - \rho u_{t-1})^2$$

$$= -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}(u-\rho u_{-1})'(u-\rho u_{-1})$$

$$= -\frac{T}{2}\log 2\pi - \frac{T}{2}\log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2}(\tilde{y}-\tilde{X}\beta)'(\tilde{y}-\tilde{X}\beta)$$

en donde:

$$u = \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_T \end{bmatrix} \quad y \quad u_{-1} = \begin{bmatrix} u_0 \\ u_1 \\ . \\ . \\ u_{T-1} \end{bmatrix}$$

Sea $\theta' = (\beta', \rho, \sigma_\varepsilon^2)$ un vector de k+2 parámetros.

El gradiente puede escribirse como:

$$\underset{(k+2)\times 1}{d(\theta)} = \begin{bmatrix} \dfrac{\partial l}{\partial \beta} = \dfrac{\tilde{X}'(\tilde{y}-\tilde{X}\beta)}{\sigma_\varepsilon^2} = \dfrac{\tilde{X}'(u-\rho u_{-1})}{\sigma_\varepsilon^2} \\ \dfrac{\partial l}{\partial \rho} = \dfrac{u_{-1}'(u-\rho u_{-1})}{\sigma_\varepsilon^2} \\ \dfrac{\partial l}{\partial \sigma_\varepsilon^2} = -\dfrac{T}{2\sigma_\varepsilon^2} + \dfrac{1}{2\sigma_\varepsilon^4}[u-\rho u_{-1}]'[u-\rho u_{-1}] \end{bmatrix} \tag{6.44}$$

Los elementos de la matriz hessiana de segundas derivadas son:

$$
\underset{(k+2)\times(k+2)}{D(\theta)} = \begin{bmatrix} -\dfrac{\tilde{X}'\tilde{X}}{\sigma_\varepsilon^2} & -\dfrac{\tilde{X}'u_{-1}}{\sigma_\varepsilon^2} & -\dfrac{\tilde{X}'(u-\rho u_{-1})}{\sigma_\varepsilon^4} \\ & -\dfrac{u'_{-1}u_{-1}}{\sigma_\varepsilon^2} & -\dfrac{u'_{-1}(u-\rho u_{-1})}{\sigma_\varepsilon^4} \\ & & \dfrac{T}{2\sigma_\varepsilon^4} - \dfrac{(u-\rho u_{-1})'(u-\rho u_{-1})}{\sigma_\varepsilon^6} \end{bmatrix}
$$

Bajo H_0 se cumple que:

$$
u = \varepsilon,\ u_{-1} = \varepsilon_{-1},\ \sigma_u^2 = \sigma_\varepsilon^2 \ \text{ y } \ \tilde{X} = X \tag{6.45}
$$

por lo que bajo H_0 la matriz de información puede escribirse:

$$
-ED(\theta) = \begin{bmatrix} E\dfrac{X'X}{\sigma^2} & E\dfrac{X'u_{-1}}{\sigma^2} & E\dfrac{X'u}{\sigma^4} \\ & E\dfrac{u'_{-1}u_{-1}}{\sigma^2} & E\dfrac{u'_{-1}u}{\sigma^4} \\ & & E\left[-\dfrac{T}{2\sigma^4} + \dfrac{u'u}{\sigma^6}\right] \end{bmatrix} \tag{6.46}
$$

Suponemos que, contemporáneamente, los regresores no están correlacionados con las perturbaciones, [EX'u=0], aunque si es posible que alguno de ellos esté correlacionado con el primer retardo de éstas [$EX'u_{-1} \neq 0$].

Bajo la hipótesis nula, asintóticamente se tiene:

$$
-ED(\theta) = \frac{1}{\sigma^2}\begin{bmatrix} X'X & X'u_{-1} & 0 \\ u'_{-1}X & u'_{-1}u_{-1} & 0 \\ 0 & 0 & \dfrac{T}{2\sigma^2} \end{bmatrix} \tag{6.47}
$$

El contraste LM en general puede escribirse como:

$$
LM = d(\tilde{\theta}_R)'\, I(\tilde{\theta}_R)^{-1}\, d(\tilde{\theta}_R) \tag{6.48}
$$

en donde $I(\tilde{\theta}_R)$ es la matriz de información escrita en (6.47) evaluada con los estimadores restringidos.

$$\tilde{\theta}_R' = (\hat{\beta}_R', 0, \hat{\sigma}_R^2)$$

$$\hat{\sigma}_R^2 = \frac{\hat{u}'\hat{u}}{T}; \quad \hat{u} = y - X\hat{\beta}_R; \hat{\beta}_R = (X'X)^{-1}X'y$$

Teniendo en cuenta (6.44) se tiene que:

$$d(\tilde{\theta}_R) = \frac{1}{\hat{\sigma}_R^2}\begin{bmatrix} X'\hat{u} \\ \hat{u}'\hat{u}_{-1} \\ -\frac{T}{2}+\frac{T}{2} \end{bmatrix} = \frac{1}{\hat{\sigma}_R^2}\begin{bmatrix} 0 \\ \hat{u}'\hat{u}_{-1} \\ 0 \end{bmatrix} \tag{6.49}$$

El elemento correspondiente a ρ en la inversa de la matriz de información es:

$$I(\tilde{\theta}_R)^{-1} = \begin{bmatrix} \hat{\sigma}_R^2[\hat{u}_{-1}'\hat{u}_{-1} - \hat{u}_{-1}'X(X'X)^{-1}X'\hat{u}_{-1}]^{-1} \end{bmatrix} \tag{6.50}$$

Sustituyendo (6.49) y (6.50) en (6.48) se obtiene:

$$LM = \frac{1}{\hat{\sigma}_R^2}\hat{u}'\hat{u}_{-1}\left[\hat{u}'_{-1}\hat{u}_{-1} - \hat{u}'_{-1}X(X'X)^{-1}X'\hat{u}'_{-1}\right]^{-1}\hat{u}'_{-1}\hat{u} \tag{6.51}$$

En la sección anterior hemos mostrado que, bajo H_0: $\rho=0$, el estadístico LM sigue una distribución χ^2 con un grado de libertad. La región crítica del contraste será:

$$LM > \chi^2_{\varepsilon_0}(1)$$

siendo $\chi^2_{\varepsilon_0}(1)$ el punto crítico correspondiente al nivel de significación previamente elegido, ε_0.

En el caso en que la hipótesis nula fuera la escrita en (6.41), entonces el contraste de los Multiplicadores de Lagrange sería:

$$LM = \frac{1}{\hat{\sigma}_R^2}\hat{u}'\hat{U}_{-p}[\hat{U}'_{-p}\hat{U}_{-p} - \hat{U}'_{-p}X(X'X)^{-1}X'\hat{U}_{-p}]^{-1}\hat{U}_{-p}'\hat{u} \tag{6.52}$$

en donde:

$$\hat{U}_{-p} = [\hat{u}_{-1}, \hat{u}_{-2}, ..., \hat{u}_{-p}]$$

En este caso los grados de libertad serían p por lo que la región crítica vendría dada por:

$$LM > \chi^2_{\varepsilon_0}(p)$$

Veamos ahora como el contraste derivado en (6.51) puede interpretarse como el contraste de la hipótesis nula H_0: $\gamma=0$, en el modelo:

$$y = X\beta + Z\gamma + u \qquad (6.53)$$

en donde $Z = \hat{U}_{-p}$

Para contrastar la hipótesis nula H_0: $\gamma = 0$, hay que tener en cuenta que los estimadores MCO de γ en (6.53) son:

$$\hat{\gamma} = (Z'M_XZ)^{-1}Z'M_Xy \qquad (6.54)$$

por lo que:

$$\hat{\gamma} \sim N(\gamma, \sigma^2(Z'M_XZ)^{-1}) \qquad (6.55)$$

Bajo H_0 se tiene que:

$$\hat{\gamma}' (Var(\hat{\gamma}))^{-1}\hat{\gamma} \sim \chi^2(p) \qquad (6.56)$$

o bien:

$$\frac{1}{\sigma^2} y'M_XZ(Z'M_XZ)^{-1}(Z'M_XZ)(Z'M_XZ)^{-1}Z'M_Xy =$$

$$= \frac{1}{\sigma^2} y'M_XZ(Z'M_XZ)^{-1}Z'M_Xy \qquad (6.57)$$

Teniendo en cuenta que $M_Xy=\hat{u}$, basta definir $Z = \hat{U}_{-p}$ y sustituir a σ^2 por $\hat{\sigma}^2$ (asintóticamente no supone ningún cambio) para derivar (6.52) a partir de (6.57).

6.4.2. CONTRASTE DE HETEROSCEDASTICIDAD

Partiremos del MLG escrito como:

$$y_t = X_t'\beta + u_t \qquad (6.58)$$

$$t = 1,2,....,T$$

en donde β es un vector kx1 de parámetros y las perturbaciones u_t se distribuyen normal e independientemente con media cero y varianza igual a:

$$\sigma_t^2 = h(Z_t'\alpha) = h(S_t) \tag{6.59}$$

en donde h(.) es una función no dependiente del tiempo, que posee primera y segunda derivadas, α es un vector de px1 coeficientes no relacionados funcionalmente con los β siendo el primer elemento de Z_t la unidad.

La hipótesis nula de homoscedasticidad toma la forma siguiente:

$$H_0 : \alpha_2 = \alpha_3 = ... = \alpha_p = 0 \tag{6.60}$$

en cuyo caso $Z_t'\alpha = \alpha_1$ de forma que $\sigma_t^2 = h(\alpha_1) = \sigma^2$ es constante para todo t.

La representación en (6.59) es lo suficientemente general para dar cabida a la mayoría de los modelos heteroscedásticos contemplados en la literatura. Por ejemplo, tienen cabida representaciones como:

$$\sigma_t^2 = \exp(Z_t'\alpha)$$

ó

$$\sigma_t^2 = (Z_t'\alpha)^m$$

si llamamos $\theta'=(\beta',\alpha')$, el logaritmo de la función de verosimilitud puede escribirse como:

$$l(\theta) = -\frac{T}{2}\log 2\pi - \frac{1}{2}\sum \log \sigma_t^2 - \frac{1}{2}\sum \frac{(y_t - X_t'\beta)^2}{\sigma_t^2} \tag{6.61}$$

Los elementos del gradiente vienen dados por:

$$d_\beta(\theta) = \frac{\partial l(\theta)}{\partial \beta} = \sum \frac{X_t'u_t}{\sigma_t^2}$$

$$d_\alpha(\theta) = \frac{\partial l(\theta)}{\partial \alpha} = -\frac{1}{2}\sum \frac{h'(S_t)Z_t}{\sigma_t^2} + \frac{1}{2}\sum \frac{h'(S_t)Z_t u_t^2}{\sigma_t^4} =$$

$$= \frac{1}{2}\sum h'(S_t)Z_t(\sigma_t^{-4}u_t^2 - \sigma_t^{-2}) \tag{6.62}$$

Respecto a las segundas derivadas, hay que tener en cuenta que:

$$\frac{\partial l(\theta)}{\partial\alpha\partial\beta'} = - \sum \frac{h'(S_t)Z_tX_t'u_t}{\sigma_t^4}$$

La esperanza matemática de esta expresión es igual a cero de forma que la matriz de información es diagonal en bloques. Por lo tanto, sólo tiene interés la matriz de segundas derivadas respecto a α.

Esta matriz viene dada por:

$$\frac{\partial l(\theta)}{\partial\alpha\partial\alpha'} = - \frac{1}{2} \sum \left[\frac{h''(S_t)Z_tZ_t'\sigma_t^2 - h'(S_t)Z_tZ_t'h'(S_t)}{\sigma_t^4}\right] +$$

$$+ \frac{1}{2} \sum \left[\frac{h''(S_t)Z_tZ_t'u_t^2\sigma_t^4 - h'^2(S_t)Z_tZ_t'u_t^2(2\cdot\sigma_t^2)}{\sigma_t^8}\right] \qquad (6.63)$$

Tomando esperanzas de (6.63), el primer término de la derecha queda como está, mientras que el segundo queda como:

$$\frac{1}{2} \sum \left[\frac{h''(S_t)Z_tZ_t'\sigma_t^6 - 2h'^2(S_t)Z_tZ_t'\sigma_t^4}{\sigma_t^8}\right] = \frac{1}{2} \sum \left[\frac{h''(S_t)Z_tZ_t'\sigma_t^2 - 2h'^2(S_t)Z_tZ_t'}{\sigma_t^4}\right]$$

Por lo tanto, (6.63) bajo la hipótesis nula [$\sigma_t^2 = h(\alpha_1) = \sigma^2$; $h'(S_t) = h'(\alpha_1)$] y evaluada con los estimadores restringidos, puede escribirse:

$$I_{\alpha\alpha} = \tfrac{1}{2}\left[\hat{\sigma}^{-4} h'^2 \left(\hat{\alpha}_1\right)\right] \sum Z_t Z_t' \qquad (6.64)$$

Por otra parte, (6.62) puede escribirse como:

$$d_\alpha\left(\tilde{\theta}_R\right) = \frac{1}{2}\left[\hat{\sigma}^{-2} h'\left(\hat{\alpha}_1\right)\right] \sum Z_t \left(\hat{\sigma}^{-2} \hat{u}_t^2 - 1\right) \qquad (6.65)$$

Teniendo en cuenta que, en este caso, también se cumple que:

$$\tilde{\lambda} = - d_\alpha(\tilde{\theta}_R)$$

la expresión (6-25) para el contraste de los multiplicadores de Lagrange puede escribirse como:

$$LM = d_\alpha(\tilde{\theta}_R)' I_{\alpha\alpha}^{-1} d_\alpha(\tilde{\theta}_R) \qquad (6.66)$$

y teniendo en cuenta (6.64) y (6.65):

$$LM = \frac{1}{2}(\sum Z_t f_t)'\left(\sum Z_t Z_t'\right)^{-1}(\sum Z_t f_t) = \frac{1}{2} f'Z\left(Z'Z\right)^{-1} Z'f \qquad (6.67)$$

en donde $f_t = (\hat{\sigma}^{-2}\ \hat{u}_t^2 - 1) = g_t - 1$

En forma vectorial: f = g-i, cumpliéndose que:

$$i'g = \frac{\Sigma \hat{u}_t^2}{\hat{\sigma}^2} = T \ \frac{\hat{\sigma}^2}{\hat{\sigma}^2} = T$$

$$i'f = i'g - i'i = 0$$

Como consecuencia de estos resultados se obtiene que:

f'Z es un vector de p elementos tal que el primero de ellos es cero porque la primera columna de Z es un vector con todos los elementos iguales a 1.

También se tiene que:

$$(Z'Z)^{-1}Z'i = \begin{pmatrix} 1 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \qquad (6.68)$$

Por lo tanto:

$$f'Z(Z'Z)^{-1}Z'i = (0......)\begin{pmatrix} 1 \\ 0 \\ . \\ . \\ 0 \end{pmatrix} = 0$$

de forma que, (6.67) puede escribirse como:

$$LM = \frac{1}{2}\ f'Z(Z'Z)^{-1}Z'g = \frac{1}{2}\ [g'Z(Z'Z)^{-1}Z'g - \frac{1}{T}(i'g)^2] \qquad (6.69)$$

pues teniendo en cuenta (6.68) se tiene que:

$$-i'Z(Z'Z)^{-1}Z'g \;=\; -(1\ 0\ .\ .\ .\ 0)\begin{pmatrix} i'g \\ z_2'g \\ . \\ . \\ z_p'g \end{pmatrix} = i'g$$

y por ser i'g = T se obtiene (6.69).

Consideremos la regresión de g sobre Z. Los valores ajustados de esta regresión vienen dados por:

$$\hat{g} = Z(Z'Z)^{-1}Z'g$$

y la suma de los cuadrados de las desviaciones de estos valores respecto a sus medias vienen dados por:

$$\sum (\hat{g}_t - \hat{g})^2 = g'Z(Z'Z)^{-1}Z'g - \frac{1}{T}(i'g)^2$$

que es dos veces la expresión escrita en (6.69).

Por lo tanto, el contraste LM para la hipótesis nula escrita en (6.60) puede calcularse como la mitad de la suma de cuadrados de los valores ajustados respecto a sus medias de la regresión de g sobre las Z.

El estadístico LM se distribuye como una chi-cuadrado con (p-1) grados de libertad, por lo que la región crítica del contraste será: LM > χ^2(p-1).

6.4.3. CONTRASTE DE NORMALIDAD

Comenzaremos recordando que la función de densidad de una variable normal es la solución a:

$$\frac{1}{f} \cdot \frac{df}{dX} = \frac{(X-\mu)}{\sigma^2} \qquad (6.70)$$

de forma que:

$$\ln f = \int \frac{-(X-\mu)}{\sigma^2} dx + c \qquad (6.71)$$

De la misma forma, podemos decir que la función de densidad de la familia de Pearson viene dada por:

$$\frac{1}{f} \cdot \frac{df}{dX} = \frac{(X-a)}{(c_0+c_1X+c_2X^2)}$$

Entonces, podemos decir que la distribución normal es un caso particular de la familia de Pearson para $c_1 = c_2 = 0$.

Por lo tanto, se parte de la familia de Pearson y se contrasta la hipótesis nula H_0: $c_1 = c_2 = 0$.

Si la perturbación aleatoria del modelo sigue una distribución de la familia de Pearson su distribución de probabilidad puede escribirse como:

$$g(\varepsilon_t) = \exp\,[\phi(\varepsilon_t)] \,/ \int_{-\infty}^{+\infty} \exp[\phi(\varepsilon_t)]d\varepsilon_t$$

en donde:

$$\phi(\varepsilon_t) = \int \frac{(c_1-\varepsilon_t)}{\sigma^2-c_1\varepsilon_t+c_2\varepsilon_t^2}\, d\varepsilon_t$$

en donde se ha escrito $c_0 = \sigma^2$.

El logaritmo de la función de verosimilitud para las T observaciones es:

$$l(\theta) = -\sum \ln\left[\int_{-\infty}^{\infty} \exp\phi(\varepsilon_t)d\varepsilon_t\right] + \sum_{1}^{T}\phi(\varepsilon_t)$$

Los elementos del gradiente serán:

$$\frac{\partial l(\theta)}{\partial c_1} = -\sum \frac{1}{G_t}\int_{-\infty}^{\infty}[\exp\phi(\varepsilon_t)]\left[\int \frac{V_{t^-}(c_1-\varepsilon_t)(-\varepsilon_t)}{V_t^2}d\varepsilon_t\right]d\varepsilon_t +$$

$$+\sum \int \frac{V_{t^-}(c_1-\varepsilon_t)(-\varepsilon_t)}{V_t^2}d\varepsilon_t \qquad (6.72)$$

en donde:

$$V_t = \sigma^2 - c_1\varepsilon_t + c_2\varepsilon_t^2$$

$$G_t = \int_{-\infty}^{\infty} \exp\phi(\varepsilon_t)d\varepsilon_t$$

$$\frac{\partial l(\theta)}{\partial c_2} = - \sum \frac{1}{G_t} \int_{-\infty}^{\infty} [\exp\phi(\varepsilon_t)] \left[\int \frac{-(c_1-\varepsilon_t)\varepsilon_t^2}{V_t^2} d\varepsilon_t \right] + \sum \int \frac{-(c_1-\varepsilon_t)\varepsilon_t^2}{V_t^2} d\varepsilon_t$$

(6.73)

Bajo H_0 y dada la definición de G_t se tiene que:

$$G_t = \int_{-\infty}^{\infty} \exp[\frac{-\varepsilon_t^2}{2\sigma^2}]d\varepsilon_t = (2\pi\sigma^2)^{1/2} \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-1/2} \exp \left\{ \frac{-\varepsilon_t^2}{2\sigma^2} \right\} d\varepsilon_t = (2\pi\sigma^2)^{1/2}$$

Bajo H_0 también se tiene que:

$$\int \frac{V_{t^-} (c_1-\varepsilon_t)(-\varepsilon_t)}{V_t^2} d\varepsilon_t = \int \frac{\sigma^2 - \varepsilon_t^2}{\sigma^4} d\varepsilon_t = \frac{\sigma^2\varepsilon_t - \frac{\varepsilon_t^3}{3}}{\sigma^4}$$

Por lo tanto, (6.72) puede escribirse como:

$$-\sum_1^T \left(2\pi\sigma^2\right)^{-1/2} \int_{-\infty}^{\infty} e^{-\frac{\varepsilon_t^2}{\sigma^2}} \frac{\sigma^2\varepsilon_t - \frac{\varepsilon_t^3}{3}}{\sigma^4} d\varepsilon_t +$$

$$+\sum \frac{\sigma^2\varepsilon_t - \frac{\varepsilon_t^3}{3}}{\sigma^4} = \frac{1}{\sigma^2}\sum_1^T \varepsilon_t - \frac{\sum \varepsilon_t^3}{3\sigma^4} \qquad (6.74)$$

Teniendo en cuenta que:

$$\int \left(2\pi\sigma^2\right)^{-1/2} e^{-\varepsilon_t^2} / \sigma^4 \varepsilon_t d\varepsilon_t = 0$$

$$\int \left(2\pi\sigma^2\right)^{-1/2} e^{-\varepsilon_t^2} / \sigma^2 \varepsilon_t^3 d\varepsilon_t = 0$$

Bajo H_0 se tiene que:

$u_t = \varepsilon_t$ siendo u_t la perturbación del Modelo Lineal General.

Introduciendo en (6.74) las estimaciones restringidas, la estimación del primer elemento del gradiente viene dada por:

$$\left(\frac{\partial l(\theta)}{\partial c_1}\right)_{\theta=\tilde{\theta}_R} = \frac{1}{\tilde{\sigma}_R^2}\sum \hat{u}_t - \frac{1}{3\tilde{\sigma}_R^4}\sum \hat{u}_t^3 = T\left[\frac{\hat{\mu}_1}{\hat{\mu}_2} - \frac{\hat{\mu}_3}{3\hat{\mu}_2^2}\right]$$

en donde: $\hat{\mu}_j = \dfrac{\sum \mu_t^j}{T}$.

Por otra parte, hay que tener en cuenta que:

$$\int \frac{-(c_1-\varepsilon_t)\varepsilon_t^2}{V_t^2}\, d\varepsilon_t \overset{\text{bajo } H_0}{=} \int \frac{u_t^3}{\sigma^4}\, du = \frac{u_t^4}{4\sigma^4}$$

Por lo tanto, (6.73) puede escribirse como:

$$-\sum \frac{E\varepsilon_t^4}{4\sigma^4} + \sum \frac{\varepsilon_t^4}{4\sigma^4} \overset{\text{bajo } H_0}{=} \sum\left[-\frac{3}{4} + \frac{\hat{u}_t^4}{4\hat{\mu}_2^2}\right] = T\left[\frac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \frac{3}{4}\right]$$

teniendo en cuenta que, para una Normal:

$$\frac{\mu^4}{\sigma^4} = 3$$

Obteniendo las segundas derivadas respecto a los parámetros se obtiene que la submatriz que corresponde a c_1 , c_2 en la inversa de la matriz de información evaluada con los estimadores restringidos es:

$$\begin{bmatrix} \dfrac{3\hat{\mu}_2}{2T} & 0 \\ 0 & \dfrac{2}{3T} \end{bmatrix}$$

Por lo que el contraste de los multiplicadores de Lagrange viene dado por:

$$\begin{bmatrix} -T\dfrac{\hat{\mu}_3}{3\hat{\mu}_2^2} & T\left(\dfrac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \dfrac{3}{4}\right) \end{bmatrix} \begin{bmatrix} \dfrac{3\hat{\mu}_2}{2T} & 0 \\ 0 & \dfrac{2}{3T} \end{bmatrix} \begin{bmatrix} -T\dfrac{\hat{\mu}_3}{3\hat{\mu}_2^2} \\ T\left(\dfrac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \dfrac{3}{4}\right) \end{bmatrix} =$$

$$= \left[-T\frac{\hat{\mu}_3}{3\hat{\mu}_2^2} \, x \, \frac{3\hat{\mu}_2}{2T} \quad T\left(\frac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \frac{3}{4}\right)\frac{2}{3T} \right] \left[\begin{array}{c} -T\dfrac{\hat{\mu}_3}{3\hat{\mu}_2^2} \\ T\left(\dfrac{\hat{\mu}_4}{4\hat{\mu}_2^2} - \dfrac{3}{4}\right) \end{array} \right] =$$

$$= T \, \frac{\hat{\mu}_3^2}{6\hat{\mu}_2^3} + \frac{T}{24}\left(\frac{\hat{\mu}_4}{\hat{\mu}_2^2} - 3\right)^2$$

y éste es el contraste de Jarque y Bera.

EJERCICIOS

6.1). En el modelo con k regresores:

$$y = X\beta + u$$

se va a contrastar la hipótesis nula de no autocorrelación frente a la alternativa de un proceso autoregresivo de segundo orden .

Para cinco observaciones, el vector de residuos MCO es:

$$\hat{u}' = (5\ \ 0\ \ 1\ \ -3\ \ -3\ )$$

A continuación, se hacen las regresiones de este vector,retardado uno y dos periodos, sobre los k regresores del modelo resultando los siguientes vectores de residuos MCO :

$$\hat{u}_1' = (1\ \ 0\ \ -1\ \ 0\ \ 0\ )$$

$$\hat{u}_2' = (\ 0\ \ 2\ \ 0\ \ -1\ \ -1\ )$$

1). Calcular el valor que toma el estadístico de Durbin-Watson e indicar como se utilizaria para contrastar la hipótesis nula.

2). En el marco del contraste de los Multiplicadores de Lagrange

$$\text{LM}=\ d(\tilde{\theta}_R)'\, I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R)$$

escribir los elementos de θ y la forma que adopta $\tilde{\theta}_R$.

3). Para el contraste de los Multiplicadores de Lagrange formular la región crítica y, utilizando la información comentada anteriormente, concluir aceptando o rechazando la hipótesis nula . Recordar que para ε=0.05 el punto crítico para una variable χ^2 con dos grados de libertad es 6.

6.2). Considerar el modelo:

$$y = X\beta + u$$

en donde X tiene k regresores.La esperanza de u es 0 y se quiere contrastar la hipótesis nula de homoscedasticidad frente a:

$$\sigma_t^2 = (\alpha_1 + \alpha_2 z_t)^2$$

utilizando el contrasta de los Multiplicadores de Lagrange:

$$\text{LM}= d\left(\tilde{\theta}_R\right)' I\left(\tilde{\theta}_R\right)^{-1} d\left(\tilde{\theta}_R\right)$$

1). Determinar los elementos del vector θ y escribir la forma que toma el vector $\tilde{\theta}_R$ en función de los datos.

2). Indicar como calcularia el estadístico LM utilizando una regresión auxiliar.

3). Establecer la región crítica del contraste utilizando el estadístico derivado en 2).

6.3). Para el modelo:

$$y_t = \phi y_{t-1} + \beta x_t + u_t \qquad |\phi| < 1$$

Se va a contrastar la hipótesis nula de no autocorrelación frente a la alternativa de un proceso autoregresivo de segundo orden, utilizando el procedimiento de los Multiplicadores de Lagrange. Se pide:

1). El número de elementos del gradiente indicando cuales de ellos seran cero en dos situaciones: sustituyendo los parámetros por los estimadores restringidos y cuando se utilizan los estimadores sin restringir.

2). Escribir la forma que toma el estadístico de los Multiplicadores de Lagrange indicando como se obtiene cada uno de sus elementos en función de las observaciones muestrales.Escribir la forma que adopta la región crítica

3). Indicar los tamaños del error tipo 1 y tipo2 que se cometen cuando se utiliza este contraste.

4).Demostrar que el contraste de la hipótesis nula de no autocorrelación utilizando los Multiplicadores de Lagrange equivale a contrastar la hipótesis nula $: H_0 : \delta = 0$ en el modelo:

$$y_t = \phi y_{t-1} + \beta x_t + \delta z_t + u_t$$

Especificar la forma que debe adoptar z_t.

6.4). Suponer que para el modelo:

$$y_t = \beta t + u_t \qquad t=1, 2, 3, 4, 5.$$

el vector de residuos MCO es el siguiente:

$$\hat{u}' = (\ 2, -4, 3, 1, -2\ )$$

1). Se pide contrastar la hipótesis nula de no autocorrelación frente a la alternativa de un proceso autorregresivo de orden dos utilizando el procedimiento de los Multiplicadores de Lagrange. Para el cálculo de este estadístico utilizar la siguiente aproximación:

$$[\hat{U}'_p \hat{U}_p - \hat{U}_p' X (X'X)^{-1} X' \hat{U}_p] = \begin{pmatrix} 29 & -18 \\ -18 & 29 \end{pmatrix}$$

Definir $\hat{U}_p$ y $\hat{U}'_P X$ utilizando los datos del problema.

2). Utilizando el mismo contraste, contrastar la hipótesis nula de homoscedasticidad frente a una alternativa del siguiente tipo: $\sigma_t^2 = (\alpha_0 + \alpha_1 t)^2$

3). Utilizando el mismo procedimiento cotrastar la hipótesis nula de normalidad.

6.5). Un investigador intenta estimar el parámetro β en la relación:

$$y_t = \beta x_t + u_t$$

El investigador sospecha que en la observación t_0 existe un atípico en la variable dependiente de forma que en esa observación el modelo es:

$$y_{t_0} = \delta + \beta x_{t_0} + u_{t0}$$

Para estimar el parámetro se piensa en una de las siguientes estrategias.

A). Se estima β sustituyendo el valor de y en t_0 por 0.
B). Se estima β sustituyendo el valor de x en t_0 por 0.
C). Se estima β sustituyendo los valores de x e y en t_0 por 0.
D). Se estima β añadiendo como regresor una variable ficticia que toma el valor 1 en la observación t_0 y cero en el resto de las observaciones.

Se pide:

1). Definir los estimadores de las cuatro estrategias, demostrando si existe alguna coincidencia entre algunos de los estimadores.

2). Obtener la esperanza y varianza de dichos estimadores.

3). Comparar los resultados en 2). con las propiedades del estimador MCO aplicado al modelo original sin modificar los datos.

4). Razonar cual de las cuatro estrategias debería recomendarse e indicar si podría pensarse en otra estrategia que fuera más satisfactoria diferente de las cuatro comentadas.

6.6). Suponer que para el modelo:

$$y_t = \phi y_{t-1} + \beta x_t + u_t$$

se va a cotrastar la hipótesis nula de no autocorrelación frente a la alternativa del siguiente modelo:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$$

Para ello se va a utilizar el contraste de los Multiplicadores de Lagrange cuyo estadístico es:

$$LM= d(\tilde{\theta}_R)' I(\tilde{\theta}_R)^{-1} d(\tilde{\theta}_R)$$

Se pide:

1). Las ventajas e inconvenientes de utilizar el prodedimiento LM frente al contraste de Durbin-Watson.

2). Escribir los elementos del vector $\tilde{\theta}_R$ en función de momentos muestrales.

3). Indicar qué elementos del gradiente evaluados con los estimadores restringidos serán cero.

4). Indicar bajo qué condiciones se puede derivar la distribución de probabilidad de LM. Especificar como se calcularían los tamaños de los dos errores.

5). Suponer que $\hat{u}'\hat{u}$=12, y que $\hat{u}'\hat{U}_{-p}$=(6 0) con T=10. Indicar como deberían ser los elementos que faltan del estadístico LM para que la hipótesis nula fuera rechazada, sabiendo que el punto crítico es igual a 12.59.

6.7). Supongamos el modelo lineal:

$$y_t = \beta x_t + u_t$$

con $E u_t = 0$ $\quad Cov(u_t\, u_{t'}) = 0$ $\quad Var(u_t) = \sigma_t^2$

1). Definir el estimador MCO de β y derivar su media y su varianza.

2). Si suponemos que: $\sigma_t^2 = \sigma^2 z_t^2$ y dividimos el modelo original por z_t se obtiene:

$$\frac{y_t}{z_t} = \beta \frac{x_t}{z_t} + \frac{u_t}{z_t}$$

Aplicar MCO a este nuevo modelo derivar la media y varianza del nuevo estimador de β y compararlo con el obtenido en 1).

3). Suponer que, para estimar la varianza del estimador MCO de β, se utiliza la siguiente expresión:

$$\hat{Var}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum x_t^2}$$

Derivar y comentar las propiedades de este estimador.

6.8). Suponer que dos investigadores utilizan el modelo lineal general con k regresores pero los resultados que obtienen indican que el modelo sufre algún problema de especificación. El primero de los investigadores cree que el problema reside en la autocorrelación por lo que propone contrastar que no existe autocorrelación frente a la alternativa de un proceso de primer orden. El segundo investigador cree que el problema reside en haber omitido una variable relevante por lo que propone contrastar si su coeficiente es o no significativo en un modelo extendido.

1). Indicar como se llevarían a cabo ambos contrastes.

2). Suponiendo que el segundo investigador tuviera razón, ¿Podría obtener el primero una potencia similar a la que alcanza el segundo en su proceso de contraste?. Comentar en que casos podría darse esa coincidencia.

3). A la luz de los resultados anteriores, comentar la siguiente afirmación de Davidson y MacKinnon (1985): "Este ejemplo muestra con claridad que nunca debemos rechazar una hipótesis en favor de la alternativa que estamos contrastando. Cuando obtenemos un valor significativo en un contraste de autocorrelación o para cualquier otro tipo de contraste de falta de especificación, lo único que aprendemos es que nuestra hipótesis nula es muy improbable que sea correcta. Pero no podemos decir nada respecto al modelo que realmente generó los datos".