Documento de Trabajo

8 9 0 7

CONTRASTES M Y DE LA MATRIZ
DE INFORMACION DINAMICA
CON UNA APLICACION A REGRESION LINEAL

Teodosio Pérez Amaral

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES.- UNIVERSIDAD COMPLUTENSE
Campus de Somosaguas. 28023 - MADRID

CONTRASTES M Y DE LA MATRIZ DE INFORMACION DINAMICA CON UNA APLICACION A REGRESION LINEAL

Teodosio Pérez Amaral

Universidad Complutense de Madrid,
Departamento de Economía Cuantitativa
y Servicio de Estudios del Banco de España

Versión preliminar. Se agradecen los comentarios.

Resumen

En este trabajo se demuestra que los contrastes m (basados en momentos) son, bajo condiciones generales, contrastes de los Multiplicadores de Lagrange (LM).

También obtenemos contrastes m que son fácilmente computables e interpretables. Destacamos una manera simple de computar los contrastes LM bajo heteroscedasticidad condicional de forma desconocida y/o de no diagonalidad por bloques de la Matriz de Información.

Aplicamos estos resultados al caso de regresión lineal y obtenemos contrastes m y LM que son consistentes a heteroscedasticidad condicional.

Algunos de estos contrastes podrían tener un buen comportamiento en pequeñas muestras.

Proponemos contrastes de especificación de la matriz de información dinámica que detectarían forma funcional incorrecta y/o una estructura dinámica insuficientemente rica. Estos contrastes se pueden considerar como contrastes de autocorrelación serial que varía sistemáticamente, de efectos ARCH en media que varían sistemáticamente y efectos ARCH no simétricos que varían sistemáticamente.

También proponemos contrastes de autocorrelación, efectos ARCH en media y restricciones de factores comunes que son válidos bajo heteroscedasticidad condicional de forma desconocida. Todos los contrastes se pueden computar con una regresión de mínimos cuadrados ordinarios.

Identificamos una alternativa general con respecto a la cual se puede considerar que los contrastes de la matriz de información dinámica son contrastes LM. Esta alternativa incluye muchos modelos conocidos y algunos nuevos. Esta alternativa sugiere, además, generalizaciones del modelo de regresión lineal que pueden ser útiles para predicción.

En este trabajo damos varias versiones asintóticamente equivalentes de los contrastes. Algunas de ellas deberían tener un buen comportamiento en pequeñas muestras. En las primeras secciones se prueban teoremas que son útiles en situaciones mucho más generales que las que se consideran específicamente en este trabajo.

NOTA BIOGRAFICA

Teodosio Pérez Amaral es Catedrático Interino de Econometría del Departamento de Análisis Económico II de la Facultad de Ciencias Económicas y Empresariales de la Universidad Complutense de Madrid, Titulado del Servicio de Estudios del Banco de España, Doctor en Ciencias Económicas, Master de Economía por la Universidad de California en San Diego y Licenciado en Ciencias Económicas. Ha sido profesor del Centro de Formación del Banco de España, del Centro de Estudios Monetarios y Financieros del Banco de España (CEMFI) y profesor ayudante de la Universidad de California en San Diego.

1. Introducción y motivación

En los últimos años se ha dedicado considerable atención al estudio de los contrastes m de especificación. Los trabajos que han estudiado estos contrastes son los de Newey (1985), Tauchen (1985) y White (1985). White propone una clase particular de contrastes m llamados contrastes de la matriz de información dinámica.

En este trabajo abordamos muchas de las cuestiones que esta literatura reciente ha dejado sin resolver. Las más importantes son: ¿Qué es lo que contrastan los contrastes m? ¿Cuál es el poder de los contrastes m? y ¿Cuál es la relación entre contrastes m y los contrastes de los multiplicadores de Lagrange? También proponemos nuevos contrastes y damos versiones simplificadas de otros; algunas de ellas posiblemente tendrán un comportamiento mejor en pequeñas muestras que otras versiones consideradas.

Expliquemos primero intuitivamente qué se entiende por contrastes m. Un conjunto de medidas de fenómenos económicos se puede considerar como la realización de un proceso estocástico sobre un espacio de probabilidad $(\Omega,\ F,\ P_0)$. Tratamos de aprender acerca del verdadero modelo P_0 usando una colección de medidas de probabilidad tentativas P, indiciadas por un conjunto de parámetros $\theta \in \Theta$. Si un modelo P está correctamente especificado esto implicará, en general, la validez de un número (quizás infinito) de condiciones de momentos de la forma

[1.1] $E_\theta\ (m_t\ (\omega,\ \theta)) = 0, \qquad t = 1,2,\ldots,$

donde m_t: $\Omega \times \Theta \to \mathbb{R}^p$ con $\omega \in \Omega$. Si para todo $\theta \in \Theta$ tenemos que

$$E_0\ (m_t\ (\omega,\ \theta)) \neq 0 \qquad t = 1,2,\ldots,$$

entonces P_0 no está en P. En la fórmula anterior la esperanza está tomada con respecto a P_0.

Cuando realizamos contrastes m estamos comprobando si las condiciones [1.1] se cumplen en nuestra muestra.

Para hacer esto estudiamos el comportamiento de su análogo muestral.

$$\hat{m}_n = n^{-1} \sum_{t=1}^{n} \hat{m}_{nt} \equiv n^{-1} \sum_{t=1}^{n} m_t \left(\omega^t, \hat{\theta}_n\right)$$

y buscamos que estén suficientemente próximos a cero. Un ejemplo simple de contraste m es un test de autocorrelación serial de primer orden. Si llamamos $(r_1, r_2, \ldots, r_n)$ a los residuos estimados, entonces investigamos si $n^{-1} \sum_{t=2}^{n} r_t r_{t-1}$ está próximo a cero en nuestra muestra.

Denotemos por $\hat{l}_{nt}$ a un vector 1 x k. Por sencillez podemos pensar que $\hat{l}_{nt}$ es el vector gradiente de la función de verosimilitud para la observación t, evaluado en el estimador asintóticamente eficiente $\hat{\theta}_n$. Sea $\hat{m}_{nt}$, como lo definimos antes, el vector p x 1 de funciones m para la observación t, evaluado también en $\hat{\theta}_n$.

White (1985) ha demostrado que bajo condiciones generales, los contrastes m se pueden computar haciendo la siguiente regresión de mínimos cuadrados ordinarios:

1 sobre $\hat{m}'_{nt}$, $\hat{l}_{nt}$.

Bajo la hipótesis de correcta especificación, nR_0^2 (el número de observaciones por el R^2 sin ajustar por una constante) se comportará asintóticamente como una variable aleatoria χ^2 con p grados de libertad.

Intuitivamente, lo que la regresión anterior intenta hacer es "explicar" una constante con m_{nt}. La forma en que esto puede ocurrir es que m_{nt} tenga medias diferentes de cero. Los términos contenidos en l_{nt} no pueden explicar la constante porque sus medias en general han sido anuladas por el proceso de estimación. Sólo la presencia de m_{nt} puede explicar la constante; sin embargo, queremos mantener l_{nt} en la regresión porque va a darnos la forma del producto exterior de la matriz de varianzas y covarianzas de los coeficientes estimados, $\hat{\theta}_n$ que se han usado al evaluar m_{nt}.

En la sección 2 damos condiciones que permitirán simplificaciones muy útiles en la computación de los contrastes m. La diagonalidad por bloques de la Matriz de Información es la primera de estas condiciones. Además, si las funciones m se pueden escribir como el producto de residuos (r_t) y otras cosas ($\hat{w}_{nt}$) que son funciones de las variables predeterminadas, esto es $\hat{m}_{nt} = r_t \hat{w}_{nt}$, y también que el gradiente se puede escribir como el producto de residuos y otras cosas (x_{nt}), que también son funciones de las variables predeterminadas, $\hat{l}_{nt} = r_t \hat{x}_{nt}$, entonces bajo homoscedasticidad condicional, los contrastes m se pueden computar regresando.

(1.2) $\qquad r_t$ sobre $\hat{w}'_{nt}, \hat{x}_{nt}$

Esta es una simplificación muy útil, que nos recuerda los contrastes de los multiplicadores de Lagrange. De hecho los contrastes m también se pueden considerar bajo ciertas condiciones como contrastes que buscan en los residuos algunos tipos particulares de no aleatoriedad. Esto nos permite identificar direcciones en las cuales los contrastes m tendrán poder local máximo asintóticamente. Resulta que bajo condiciones generales se puede demostrar que los contrastes m son también contrastes LM; por lo tanto la analogía que sugiere la regresión (1.2) es apropiada.

La regresión (1.2) intenta encontrar una correlación significativa entre r_t y w'_{nt}. De nuevo necesitamos x'_{nt} (aunque es normalmente ortogonal a r_t) para obtener la matriz de varianzas y covarianzas de los coeficientes estimados. Aquí usamos implícitamente la forma del Hesiano de la matriz de información.

El resto del trabajo está organizado como sigue: la sección 3 analiza el caso simple de la regresión lineal y contiene una gran variedad de contrastes m consistentes a heteroscedasticidad e inconsistentes a heteroscedasticidad. Algunos de estos contrastes posiblemente tendrán un buen comportamiento en pequeñas muestras.

La sección 4 trata los contrastes de la matriz de información dinámica en el contexto de la regresión lineal. Proponemos nuevos contrastes de especificación dinámica. Estos contrastes serán sensibles a una forma funcional incorrecta y/o una especificación dinámica insuficientemente rica.

La sección 5 presenta un modelo general que se puede considerar como la alternativa general que los contrastes de la matriz de información dinámica están considerando en el caso de regresión lineal. Este modelo general contiene como casos especiales varios modelos ya conocidos y muy usados, así como algunos modelos nuevos que pueden ser útiles para predicción.

La sección 6 contiene otras aplicaciones del marco de los contrastes m. En concreto consideramos un contraste de restricciones de factores comunes en presencia de heteroscedasticidad de forma desconocida.

En la sección 7 se encuentran las principales conclusiones y sugerencias para la investigación futura.

2. Resultados generales

El Profesor White (1985), en el Teorema 3.3 ha propuesto una manera simple para computar los contrastes m. En esta sección demostramos que bajo ciertas condiciones generales existen versiones de los contrastes m que son asintóticamente equivalentes a las que él propone, que son computacionalmente más sencillas y además tienen interpretaciones útiles y familiares. Una de estas versiones viene dada por el Teorema 2.1 que está a continuación, bajo la condición de que la matriz de información del modelo sea diagonal por bloques bajo la hipótesis nula.

Vamos a suponer directamente que se cumplen algunas condiciones suficientes para el Teorema 3.3 de White. Estas condiciones aseguran que un contraste m se puede computar como nR_0^2 (el R^2 no ajustado por una constante) de la regresión de mínimos cuadrados ordinarios donde la variable dependiente es la constante unidad a lo largo de todo el tiempo y los regresores son el gradiente del logaritmo de la verosimilitud condicional y las funciones m, ambos evaluados en el estimador asintóticamente eficiente bajo la hipótesis nula.

Primero estableceremos la notación y definiciones. Las mediciones de los fenómenos económicos observables se pueden considerar como la realización de un proceso estocástico sobre un espacio de probabilidad (Ω, F, P_0). La medida de probabilidad desconocida P_0 nos da una descripción completa del comportamiento estocástico del mecanismo que genera los datos, y por lo tanto puede considerarse como el verdadero mecanismo generador de los datos.

Un punto de partida útil para intentar aprender acerca del mecanismo que genera los datos, que es desconocido, P_0, es especificar un modelo de probabilidad indiciado por algunos parámetros de interés.

DEFINICION D.1: Un modelo de probabilidad P es una colección de medidas de probabilidad diferentes sobre un espacio mesurable (Ω, F).

Los parámetros de interés se definen por una correspondencia ν: $P \rightarrow \Theta$, $\Theta \subseteq \mathbb{R}^k$, $k \in N_+ \equiv \{1,2,\ldots\}$

El modelo P se puede considerar como la colección de todas las medidas de probabilidad que uno está dispuesto a considerar como candidatos para haber generado los datos, y debería ser suficientemente amplia como para que sea plausible que $P_0 \in P$. Cuando efectivamente $P_0 \in P$, decimos que el modelo P está correctamente especificado o simplemente que P es el verdadero modelo.

Una manera de formular los modelos de probabilidad es considerar una familia que genera funciones de verosimilitud en el sentido de la siguiente definición.

DEFINICION D.2: (Familia que Genera Verosimilitudes $\{f_n(\omega^n,\theta)\}$)

Sea $\{(A_t, D_t)\}$ una secuencia de espacio mesurables y definamos

$$\Omega^n \equiv X_{t=1}^{n} A_t \ , \ F^n \equiv X_{t=1}^{n} D_t \ , \ \Omega \equiv X_{t=1}^{\infty} A_t \ , \ F \equiv X_{t=1}^{\infty} D_t$$

Sea $P \equiv \{P_\theta: \theta \in \Theta\}$ una familia de medidas de probabilidad diferentes sobre (Ω,F), donde $\Theta \subseteq \mathbb{R}^k$, $k \in N$, de manera qué $\Theta = \nu$ (P) con ν definida como la función tal que $\theta = \nu\ (P_0)$.

La familia P es una familia que genera verosimilitudes $\{f_n(\omega^n,\theta)\}$ si y sólo si existe una medida μ sobre (Ω,F) que no depende de θ tal que para cada n = 1, 2, ... la restricción de μ a (Ω^n, F^n) definida como

$$\mu^n(F) \equiv \mu\ (F \mathop{X}_{t=n+1}^{\infty} A_t), \quad F \in F^n$$

es σ finita, y para cada Θ en Θ la restricción de P_O a (Ω^n, F^n), definida como

$$P_{\Theta n}(F) \equiv P_{\Theta}(F \mathop{X}_{t=n+1}^{\infty} A_t) , F \Omega F^n$$

es absolutamente contínua con respecto a μ^n de forma que $P_{\Theta n}$ tiene una derivada de Radon-Nikodỳm

$$f_n(\omega^n, \Theta) = d P_{\Theta n}/d\mu^n, \quad \omega^n \in \Omega^n$$

La densidad f_n es la función de verosimilitud generada por la familia.

Consideremos una familia de verosimilitudes $f_n(\omega^n,\Theta)$ indiciada por un vector de parámetros Θ, de dimensión kx1, donde se puede interpretar n como el tamaño de la muestra. La notación sugiere que la verosimilitud puede cambiar con el tamaño de la muestra. Sea Θ_n un estimador asintóticamente eficiente de Θ.

Definimos $L_n(\omega,\Theta) \equiv -n^{-1} \ln f_n(\omega^n,\Theta)$. Podemos obtener una representación útil de $f_n(\omega^n,\Theta)$ usando las verosimilitudes condicionales, definidas como

$$f_{t|t-1}(\omega^t, \Theta) \equiv \frac{f_t(\omega^t, \Theta)}{f_{t-1}(\omega^{t-1},\Theta)}, \qquad f_{t-1}(\omega^t, \Theta) > 0$$

$$\equiv \int f_t(\omega^t, \Theta)\, d\, \mu^{t-1} (\omega^{t-1}), \quad f_{t-1}(\omega^{t-1}, \Theta) \doteq 0,\ t = 2,\ldots$$

y $f_{1/0}(\omega^1, \Theta) \equiv f_1(\omega^1, \Theta)$. La función $f_{t/t-1}$ nos da la verosimilitud condicional de la observación t-ésima dada la información disponible en el momento t-1.

Usando esta definición podemos escribir

$$L_n(\omega, \theta) = -n^{-1} \sum_{t=1}^{n} \ln f_{t/t-1}(\omega^t, \theta)$$

Particionamos el vector kx1 de parámetros $\theta' \equiv (\beta', \alpha')$ donde β es un vector 1x1 y α es un vector (k-1)x1, y podemos escribir el gradiente de la verosimilitud condicional como el vector 1xk

$$\nabla_\theta \ln f_{t/t-1} (\omega^t,\theta) = (\nabla_\beta \ln f_{t/t-1}(\omega^t,\theta), \nabla_\alpha \ln f_{t/t-1}(\omega^t,\theta))$$

Sea $m_{nt}(\omega^t,\theta)$ un vector px1 de condiciones de momentos tal que bajo especificación correcta

$$\int m_{nt}(\omega^t,\theta)\, f_n(\omega^t\theta)\, d\mu^t = 0$$

Bajo condiciones suficientes para que se cumpla el Teorema 3.3 de White (1985), los contrastes m se pueden computar como nR_0^2 de la regresión de mínimos cuadrados ordinarios donde la variable dependiente es un vector columna de unos y los regresores son

$$m_{nt}(\omega^t,\hat{\theta}_n)', \quad \nabla_\beta \ln f_{t/t-1}(\omega^t, \hat{\theta}_n), \quad \nabla_\alpha \ln f_{t/t-1}(\omega^t, \hat{\theta}_n).$$

Para simplificar la notación definamos.

$\hat{m}_{nt} \equiv m_t(\omega^t, \hat{\theta}_n)$, un vector px1

$\hat{l}_{\beta nt} \equiv \nabla_\beta \ln f_{t/t-1}(\omega^t, \hat{\theta}_n)$, un vector 1x1

$\hat{l}_{\alpha nt} \equiv \nabla_\alpha \ln f_{t/t-1}(\omega^t, \hat{\theta}_n)$, un vector 1x(k-1) y

$\hat{m} \equiv (\hat{m}_{n1}, \hat{m}_{n2}, \ldots, \hat{m}_{nn})'$, una matriz nxp

$\hat{l}_\beta \equiv (\hat{l}_{\beta n1}', \hat{l}_{\beta n2}', \ldots, \hat{l}_{\beta nn}')$, una matriz nx1

$\hat{l}_\alpha \equiv (\hat{l}_{\alpha n1}', \hat{l}_{\alpha n2}', \ldots, \hat{l}_{\alpha nn}')$, una matriz nx(k-1)

Hemos suprimido los subíndices n para simplificar la notación. Denominamos $i = (1,\ldots, 1)'$ a un vector columna de dimensión nx1, cuyos elementos son todos iguales a la unidad.

A continuación, establecemos las condiciones que caracterizan las situaciones en las que nuestro primer Teorema, el 2.1, se puede aplicar.

Lo primero que haremos será enumerar los supuestos que White (1985) usa para demostrar su Teorema 3.3. En el trabajo original de White se puede encontrar una discusión detallada de estos supuestos.

<u>SUPUESTO S.1</u>: Los datos observados son una realización de un proceso estocástico sobre un espacio de probabilidad (Ω, F, P_O).

<u>SUPUESTO S.2</u>: Existe una secuencia $\{\hat{\theta}_n : \Omega \to \Theta \subseteq \mathbb{R}^k,\ k \in n\}$ de funciones F mesurables, una secuencia no estocástica $\{\theta_n^*\}$, una formación doble $\{l_{nt} : \Omega \times \Theta \to \mathbb{R}^k\}$ mesurable $-F \times B(\mathbb{R}^k)$ e integrable- P_O para cada θ en Θ, y una secuencia no estocástica O(1) de matrices kxk uniformemente no singulares $\{H^*_n\}$ tal que

$$\sqrt{n}(\hat{\theta}_n - \theta_n^*) + H^{*-1}_n\, n^{-1/2} \sum_{t=1}^{n} l^*_{nt} \xrightarrow{P_O} 0,$$

donde $l^*_{nt} \equiv l_{nt}(\omega, \theta^*_n)$.

El siguiente supuesto se refiere a un conjunto de parámetros irrelevantes (nuisance) Π, y sus estimadores Π_n.

En este trabajo no los trataremos explícitamente, pero el análisis que hacemos a continuación es también válido si los hubiésemos considerado explícitamente.

SUPUESTO S.3: Existe una secuencia $\{\hat{\Pi}_n: \Omega \to \Pi \subseteq |R^j,\ j \in N\}$ de funciones mesurables F, una secuencia no estocástica $\{\Pi^*_n\}$ tal que $\hat{\Pi}_n - \Pi^*_n \xrightarrow{P_0} 0$, y una formación doble $\{m_{nt}: \Omega \times \Theta \times \Pi \to |R^p,\ p \in N\}$ mesurable $-F \times B(|R^k) \times B(|R^j)$ tal que $m_{nt}(\omega, \theta, \Pi^*_n)$ es integrable- P_0 para θ en Θ, n, t = 1, 2,...

El modelo P es típicamente el modelo que se está sometiendo a contraste. El modelo recibe el nombre de modelo explícitamente sometido al contraste. Sin embargo, es importante reconocer que existirán, en general, modelos alternativos para los que el poder del contraste no es mayor que su tamaño. Consecuentemente, definimos el modelo subyacente al contraste de la siguiente manera.

DEFINICION D.3: (Modelo subyacente al contraste): Definimos como modelo subyacente al contraste a la colección R de todas las medidas de probabilidad P^0 tal que si se cumplen los supuestos S.1 - S.3 para $P_0 = P^0$ entonces

(i) Para cualquier secuencia $\{\theta_n\}$ que satisface el Supuesto S.2 existe una secuencia no estocástica que es O(1) de matrices pxk $\{G^*_n\}$ (que dependen de P^0) tal que

$$\sqrt{n}\, \hat{m}_n - n^{-1/2} \sum_{t=1}^{n} m^*_{nt} - G^*_n \sqrt{n}\, (\hat{\theta}_n - \theta_n) \xrightarrow{P^0} 0;$$

y

(ii) La matriz V^*_n se define como

$$V^*_n \equiv \text{var} \ [n^{-1/2} \sum_{t=1}^{n} (m^*_{nt} - G^*_n H^{*-1}_n l^*_{nt}]$$

existe para $n = 1,2,\ldots.$ Además, $\{V^*_n\}$ es $O(1)$, uniformemente definida positiva y

$$V^{*-1/2}_n n^{-1/2} \sum_{t=1}^{n} (m^*_{nt} - G^*_n H^{*-1}_n l^*_{nt}) \xrightarrow{L(P^0)} N(0, I_p)$$

donde $\xrightarrow{L(P^0)}$ denota convergencia en ley bajo P^0.

En general, el modelo subyacente al contraste incluye todas las medidas de probabilidad para las cuales

$$n^{-1/2} \sum_{t=1}^{n} E^0 (l^*_{nt}) \to 0$$

$$n^{-1/2} \sum_{t=1}^{n} E^0 (m^*_{nt}) \to 0$$

donde E^0 denota la esperanza bajo P^0, de manera que $E^0(l^*_{nt})=0$ y $E^0(m^*_{nt})=0$ para todo n y t es una condición suficiente pero no necesaria para la inclusión en R.

SUPUESTO S.4: Sea P un subconjunto determinado de R, y sea $\{\hat{V}_n : \Omega \to \mathbb{R}^{p \times p}\}$ una secuencia de matrices semidefinidas positivas mesurables -F. Entonces $P \subseteq Q$, donde Q es el conjunto de todos los elementos P^0 de R tales que $\hat{V}_n - V^*_n \xrightarrow{P^0} 0$.

LLamaremos Q al modelo contrastado implícitamente, puesto que se pueden realizar en Q contrastes para P^0 del tamaño (asintótico) apropiado. Q puede ser un subconjunto propio de R; por lo tanto pueden existir medidas de probabilidad P^0 en R pero no en Q para las cuales uno no tendrá un contraste del tamaño apropiado. Esto puede ocurrir bien porque uno no ha usado o porque no puede usar un estimador consistente de V^*_n.

SUPUESTO A. 4': Sea P un subconjunto específico de R. Entonces $P \subseteq Q'$, donde Q' es el conjunto de todos los elementos de R tales que si $P_0 \in Q'$

(i) $\quad \mathrm{var}\,(n^{-1/2} \sum_{t=1}^{n} l^*_{nt}) = H^*_n;$

(ii) $\quad \mathrm{cov}(n^{-1/2} \sum_{t=1}^{n} m^*_{nt},\ n^{-1/2} \Sigma_{t=1}^{n} l^*_{nt}) = G^*_n;$

(iii) $\quad n^{-1} \Sigma_{t=1}^{n} \hat{m}'_{nt} \hat{m}_{nt} - F^*_n \xrightarrow{P_0} 0,\ F^*_n \equiv \mathrm{var}\,(n^{-1/2} \sum_{t=1}^{n} m^*_{nt});$

(iv) $\quad n^{-1} \Sigma_{t=1}^{n} \hat{m}_{nt} \hat{l}'_{nt} - G^*_n \xrightarrow{P_0} 0;$

(v) $\quad n^{-1} \Sigma_{t=1}^{n} \hat{l}_{nt} \hat{l}'_{nt} - H^*_n \xrightarrow{P_0} 0;$

(vi) $\quad n^{-1/2} \sum_{t=1}^{n} \hat{l}_{nt} \xrightarrow{P_0} 0$

Bajos los Supuestos S.1, S.2, S.3, S.4' y correcta especificación ($P_0 \subset P$), el Teorema 3.3 de White (1985) nos asegura que un contraste m se puede computar como n R_0^2 de una regresión de mínimos cuadrados ordinarios de

i sobre $\hat{m}$, $\hat{l}_\beta$, $\hat{l}_\alpha$

donde hemos particionado la matriz l, de acuerdo a los dos grupos de parámetros que consideramos.

El siguiente supuesto es el único supuesto adicional que nos permitirá probar el Teorema 2.1.

<u>SUPUESTO S.5</u>: Bajo la hipótesis nula de especificación correcta, la Matriz de Información es diagonal por bloques entre los parámetros α y β, esto es

(i) $$n^{-1}\, \hat{l}'_{\beta}\, \hat{l}_{\alpha} \xrightarrow{P} 0$$

Además

(ii) $$n^{-1}\, \hat{m}'\, \hat{l}_{\alpha} \xrightarrow{P} 0$$

El supuesto S.5 es el supuesto crucial, pues nos permite eliminar términos asintóticamente insignificantes en el estadístico del contraste m.

En el caso de la regresión lineal, β puede interpretarse como los parámetros de la media condicional y α como el parámetro de la varianza condicional y m puede ser considerado como un conjunto de condiciones que contrastan la adecuación de la especificación de la media condicional. El Supuesto S.5 (ii) requiere que bajo la hipótesis nula las condiciones m sean asintóticamente incorrelacionadas con α.

Si las condiciones m se refieren a la especificación de la varianza condicional, podemos invertir los papeles de α y β en el anterior supuesto y así también se cumpliría éste.

TEOREMA 2.1:

Bajo los Supuestos S.1, S.2, S.3, S.4' y S.5, un test m asintóticamente equivalente al $n\ R_0^2$ de la regresión de mínimos cuadrados ordinarios de

$$i \text{ sobre } \hat{m},\ \hat{l}_\beta,\ \hat{l}_\alpha$$

puede computarse como nR_0^2 de la regresión de mínimos cuadrados ordinarios de

$$i \text{ sobre } \hat{m},\ \hat{l}_\beta$$

Prueba: Véase Apéndice matemático.

El teorema de arriba nos permite eliminar algunos regresores cuando estemos computando ciertos contrastes m. La elección entre las dos formas de computar los contrastes m estará basada en consideraciones de muestras finitas.

Cabría esperar un mejor comportamiento en muestras finitas de la versión simplificada que nos da el Teorema 2.1, pero esto no es siempre así necesariamente. Más adelante volveremos sobre este tema.

Nuestro siguiente teorema nos da condiciones bajo las cuales se puede simplificar todavía más la computación de los contrastes m. Estas condiciones son

(i) Estamos considerando una sola ecuación.

(ii) Las funciones m y el gradiente condicional se pueden escribir como el producto de residuos (posiblemente generalizados) y funciones de las variables predeterminadas (y de los coeficientes estimados).

(iii) Homoscedasticidad condicional.

Este teorema se puede aplicar directamente a simplificar la computación de los contrastes m en el caso de que tengamos verosimilitudes que pertenecen a la subclase C- lineal de la familia exponencial lineal; ver Gourieroux, Monfort, Renault y Trognon, (1984, p. 43). Esta subclase incluye las densidades gaussiana (con varianza conocida), Poisson, gamma y binomial. Además su rango de aplicabilidad se ve ensanchado cuando podemos aplicar previamente el Teorema 2.1 (por ejemplo el caso de una densidad gaussiana con varianza desconocida).

A continuación, establecemos notación adicional y damos suficientes condiciones explícitas para probar el Teorema 2.2.

Sea $\omega_t \equiv (Y_t, z_t)$, de modo que las variables observables en el período t son el vector (Y_t, z_t) de dimensión (1+m)x1, donde z_t es un vector mx1 de variables exógenas contemporáneas e Y_t es una variable dependiente escalar.

Entonces $f_{t/t-1}(\omega^t, \theta)$, la densidad condicional, se puede redefinir como la densidad condicional de Y_t dado z_t y todos los ω_t pasados. F_{t-1} se puede reinterpretar como $\sigma(\ldots, \omega_{t-1}, z_t)$.

<u>SUPUESTO B.1</u>: Y_t es un escalar.

<u>SUPUESTO B.2</u>: El vector $l_{\beta tn} \equiv \nabla_\beta \ln f_{t/t-1} (\omega^t, \theta)$ de dimensión 1x1 se puede escribir como

$$\nabla_\beta \ln f_{t/t-1} (\omega^t, \theta) = u_t(\omega^t, \theta) \; x_t \; (\omega^{t-1}, z_t, \theta)$$

donde $u_t(\omega^t, \theta)$ es un escalar y $x_t(\omega^{t-1}, z_t, \theta)$ es un vector 1 x 1. Además, uno de los regresores es una constante,

$$x_{t1}(\omega^{t-1}, z; \theta) \equiv 1, \qquad t = 1,.., n.$$

SUPUESTO B.3: El vector $m_t(\omega^t, r)$ de dimensión px1 se puede escribir como

$$m_t(\omega^t, \theta) = u_t(\omega^t, \theta) \; . \; w_t(\omega^{t-1}, z_t, \theta)$$

donde $u_t(\omega^t, \theta)$ es un escalar y $w_t(\omega^{t-1}, z_t, \theta)$ es un vector de dimensión px1. Hemos suprimido la dependencia de m_t de n y Π por simplificar la notación.

Nótese que las funciones $u_t(\omega^t, \theta)$ de los supuestos B.2 y B.3 son las mismas. Cuando el modelo está correctamente especificado,

$$E_0(\nabla_\beta \ln f_{t/t-1}(\omega^t, \theta_0)/F_{t-1}) = E(u_t(\omega^t, \theta_0) \; x_t(\omega^{t-1}, z_t; \theta_0)/F_{t-1})$$

$$= E_0(u_t(\omega^t, \theta_0)/F_{t-1}) \;\; x_t(\omega^{t-1}, z_t, \theta_0) = 0,$$

de manera que

$$E_0(u_t(\omega^t, \theta_0)/F_{t-1}) = 0, \qquad t = 1, 2, \ldots \; .$$

Podemos considerar que $u_t(\omega^t, \theta)$ son residuos. En general serán residuos generalizados o estandarizados. Este punto se verá más claro en los ejemplos que presentaremos en las secciones siguientes.

$x_t(\omega^{t-1}, z_t, \theta)$ y $w_t(\omega^{t-1}, z_t, \theta)$ son funciones de las variables exógenas y predeterminadas y de los parámetros. Nótese que no dependen de la variable endógena contemporánea Y_t, pero pueden depender de valores retrasados de Y_t.

Llamemos al vector columna de unos $i = (1, \ldots, 1)'$, de dimensión nx1. Sea θ_n un estimador consistente y asintóticamente eficiente de θ. Definimos R como la matriz diagonal de dimensión nxn tal que

$$R = \text{diag}\ (u_1(w^1, \hat{\theta}_n),\ u_2(w^2, \hat{\theta}_n), \ldots,\ u_n(w^n, \hat{\theta}_n))$$

W es la matriz nxp $W = (w_1(Z_1, \hat{\theta}_n),\ w_2(\omega^1, Z_2, \hat{\theta}_n), \ldots,\ w_n(\omega^{n-1}, Z_n, \hat{\theta}_n))'$

X es la matriz nx1 $X = (x_1(Z_1, \hat{\theta}_n)',\ x_2(\omega^1, Z_2, \hat{\theta}_n), \ldots,\ x_n(\omega^{n-1}, Z_n, \hat{\theta}_n))'$

r es el vector nx1 $r = (u_1(\omega^1, \hat{\theta}_n),\ u_2(\omega^2, \hat{\theta}_n), \ldots,\ u_n(\omega^n, \hat{\theta}_n))'$

Decimos que nuestro modelo es condicionalmente homoscedástico si bajo especificación correcta tenemos que

$$E_0(u_t(\omega^t, \theta_0)^2/F_{t-1}) \equiv \sigma_t^2 = \sigma_0^2 < \infty \quad \text{para todo} \quad t = 1, 2, \ldots$$

donde $F_{t-1} \equiv \sigma\ (\ldots, \omega^{t-2}, \omega^{t-1}, z_t)$.

Definimos $\hat{\sigma}_n^2 \equiv n^{-1} \sum_{t=1}^{n} u_t\ (\omega^t, \hat{\theta}_n)^2$

Nuestro siguiente supuesto, B4, requiere esencialmene que ocurran ciertas convergencias útiles. Las convergencias supuestas en B.4 ocurrirán típicamente bajo homoscedasticidad condicional cuando se cumplan las leyes de los grandes números apropiadas.

<u>SUPUESTO B.4</u>: Bajo especificación correcta ($P_0 \in P$)

(i) $\hat{\sigma}_n^2 - \sigma_0^2 \xrightarrow{P_0} 0$;

(ii) $$\hat{\sigma}_0^2\, W'\, W/n - W'\, R\, R\, W/n = o_{P_0}(1);$$

$$\hat{\sigma}_0^2\, W'\, X/n - W'\, R\, R\, X/n = o_{P_0}(1);$$

$$\hat{\sigma}_0^2\, X'\, X/n - X'\, R\, R\, X/n = o_{P_0}(1);$$

Estas condiciones también podrían cumplirse bajo heteroscedasticidad condicional, pero este caso será en principio poco frecuente.

Aunque estas condiciones se cumpliesen bajo heteroscedasticidad condicional, sus consecuencias no serían muy serias, puesto que bajo heteroscedasticidad condicional usaríamos contrastes m consistentes a heteroscedasticidad, y las inferencias serían asintóticamente correctas.

TEOREMA 2.2

Bajos los Supuestos S.1 - S.4', B.1 - B.4, podemos computar los contrastes m como

$$M''_n = n\, R^2 \overset{A}{\sim} \chi^2_p$$

donde R^2 es el R^2 ajustado por el uso de una constante de una regresión de mínimos cuadrados ordinarios de

$$r_{nt} \text{ sobre } \hat{w}_{nt},\ \hat{x}_{nt}$$

DEMOSTRACION: Ver Apéndice Matemático.

Este Teorema nos permite reconocer el hecho de que bajo ciertas condiciones el Teorema 2.1 está usando la matriz de covarianzas de los coeficientes estimados que es consistente a heteroscedasticidad.

Bajo homoscedasticidad condicional, no hay necesidad de usar el estimador de la matriz de covarianzas que es consistente a heteroscedasticidad, y por lo tanto los contrastes m se pueden computar de la manera familiar en que se computan los contraste LM.

Merece la pena resaltar que bajo heteroscedasticidad condicional los contrastes m se deberían computar como se propone en el Teorema 2.1, si queremos obtener un contraste con el tamaño asintótico correcto. Además, bajo heteroscedasticidad condicional y no diagonalidad por bloques de la matriz de información bajo la hipótesis nula, los contrastes m y los contrastes de los multiplicadores de Lagrange se deben computar como se propone en el Teorema 3.3 de White (1985).

En el Teorema 2.2 usamos el R^2 ajustado por el uso de una constante. Esto supone implícitamente que usamos una constante como uno de los regresores.

Bajo las condiciones del Teorema 2.2, los contrastes m pueden considerarse como contrastes que analizan los residuos (posiblemente generalizados) con respecto a formas específicas de no aleatoriedad.

El procedimiento de analizar los residuos con respecto a formas concretas de no aleatoriedad no será apropiado en todas las circunstancias. En particular, puede no ser apropiado cuando el Teorema 2.2 no se cumpla.

El análisis de los residuos sí será apropiado, sin embargo, cuando se cumpla el Teorema 2.2 (y posiblemente también el 2.1). La utilidad del análisis de los residuos proviene del hecho de que bajo ciertas condiciones, analizar los residuos es equivalente a llevar a cabo contrastes m y de los multiplicadores de Lagrange.

En este trabajo no damos condiciones primitivas sobre el comportamiento de las observaciones, sino que por el contrario suponemos que se cumplan las leyes de los grandes números y los teoremas centrales del límite apropiados. Domowitz y Hakkio (1984), página 7, dan condiciones generales suficientes para que estos se cumplan.

Recientemente Newey (1985), Tauchen (1985) y White (1985) han demostrado que muchos contrastes de especificación pueden ser considerados casos particulares de contrastes m. Esos contrastes m están basados en condiciones de momentos que deberían ser cero bajo especificación correcta. Las condiciones de momentos (m) están evaluadas bajo la hipótesis nula. Esto sugiere una analogía con los contrastes de los multiplicadores de Lagrange (LM). Si pudiésemos demostrar que los contrastes m son contrastes LM, entonces heredarían las propiedades de optimalidad de los contrastes LM y podríamos aplicarles nuestros conocimientos sobre contrastes LM. Este es precisamente el contenido del Teorema 2.3, que va a continuación. Primero estableceremos alguna notación y supuestos.

La especificación correcta de un modelo dado P implicará, en general, la validez de un número, quizás infinito, de condiciones de momentos de la forma.

$$E_0 (m_{nt} (\omega, \theta)) = 0, \qquad t = 1, 2, \ldots,$$

donde $m_{nt}: \Omega \times \Theta \to \mathbb{R}^p$ con $\omega \in \Omega$.

Si para cada θ en Θ uno tiene que

$$E_0(m_{nt}(\omega, \theta)) \neq 0, \qquad t = 1, 2, \ldots,$$

entonces P_0 no está en P. Ahora las esperanzas están tomadas con respecto a P_0.

Las anteriores condiciones que deben ser cero nos recuerdan las condiciones del gradiente con respecto a las variables adicionales que deben ser cero cuando hacemos un contraste de los multiplicadores de Lagrange.

A continuación vamos a mostrar que las dos situaciones son análogas.

DEFINICION D.4: (Verosimilitud Artificial)

Si tenemos una familia generadora de verosimilitudes $\{f_n(\omega^n,\theta)\}$ y un vector de dimensión px1 de condiciones de momentos $m_{nt}(\omega^t,\theta)$ tal que bajo correcta especificación $(P_0 \in P)$.

$$\int m_{nt}(\omega^t,\theta)\ f_t(\omega^t,\theta)\ d\ \mu^t = 0$$

definimos la verosimilitud artificial $h_n(\omega^n;\theta,\lambda)$, donde $\lambda \in \Lambda \subset \mathbb{R}^p$ y definimos Λ como el conjunto de todos los λ tales que

$$0 < \int \exp\{\lambda'\ m_n(\omega^n,\ \theta)\}\ f_n(\omega^n,\ \theta)\ d\ \mu^n < +\infty$$

para todo $\theta \in \Theta$ y todo $n \in \mathbb{N}$. Además, para todo θ,λ en $\Theta \times \Lambda$ definimos

$$h_n(\omega^n,\ \theta,\ \lambda) \equiv \exp\,[\lambda'\ m_n(\omega^n;\ \theta) - \Psi_n(\theta,\ \lambda)]\ f_n(\omega^n,\ \theta)$$

donde

$$\Psi_n(\theta,\ \lambda) \equiv \log \int \exp\,[\lambda'\ m_n(\omega^n,\ \theta)]\ f_n(\omega^n,\ \theta)\ d\ \mu^n$$

Cuando Λ contiene un entorno del origen, decimos que h_n es una verosimilitud artificial propiamente dicha, generada por f_n y m_n.

La función Ψ_n definida sobre $\Theta x\Lambda$, toma valores en $\mathbb{R}$, y se puede interpretar como el logaritmo de la transformada de laplace de $\int f_n(\omega^n,\theta)\, d\,\mu^n$.

El siguiente teorema muestra que bajo condiciones generales la construcción anterior es una densidad propiamente dicha y que satisface los requisitos que nos permiten considerar a los contrastes m como contrastes de los multiplicadores de Lagrange.

TEOREMA 2.3

Supongamos que estamos considerando una familia generadora de verosimilitudes $\{f_n(\omega^n,\theta)\}$, diferenciable en θ a.s. $-P_0$, y que tenemos un vector px1 de condiciones de momentos $m_{nt}(\omega^t,\theta)$ que es mesurable- F^t tal que bajo especificación correcta ($P_0 \in P$):

$$\int m_t(\omega^t,\ \theta)\ f_n(\omega^t,\ \theta)\ d\ \mu^t = 0 \qquad t= 1,\ 2,\ldots$$

Entonces $h_n\ (\omega^n,\ \theta,\ \lambda)$ dada por la definición D.4 es tal que

(i) $h_n(\omega^n;\ \theta,\ \lambda)$ es una verosimilitud propiamente dicha para todo $\theta,\ \lambda \in \Theta \times \Lambda$, y

$$h_n(\omega^n,\ \theta,\ \lambda)\ |_{\lambda=0} = f_n(\omega^n,\ \theta)$$

(ii) $\nabla_\theta \ln h_n(\omega^n;\ \theta,\ \lambda)\ |_{\lambda=0} = \nabla_\theta \ln f_n(\omega^n,\ \theta)$

Si Λ contiene un entorno del origen, entonces

(iii) $\nabla_\lambda \ln h_n(\omega^n;\ \theta,\ \lambda)\ |_{\lambda=0} = m_n(\omega^n,\ \theta)'$

Demostración: véase el Apéndice Matemático.

El Teorema anterior nos permite incluir un contraste m dado en el marco de los contrastes LM, anidando la densidad original y las funciones m de manera que el gradiente de la nueva densidad, bajo la hipótesis nula, se reduce al gradiente de la densidad original y las funciones m.

Esta verosimilitud artificial ha sido usada por Gourieroux, Monfort, Renault y Trognon (1984) para incluir una densidad dada en la familia lineal exponencial.

White (1984b) propone cuatro condiciones que una verosimilitud artificial debería satisfacer para ser útil. La verosimilitud artificial de la definición D.4 satisface esas condiciones si igualamos lo que White llama $\Upsilon(x;\ \Upsilon,\ \theta)$ a $-1 + \exp\,[\lambda' m(\omega;\ \theta) - \Psi(\theta,\ \lambda)]$ y tomamos en cuenta las diferencias de de notación.

Si la verosimilitud $f_n(\omega^n,\theta)$ pertenece a la familia exponencial, $h_n(\omega^n;\theta,\lambda)$ también pertenecerá a la familia exponencial; véase Gourieroux, Monfort, Renault y Trognon (1984, Annexe 3). En algunos casos, uno está interesado en considerar densidades alternativas que están en la misma familia paramétrica que $f_n(\omega^n,\theta)$. En muchos casos la verosimilitud artificial nos sugiere la forma de esas densidades alternativas. Si $f_n(\omega^n,\theta)$ es gaussiana, la verosimilitud artificial sugiere alternativas gaussianas y no gaussianas que resultan naturales y atractivas.

Esta densidad artificial revela que los constrastes m pueden ser considerados como contrastes clásicos contra alternativas paramétricas específicas, y también nos sugerirían los contrastes de Wald y del cociente de las verosimilitudes correspondientes a cada contraste m.

Para construir los contrastes de Wald y del cociente de las verosimilitudes correspondientes necesitaríamos condiciones que asegurasen la consistencia y normalidad asintótica de los estimadores de máxima verosimilitud de θ y λ que obtendríamos maximizando $h_n(\omega^n; \theta, \lambda)$ sin restricciones sobre λ.

Este problema de estimación puede ser muy complicado en general; esto resalta la facilidad con que se pueden computar los contrastes m.

3. Contrastes m en el caso de regresión lineal

En esta sección presentamos varias formas asintóticamente equivalentes de computar los contrastes m. La elección entre ellas vendrá dada fundamentalmente por su comportamiento en pequeñas muestras. Aquí comentamos varios estudios teóricos que pueden ser relevantes para nuestra elección. También discutimos algunos resultados de simulaciones que pueden resultar útiles para tomar una decisión. Esta discusión sugiere candidatos que deberían tener un buen comportamiento en muestras de tamaño moderado. De todas formas la evidencia teórica y experimental no es completamente aplicable al caso que estamos considerando, puesto que sólo trata algunos casos especiales.

Cuando queremos realizar contrastes m en el caso de regresión lineal tenemos que hacer varias elecciones. La primera consiste en elegir entre contrastes inconsistentes a heteroscedasticidad y contrastes consistentes a heteroscedasticidad.

Si decidimos usar contrastes m inconsistentes a heteroscedasticidad, tendremos que elegir de entre varias versiones asintóticamente equivalentes. De manera análoga si decidimos usar contrastes m consistentes a heteroscedasticidad, hay varias versiones asintóticamente equivalentes de las que podremos elegir.

En esta sección presentamos varias versiones de contrastes consistentes e inconsistentes a heteroscedasticidad.

También discutimos la evidencia teórica y empírica que tenemos disponible sobre su comportamiento en muestras finitas. Indicamos cuales versiones pueden ser preferidas basados en estos estudios. En Pérez Amaral (1988) hemos llevado a cabo un estudio de simulación de los contrastes m con resultados muy alentadores sobre su comportamiento en muestras de moderado tamaño.

3.1. El modelo de Regresión Lineal

Presentamos aquí el caso de regresión lineal por motivos de exposición. La extensión al caso de regresión no lineal no debería presentar complicaciones importantes.

Sea el modelo

$$y = X\beta + u$$

donde y es un vector (nx1) de observaciones de una variable dependiente, X es una matriz de observaciones de variables independientes de dimensión nxk, que suponemos es de rango completo con respecto a sus columnas y u es un vector de dimensión nx1 de errores con media cero. El estimador mínimo cuadrático es:

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

Definimos el vector r de residuos estimados y dimensión nx1 como

$$r = (I - X\,(X'\,X)^{-1}\,X')\,y$$

$$r = (r_1,\, r_2, \ldots,\, r_n)'$$

Si $u_t/F_{t-1} \sim N(0,\, \sigma^2)$ entonces el gradiente de la verosimilitud condicional será el vector de dimensión 1x(k+1):

$$\nabla \log f_{t/t-1}(\omega^t,\, \theta) = \left(\frac{u_t x_{ti}}{\sigma^2},\; \frac{u_t^2 - \sigma^2}{\sigma^3}\right), \quad i = 1, \ldots, k.$$

donde $\theta = (\beta', \sigma)$. Definamos la matriz de dimensión nxn, $R = \mathrm{diag}\,(r_1,\, r_2, \ldots,\, r_n)$.

En el caso de regresión lineal un estimador de la forma Hesiana de la Matriz de Información será $n^{-1} \hat{\sigma}^2 (X' X)^{-1}$.

Este el estimador inconsistente a heteroscedasticidad de la matriz de varianzas y covarianzas de los parámetros de la media condicional. Un estimador de la forma del producto exterior de la Matriz de Información será

$$n^{-1}(X'R\ R\ X)^{-1}.$$

Supongamos que el vector de dimensión px1 de funciones m se puede escribir como

$$m_t(\omega,\ \theta)=\ u_t(\omega^t,\ \theta).\ w_t(\omega^{t-1},\ x_t;\ \theta)$$

donde $w_t(\omega^{t-1}, x_t; \theta)$ es un vector de dimensión px1 y puede ser considerado como un vector de variables posiblemente excluídas.

$$\text{Sea } \hat{w}_t \equiv w_t(\omega^{t-1},\ x_t,\ \hat{\theta}_n) \quad \text{y} \quad \hat{\sigma}^2 \equiv \frac{1}{n} \sum_{t=1}^{n} r_t^2.$$

Entonces se pueden computar contrastes m consistentes a heteroscedasticidad, como fueron propuestos originalmente por White (1985), regresando (bajo condiciones apropiadas)

$$(3.0) \qquad 1 \quad \text{sobre} \quad r_t\ \hat{w}_t',\ r_t\ x_t/\hat{\sigma}^2,\ (r_t^2 - \hat{\sigma}^2)/\hat{\sigma}^3$$

y n veces el coeficiente de determinación no adjustado por una constante (R_0^2) de esta regresión se comportará asintóticamente como una χ^2 con p grados de libertad bajo la hipótesis nula de especificación correcta. Nótese que si w_t incluye errores retrasados, variables dependientes retrasadas o regresores retrasados, el tamaño de la muestra tendría que ser reajustado.

En la regresión anterior podemos multiplicar los últimos k+1 regresores por potencias apropiadas de σ^2 sin que esto cambie el R_0^2. Esto nos lleva a la regresión simplificada.

$$(3.1) \qquad 1 \quad \text{sobre} \quad r_t \hat{w}'_t \,,\; r_t x_t ,\; (r_t^2 - \hat{\sigma}^2)$$

y entonces $M1 \equiv n R_0^2$ donde R_0^2 es el R^2 no ajustado por una constante, de la regresión (3.1).

Puesto que la Matriz de Información es diagonal por bloques, podemos usar el Teorema 2.1 para computar un contraste asintóticamente equivalente a M1 llevando a cabo la regresión

$$(3.2) \qquad 1 \quad \text{sobre} \quad r_t \hat{w}'_t ,\; r_t x_t$$

y $M2 \equiv n R_0^2$, donde R_0^2 es el R^2 no ajustado por el uso de una constante de la regresión (3.2). M2 será una variable aleatoria asintóticamente χ_p^2 bajo la hipótesis nula de especificación correcta. Nótese que puesto que nos hallamos en un contexto dinámico, muchos de estos contrastes serán del tipo de correlación serial y las x_t generalmente incluirán una constante como uno de los regresores.

El contraste M2 es un contraste consistente a heteroscedasticidad. Usa un estimador de la forma de Producto Exterior de la Matriz de Información.

Un contraste asintóticamente equivalente, posiblemente con mejor comportamiento en muestras finitas se podrá computar regresando

$$(3.3) \qquad 1 \quad \text{sobre} \quad r^*_t \hat{w}'_t ,\; r^*_t x_t$$

donde $r^*_t \equiv r_t/\sqrt{(1-k_{tt})}$ donde k_{tt} es el elemento diagonal t-ésimo de la matriz de predicción $X(X'X)^{-1}X'$. $M3 \equiv n R_0^2$ se distribuirá asintóticamente como una variable aleatoria χ^2_p bajo la hipótesis nula de especificación correcta.

Este contraste también es consistente a heteroscedasticidad. Cuando r_t es cero, será apropiado omitir la observación t-ésima de la muestra, sea k_{tt} igual a uno o no. A continuación pasamos a motivar el uso de esta corrección de los residuos.

Un resultado bien conocido, ver Theil (1971, p. 196) es que cuando las x_t son no aleatorias,

$$\text{var}\ (r_t) = \sigma^2(1 - k_{tt})$$

Concretamente, hay autores que han señalado que en vez de estudiar los residuos r_t, deberíamos usar los residuos estandarizados.

$$r_{st} \equiv r_t/(s\ \surd(1-k_{tt}))$$

donde s sería un estimador de la desviación estándar de la regresión; véase Belsley, Kuh y Welsch (1980, p. 19). Cook y Weisberg (1982, p. 18) denominan a r_{st}, residuos estudentizados internamente. En ese trabajo ellos dan una lista de referencias que estudian esta transformación de los residuos, aunque se ocupan fundamentalmente de los problemas causados por las observaciones atípicas. El uso de r^*_t en este contexto equivale al de los residuos estandarizados r_{st}, puesto que s es la misma para todas las observaciones.

El uso de r^*_t en vez de r_t se propone como una corrección de pequeñas muestras, puesto que r^*_t pondera más fuertemente los residuos de las observaciones influyentes. Esta corrección parece apropiada en vista de los resultados obtenidos en los estudios de Horn, Horn y Duncan (1975), Mackinnon y White (1985), Davidson y Mackinnon (1985) y Chesher y Jewitt (1987). Asintóticamente max k_{tt} -> 0 y por lo tanto usar r_t o r^*_t no tendrá ningún efecto asintótico.

En el contexto de la regresión (3.3) el uso de r^*_t tiene dos efectos:

<u>Primero</u>.- Usamos residuos estandarizados en el numerador del contraste.

<u>Segundo</u>.- Usamos en el denominador del contraste una matriz de varianzas y covarianzas consistente a heteroscedasticidad con buenas propiedades en muestras finitas tanto bajo homoscedasticidad condicional como bajo heteroscedasticidad condicional.

El usar r^*_t en vez de r_t es equivalente a omitir la observación t-ésima de la muestra y estimar los coeficientes basados solamente en el resto de la muestra. Esta técnica está relacionada con la técnica que se conoce por el nombre de validación cruzada.

Definamos $r_{(t)} = y_t - x_t' \hat{\beta}_{(t)}$, donde el subíndice (t) significa que los coeficientes se han obtenido eliminando la observación t de la muestra. Se puede demostrar que $r^*_t = r_{(t)}$. También se puede demostrar que si los residuos r_t se distribuyen normalmente, entonces r^*_t también se distribuirán normalmente con la misma estructura de autocorrelación. Véase Cook y Weisberg (1982, pp. 33-34). Por lo tanto usar r^*_t en vez de r_t nos dará resultados asintóticamente equivalentes con posiblemente mejores propiedades en muestras finitas.

Si podemos justificar la homoscedasticidad condicional, podemos aplicar el Teorema 2.2. De manera que podemos computar un contraste m asintóticamente equivalente a los anteriores con propiedades de poder local posiblemente mejores, (véase Cavanagh (1985)) llevando a cabo la siguiente regresión

(3.4) $\quad r_t \quad$ sobre $\quad \hat{w}'_t,\ x_t$

Sea R^2 el coeficiente de determinación ajustado por el uso de una constante entre los regresores. $M4 \equiv n\ R^2$ de esta regresión se comportará asintóticamente como una variable aleatoria χ^2_p bajo la hipótesis nula de especificación correcta. Este contraste es un contraste m inconsistente a heteroscedasticidad.

Un contraste que es asintóticamente equivalente a M4 se puede computar con la ayuda de la siguiente regresión

(3.5) $\quad r^*_t \quad$ sobre $\quad \hat{w}'_t,\ x_t$

$M5 \equiv n\ R^2$ de la regresión (3.5)

Este contraste tampoco es consistente a heteroscedasticidad. También podemos obtener otras versiones de los contrastes anteriores aplicándoles correcciones de los grados de libertad. Estas correcciones implicarían multiplicar los constrastes anteriores por los factores $\frac{n-k}{n}, \frac{n-k-p}{n}$ o bien $\frac{n-k-2p}{n}$ según sea apropiado.

Un contraste asintóticamente equivalente al contraste anterior se puede computar usando los contrastes F de significatividad conjunta de los coeficientes apropiados en las regresiones anteriores. La legitimidad de las versiones F de los contrastes (3.1), (3.2) y (3.3) se basa en el Teorema 3.6 de White (1985). La legitimidad de usar las versiones F en los contrastes (3.4) y (3.5) se basa en el Teorema 1.(a) de Engle (1982b).

Estas versiones F de los contrastes anteriores se pueden computar con gran sencillez. Cuando $\hat{w}'_t$ consta de un solo elemento, entonces un test t de significatividad de su coeficiente en las regresiones (3.1) a (3.5) sería apropiado.

Recordemos que los contrastes m consistentes a heteroscedasticidad son los contrastes M1, M2 y M3, y sus versiones correspondientes corregidas de grados de libertad.

Los contrastes M4 y M5 y sus versiones correspondientes corregidas de grados de libertad son contrastes m inconsistentes a heteroscedasticidad.

En la sección siguiente discutimos la elección entre contrastes m consistentes e inconsistentes a heteroscedasticidad.

3.2. La elección entre contrastes m consistentes e inconsistentes a heteroscedasticidad

Hemos visto que en el caso de regresión lineal, muchos contrastes m se pueden considerar contrastes LM (Teorema 2.3) que usan un estimador de la Forma de Producto Exterior de la Matriz de Información.

En este caso esto es equivalente a usar los estimadores de la matriz de varianzas y covarianzas consistentes a heteroscedasticidad de Eicker (1967) y White (1980).

En este momento tenemos disponibles esencialmente dos métodos diferentes de computar los contrastes m en el caso de regresión lineal. Estos dos métodos se diferencian en el estimador de la matriz de varianzas y covarianzas que usa cada uno de ellos.

- La forma de Producto Exterior (contrastes consistentes a heteroscedasticidad o robustos a heteroscedasticidad).

- La forma del Hesiano (contrastes que no son válidos bajo heteroscedasticidad condicional, contrastes que son inconsistentes a heteroscedasticidad).

Bajo homoscedasticidad condicional estas dos formas de computar los contrastes m son asintóticamente equivalentes, de modo que la elección entre las dos maneras de computar los contrastes se basará en su comportamiento en muestras finitas. Afortunadamente tenemos alguna literatura reciente que trata problemas relacionados con este, y puede servirnos de base para esta elección. De todas formas la necesidad de un estudio de simulación más específico es clara. Este estudio ha sido llevado a cabo por Pérez Amaral (1988) encontrándose resultados altamente satisfactorios sobre el comportamiento de los contrastes m en sus versiones consistentes e inconsistente a heteroscedasticidad. Seguidamente pasamos a comentar otra literatura relevante.

Cavanagh (1985) ha demostrado que bajo condiciones generales, los contrastes de los multiplicadores de Lagrange, de Wald y del Cociente de las Verosimilitudes que usan un estimador de la forma Hesiana de la Matriz de Información tienen mejores propiedades en cuanto a su poder local en muestras finitas que sus homónimos que usan un estimador de la Forma del Producto Exterior de la matriz de información. Cavanagh trabaja con observaciones independientemente no idénticamente distribuidas, usando verosimilitudes que son suficientemente regulares. De esta forma estudia el poder local de los contrastes con respecto a una variable adicional cada vez, usando aproximaciones de Edgeworth y experimentos de Monte Carlo. El considera el caso de un modelo correctamente especificado excepto, posiblemente, en la dirección de la variable omitida en la que se está llevando a cabo el contraste.

Sus resulados, por tanto no serían aplicables directamente al caso de contrastes de autocorrelación en un modelo dinámico de regresión cuando estamos en presencia de heteroscedasticidad de forma desconocida.

Sin embargo si uno cree que los resultados de Cavanagh también se aplican al caso de observaciones dependientes y contrastes de varias funciones m al mismo tiempo, tendrían consecuencias para la forma en que uno computaría los contrastes m.

Los resultados de Cavanagh sugieren que bajo homoscedasticidad condicional uno debería usar las versiones de los contrastes m que usan implícitamente el estimador de la forma Hesiana de la Matriz de Información, esto es, la versión de los contrastes m que es inconsistente a heteroscedasticidad. Nuestro caso es más general que el que considera Cavanagh y puede ser útil, como ya hemos dicho, hacer algunas simulaciones más específicas. Si uno decide usar contrastes m inconsistentes a heteroscedasticidad, todavía tiene disponible una amplia gama de contrastes m asintóticamente equivalentes.

Kiviet(1981) estudia el comportamiento en pequeñas muestras de los contrastes del tipo de los multiplicadores de Lagrange para autocorrelación serial. En ese estudio encuentra que los que tienen un mejor comportamiento en cuanto a poder y tamaño es la versión que nosotros llamaríamos MF4, esto es, el contraste m en su versión F, que es inconsistente a heteroscedasticidad.

Hay que tener en cuenta, sin embargo, que este estudio es muy limitado y no trata muchas de las versiones que hemos propuesto en este trabajo.

Los resultados del estudio de Monte Carlo de Pérez Amaral (1988) confirman la superioridad de este tipo de contrastes bajo homoscedasticidad condicional, e incluso bajo ciertas formas de heteroscedasticidad condicional. Asimismo en el caso de que haya cambios bruscos en la varianza, el contraste a elegir sería el MF2, esto es, la versión F del contraste consistente a heteroscedasticidad (<u>sin</u> modificar por r^*_t).

3.3 La elección entre contrastes m Consistentes a Heteroscedasticidad

Cuando estamos en presencia de heteroscedasticidad condicional (bien la hemos detectado o bien sospechamos de su existencia) entonces se deberían usar contrastes m consistentes a heteroscedasticidad.

Cuando los residuos de la regresión contienen heteroscedasticidad condicional de forma desconocida, la matriz de varianzas y covarianzas que los contrastes m inconsistentes a heteroscedasticidad están usando es inconsistente y por tanto estos contrastes no son válidos ni siquiera asintóticamente. Esto se aplica en particular a la manera habitual de computar los contrastes LM como una regresión de los residuos de la regresión sobre las variables posiblemente excluídas y los regresores.

Según MacKinnon y White(1985), los contrastes de heteroscedasticidad que se usan normalmente tienen un poder bajo, incluso en casos donde hay bastante heteroscedasticidad en los errores como para ocasionar serios problemas de inferencia cuando se usan los estadísticos t obtenidos por mínimos cuadrados oridinarios.

Basados en estas consideraciones, sería interesante tener contrastes robustos a heteroscedasticidad que tuviesen un buen comportamiento también en casos de homoscedasticidad condicional y heteroscedasticidad condicional en muestras finitas. Esto lo podemos lograr buscando estimadores de la matriz de varianzas y covarianzas consistentes a heteroscedasticidad con mejor comportamiento en muestras finitas. Hay varios trabajos que tratan este tema y otros temas relacionados.

MacKinnon y White(1985) estudian el comportamiento en pequeñas muestras de varios estimadores de la matriz de varianzas y covarianzas consistentes a heteroscedasticidad. Para ello usan métodos de simulación, y consideran el modelo de regresión lineal

$$y = X \beta + u$$

donde y es un vector nx1 de observaciones de una variable dependiente, X es una matriz nxk de observaciones sobre variables independientes, cuyas columnas son linealmente independientes, y u es un vector nx1 de errores con media cero. El estimador de mínimos cuadrados ordinarios para este modelo es

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

y $\sqrt{n}(\hat{\beta} - \beta)$ tiene media cero y matriz de varianzas y covarianzas $(X' X)^{-1} X' \Omega X (X' X)^{-1}$. Definamos $r = (I - X (X' X)^{-1} X')$ y entonces la matriz de varianzas y covarianzas anterior se puede estimar consistentemente por

$$(X' X)^{-1} X' \hat{\Omega} X (X' X)^{-1}.$$

donde $\hat{\Omega} = \text{diag } (r_1^2, r_2^2, \ldots, r_n^2)$; véase White(1980). Este es el estimador de Eicker-White que es el estimador de la matriz de varianzas y covarianzas consistente a heteroscedasticidad. Al estimador anterior lo llaman MacKinnon y White(1985) el estimador HC0, (consistente a heteroscedasticidad versión cero).

Una versión corregida de grados de libertad es llamada por estos autores HC1

$$(n/(n - k)) (X' X)^{-1} X' \hat{\Omega} X (X' X)^{-1}$$

Una corrección de grados de libertad más sofisticada tiene en cuenta que las observaciones influyentes tienden a aparecer con errores mínimo cuadráticos relativamente pequeños.

Denotemos como k_{tt} el t-ésimo elemento diagonal de la matriz $X(X' X)^{-1} X'$. La versión HC2 usa $\hat{\Omega} = \text{diag}\ (r^{*2}, r_2^{*2}, \ldots, r_n^{*2})$ donde $r_t^{*2} = r_t^2/\sqrt{(1 - k_{tt})}$. Este estimador pondera más fuertemente los errores que corresponden a las observaciones influyentes. Nótese que $k_{tt} = x_t\ (X' X)^{-1}\ x'_t$.

Una posibilidad adicional es usar el estimador de la "navaja" de la matriz de varianzas y covarianzas; véase Efron (1982). Mackinnon y White llaman a esta versión HC3, y es algo más difícil de computar. El orden de preferencia con respecto al sesgo que MacKinnon y White establecen es el siguiente: HC3 - HC2 - HC1 - HC0 (ordenados de menor a mayor sesgo).

Davidson y MacKinnon (1985) continúan con el trabajo de MacKinnon y White (1985) y nos ofrecen un menú de lo que ellos llaman contrastes robustos a heteroscedasticidad en las direcciones de regresión. Nosotros los llamaríamos contrastes robustos a heteroscedasticidad de la especificación de la media condicional. Davidson y MacKinnon dan versiones de HC0 a HC3 que usan residuos restringidos, esto es residuos y computados bajo la hipótesis nula de especificación correcta, como sería apropiado en el caso de los contrastes m que estiman el modelo bajo la hipótesis nula. Ellos llaman a estos estimadores HCR0 - HCR1 - HCR2 y HCR3.

En este caso concurren dos hechos afortunados: el uso de matrices de covarianzas robustas a heteroscedasticidad es computacionalmente simple y además el comportamiento en pequeñas muestras de HCR0 hasta HCR3 es mejor que el de sus homólogos sin restringir HC0 hasta HC3 (en el sentido de que su sesgo es menor).

La clasificación atendiendo a su sesgo en pequeñas muestras es, según Davidson y MacKinnon, HCR3, HCR2, HCR1 y HCR0. Las diferencias no parecen ser tan marcadas como en el caso de los residuos no restringidos, pero estos resultados están sujetos a las reservas que se aplican a todos los resultados de simulación. Además, no usan variables dependientes retrasadas como regresores.

La versión de los contrastes m robustos a heteroscedasticidad que puede resultar preferida es M3, que usa implícitamente HCR2. Esta conclusión es tentativa.

Si uno está dispuesto a incurrir en costes computacionales adicionales, una versión del estimador de la matriz de covarianzas de la navaja ("jacknife") puede ser preferida en base a su menor sesgo en pequeñas muestras. Esta línea de investigación la posponemos para ulteriores trabajos.

Otro trabajo reciente que apoyaría nuestra conclusión tentativa en favor de M3 es el de Chesher y Jewitt (1988)[2]. Estos autores proponen un estimador alternativo de la matriz de varianzas y covarianzas que coincide con el que hemos llamado HC2.

Chesher y Jewitt prueban que este estimador es insesgado bajo homoscedasticidad y tiene un sesgo menor que el estimador original de Eicker-White bajo heteroscedasticidad condicional.

4. Contrastes de la matriz de información dinámica en el caso de regresión lineal

Aplicamos los resultados de las dos secciones anteriores al caso de los contrastes de la matriz de información dinámica. Estos contrastes fueron propuestos originalmente por White (1985).

Presentamos un marco general para contrastar la especificación en el caso de regresión lineal. Este marco contiene un amplio rango de contrastes estándar y sugiere algunos nuevos que son interesantes. Entre los nuevos tenemos contrastes para correlación serial que varía sistemáticamente, contrastes de efectos del tipo ARCH-M que varían sistemáticamente, tests consistentes a heteroscedasticidad para efectos ARCH-M y contrastes de efectos ARCH no simétricos. También proponemos nuevas versiones de contrastes para efectos ARCH.

Proponemos versiones computacionalmente simplificadas para correlación serial consistentes a heteroscedasticidad. Presentamos varias versiones asintóticamente equivalentes a los contrastes anteriores con propiedades de nuestras finitas posiblemente mejoradas.

A lo largo de esta sección usaremos el siguiente ejemplo para ser más concretos y conseguir una mayor simplicidad.

Vamos a considerar la siguiente densidad:

$$y_t/F_{t-1} \sim N(\alpha+\rho\, y_{t-1}+\beta\, x_t,\ \sigma^2) \tag{4.0}$$

donde F_{t-1} representa la información contenida en las variables predeterminadas. Sólo incluimos una variable dependiente retrasada y una variable independiente. La generalización es obvia.

El logaritmo de la verosimilitud para este modelo es

$$\ln f_{t/t-1}(\omega^t, \theta) = \ln \sqrt{2\pi} - \ln \sigma - \{(y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t)^2 / 2\, \sigma^2\}$$

donde $\quad \theta \equiv (\alpha,\ \rho,\ \beta,\ \sigma)$

Definamos $u_t(\theta) \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad$ y

$r_t \equiv u_t(\hat{\theta}_n)$,

donde $\hat{\theta}_n$ es el estimador de máxima verosimilitud de θ (mínimos cuadrados ordinarios).

Así pues el gradiente condicional de la observación t es:

$$\nabla_\theta \ln f_{t/t-1}(\omega^t, \theta) = \left(\frac{u_t(\theta)}{\sigma^2}, \frac{u_t(\theta) y_{t-1}}{\sigma^2}, \frac{u_t(\theta) x_t}{\sigma^2}, \frac{u_t(\theta)^2 - \sigma^2}{\sigma^3}\right)$$

y los gradientes condicionales con λ retrasos serán:

$$\nabla_\theta \ln f_{t-\lambda/t-\lambda-1}(\omega^{t-1}, \theta) = \left(\frac{u_{t-\lambda}(\theta)}{\sigma^2}, \frac{u_{t-\lambda}(\theta) y_{t-\lambda-1}}{\sigma^2}, \frac{u_{t-\lambda}(\theta) x_{t-\lambda}}{\sigma^2}, \frac{u_{t-\lambda}(\theta)^2 - \sigma^2}{\sigma^3}\right)$$

Definimos

$$R_\lambda^0 \equiv \nabla_\theta \ln f_{t/t-1}(\omega^t, \theta)'\ \nabla_\theta \ln f_{t-\lambda/t-\lambda-1}(\omega^{t-\lambda}, \theta)$$

R_λ^0 serán los indicadores de la matriz de información dinámica que envuelve el retraso λ-ésimo del gradiente, evaluados en la hipótesis nula. Ahora centramos nuestra atención en R_1^0 que será en nuestro caso, omitiendo la dependencia entre u y θ, sacando factor común σ^{-4} y definiendo $\eta_t \equiv (u_t^2 - \sigma^2)$:

$$R_1^0 \equiv \sigma^{-4} \begin{pmatrix} u_t u_{t-1} & u_t y_{t-2} u_{t-1} & u_t x_{t-1} u_{t-1} & u_t \eta_{t-1}/\sigma \\ u_t y_{t-1} u_{t-1} & u_t y_{t-1} y_{t-2} u_{t-1} & u_t y_{t-1} x_{t-1} u_{t-1} & u_t y_{t-1} \eta_{t-1}/\sigma \\ u_t x_t u_{t-1} & u_t x_t y_{t-2} u_{t-1} & u_t x_t x_{t-1} u_{t-1} & u_t x_t \eta_{t-1}/\sigma \\ \eta_t u_{t-1}/\sigma & \eta_t y_{t-2} u_{t-1}/\sigma & \eta_t x_{t-1} u_{t-1}/\sigma & \eta_t \eta_{t-1}/\sigma^2 \end{pmatrix}$$

El valor esperado de cada uno de los elementos de las matrices R^0_λ, $\lambda=1,\ldots,$ p es cero bajo especificación correcta. Si no son cero la igualdad de la matriz de información condicional no se cumplirá y en general la matriz de varianzas y covarianzas de los coeficientes no será consistente. Para más detalles, ver White (1985).

Si basamos un contraste m en el elemento (1,1) de R^0_1 tenemos un contraste de la matriz de información dinámica para correlación serial de primer orden. Definimos IMSC1(1) (Information Matrix Serial Correlation, versión 1, orden 1) como IMSC1(1) $\equiv nR^2_0$ (el R^2 no ajustado por el uso de constante) de la regresión de mínimos cuadrados ordinarios de

$$(4.1) \qquad 1 \text{ sobre } r_t\, r_{t-1},\ \ r_t,\ \ r_t\, y_{t-1},\ \ r_t x_t,\ \ (r^2_t - \hat{\sigma}^2)$$

donde hemos multiplicado los regresores por potencias de σ por conveniencia. Observese que damos nombres a los contrastes que corresponden a los de la sección 3. IMSC1(1) $\overset{A}{\sim} \chi^2_1$ bajo la hipótesis nula de correcta especificación. Una versión simple corregida de grados de libertad puede ser definida como

$$\text{IMSC1(1)}^{**} = \frac{n-4}{n}\ \text{IMSC1(1)}$$

Un test equivalente asintóticamente a IMSC1(1) y IMSC1(1)** es un test t del coeficiente de $r_t\, r_{t-1}$ en la regresión anterior.

Otra versión utilizaría $r^*_t \equiv r_t/\surd(1-k_{tt})$ en la regresión (4.1). Otra vez, k_{tt} es el elemento diagonal t-ésimo de la matriz de predicción $X(X'X)^{-1}X'$, y X es la matriz de regresores. El test podría ser llevado a cabo regresando

$$(4.1') \qquad 1 \text{ en } r^*_t\, r_{t-1},\ \ r^*_t,\ \ r^*_t\, y_{t-1},\ \ r^*_t\, x_t,\ \ (r^{*2}_t - \hat{\sigma}^2)$$

y $IMSC1(1)' \equiv nR_0^2$ de la regresión de arriba. Una versión asintóticamente equivalente se puede obtener de la regresión

$$(4.1'') \qquad 1 \text{ en } \quad r^*_t r^*_{t-1}, \quad r^*_t, \quad r^*_t y_{t-1}, \quad r^*_t x_t, \quad (r^{*2}_t - \hat{\sigma}^2)$$

y $IMSC1(1)' \equiv nR_0^2$ de la regresión de arriba.

Dado que la matriz de información es diagonal por bloques, el Teorema 2.1 nos permite usar una versión asintóticamente equivalente de los contrastes anteriores regresando

$$(4.2) \qquad 1 \text{ en } \quad r_t r_{t-1}, \quad r_t, \quad r_t y_{t-1}, \quad r_t x_t$$

y $IMSC2(1) \equiv nR_0^2$ de la regresión anterior. Este es llamado un test LM robusto a heteroscedasticidad para correlación serial de primer orden por Alastair Hall (1985). Un test asintóticamente equivalente se obtendría regresando

$$(4.3) \qquad 1 \text{ en } \quad r^*_t r_{t-1}, \quad r^*_t, \quad r^*_t y_{t-1}, \quad r^*_t x_t$$

y $IMSC3(1) \equiv nR_0^2$ de la regresión anterior, o

$$(4.4) \qquad 1 \text{ en } \quad r^*_t r^*_{t-1}, \quad r^*_t, \quad r^*_t y_{t-1}, \quad r^*_t x_t$$

$IMSC3(1)' \equiv nR_0^2$ de la regresión anterior.

Los contrastes t de los coeficientes de los regresores $r_t r_{t-1}$, $r^*_t r_{t-1}$ y $r^*_t r^*_{t-1}$ en cada una de las anteriores regresiones serían contrastes válidos asintóticamente.

Los contrastes de la matriz de información dinámica para correlación serial de primer orden que hemos propuesto hasta ahora son válidos bajo heteroscedasticidad condicional. Los que vienen a continuación son válidos solamente bajo homoscedasticidad condicional.

Bajo homoscedasticidad condicional, podemos aplicar el Teorema 2.2 y obtener otro contraste asintóticamente equivalente a los anteriores usando la regresión de mínimos cuadrados ordinarios

(4.5) $\quad r_t$ sobre $\quad r_{t-1}, \quad 1, \quad y_{t-1}, \quad x_t.$

Llamaremos R^2 al coeficiente de determinación, ajustado por el uso de una constante, de la regresión anterior. Entonces bajo la hipótesis nula de especificación correcta, IMSC4(1) $\equiv nR^2$ se comportará asintóticamente como una variable aleatoria χ^2_1. Un contraste t del coeficiente de r_{t-1} en la anterior regresión es también un contraste válido.

Un contraste asintóticamente equivalente a los que se basan en la regresión (4.5) puede basarse en la regresión

(4.6) $\quad r^*_t$ en $\quad r^*_{t-1}, \quad 1, \quad y_{t-1}, \quad x_t$

y IMSC5(1)' $\equiv nR^2$ de la anterior regresión. De nuevo IMSC5(1)' $\overset{A}{\sim} \chi^2_1$ bajo la hipótesis nula de especificación correcta.

Uno podría llevar a cabo versiones corregidas de grados de libertad análogas a las propuestas en la sección 3. Los contrastes t de la significatividad de los coeficientes de r_{t-1} y r^*_{t-1} en las anteriores regresiones son también versiones asintóticamente válidas de los contrastes m propuestos.

El contraste IMSC4(1) es la forma usual del contraste LM de autocorrelación serial. Su versión t ha sido llamada por Durbin el procedimiento alternativo para computar la h de Durbin; véase Durbin (1970), pag. 420.

El contraste de la matriz de información dinámica para autocorrelación de primer orden se puede extender fácilmente para hacer contrastes conjuntos de p órdenes de autocorrelación serial, posiblemente en presencia de heteroscedasticidad condicional de forma desconocida. El contraste (en una de sus versiones) se puede computar regresando:

(4.8) 1 sobre $r^*_t r^*_{t-1}$, $r^*_t r^*_{t-2}, \ldots,$ $r^*_t r^*_{t-p}$, r^*_t, $r^*_t y_{t-1}$, $r^*_t x_t$

entonces $IMSC3(p)' \equiv nR^2_0 \overset{A}{\sim} \chi^2_p$ bajo la hipótesis nula.

Un test F de la significatividad de los p primeros coeficientes sería también asintóticamente válido. El contraste IMSC3(p)' está relacionado con lo que Domowitz y Hakkio (1984, pag. 7) llaman contrastes de correlación serial robustos a heteroscedasticidad. Nuestro procedimiento, sin embargo, es más simple computacionalmente y debería tener un mejor comportamiento en muestras finitas.

En el caso de homoscedasticidad condicional, el contraste IMSC3(p)' se simplifica y da lugar al contraste usual LM de autocorrelación serial de Breusch (1978), pag. 353 y Godfrey (1978), pag. 1298 (excepto por el uso de r_t en vez de r^*_t).

Como señalan Breusch y Godfrey, un contraste LM contra errores de media móvil de orden p toma exactamente la misma forma que un contraste LM contra errores autorregresivos de orden p. Este resultado también es aplicable a sus homónimos robustos a heteroscedasticidad. Nótese que los rechazos de la hipótesis nula basados en estos contrastes no nos dicen cual es la causa de que existan errores autocorrelacionados. En particular algunos modelos bilineales pueden causar errores autocorrelacionados, véase por ejemplo Granger y Andersen (1978, páginas 42, 53 y 63). La omisión de variables y/o una forma funcional incorrecta también pueden ser causa de que haya errores autocorrelacionados.

Hasta ahora hemos concentrado nuestros esfuerzos en los elementos (1,1) de las matrices R_{λ}^{0}, $\lambda=1,\ldots,p$, las matrices de retrasos λ-ésimos de indicadores de la matriz de información dinámica.

Sin embargo, no hay razón para pensar que los rechazos basados en los elementos (1,1) son más serios que aquellos basados en otros elementos de R_{λ}^{0}. En un modelo de regresión con variable dependiente retrasada, la autocorrelación en los errores hace que el estimador del parámetro de la variable dependiente retrasada sea inconsistente. Por tanto, rechazos basados en otros indicadores de la matriz de información dinámica pueden detectar también errores en la especificación que provoquen la inconsistencia de los estimadores de los parámetros de la media condicional.

Los tests de autocorrelación serial están considerando errores en la especificación en una dirección particular. Los contrastes de la matriz de información dinámica sugieren otras direcciones importantes en las que analizar la especificación de un modelo dado.

La popularidad de los tests de autocorrelación se debe en parte a su simplicidad computacional y en parte a nuestra capacidad para escribir modelos alternativos plausibles (en contra de los cuales el test tendrá poder máximo localmente) que podemos estimar.

Resulta que los contrastes de la matriz de información dinámica son fáciles de computar, podemos formular alternativas plausibles (contra las cuales serían tests LM) y podemos estimar esas alternativas.

Queremos aquí resaltar que un rechazo basado en un contraste de la matriz de información dinámica no implica de ninguna manera la aceptación de la hipótesis contra la cual el contraste tiene poder local máximo. Entre otras cosas porque típicamente podremos identificar varias alternativas diferentes contra las cuales el contraste de la matriz de información dinámica tendrá poder local máximo.

Es bien sabido que los contrastes de autocorrelación de los errores de un modelo gaussiano pueden considerarse como contrastes LM en los cuales la hipótesis nula es, digamos,

$y_t/F_{t-1} \sim N(\alpha+\rho\, y_{t-1}+\beta\, x_t,\ \sigma^2)$ o análogamente

(4.9) $\quad u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \quad$ y

$$u_t/F_{t-1} \sim N(0,\ \sigma^2)$$

Podemos escribir una hipótesis alternativa como

$$u_t \equiv y_{t-1} - \alpha - \rho\, y_{t-1} - \beta\, x_t \quad \text{y}$$

(4.10) $\quad u_t = \gamma_1\, u_{t-1} + \gamma_2\, u_{t-2} + \ldots + \gamma_p\, u_{t-p} + \xi_t \quad$ y

$$\xi_t/F_{t-1} \sim N(0,\ \sigma_\xi^2)$$

Resulta que en muchos casos los contrastes de la matriz de información dinámica se pueden reformular en un marco análogo al descrito en las ecuaciones (4.9) y (4.10).

Volviendo a nuestro ejemplo del principio de esta sección, un contraste de la matriz de información dinámica basado en el elemento (2,1) de R_1^0 se puede considerar como un contraste LM robusto a heteroscedasticidad de (4.9) contra la alternativa

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \quad \text{y}$$

(4.11) $$u_t = \gamma_1 \, y_{t-1} \, u_{t-1} + \xi_t \quad \text{y}$$

$$\xi_t / F_{t-1} \sim N(0, \sigma^2_\xi)$$

Esto puede considerarse como un contraste de autocorrelación serial que tiene un parámetro que varía sistemáticamente. Bajo heteroscedasticidad condicional, un test de la matriz de información dinámica basado en este indicador se puede computar como la regresión de

(4.12) $$1 \text{ en } \quad r^*_t \, y_{t-1} \, r^*_{t-1}, \quad r^*_t, \quad r^*_t \, y_{t-1}, \quad r^*_t \, x_t$$

donde $r^*_t \equiv r_t / \surd(1 - k_{tt})$

y $IMVSC3(1) \equiv nR^2_0 \overset{A}{\sim} \chi^2_1$ bajo la hipótesis nula de ausencia de autocorrelación serial.

IMVSC3(1) quiere decir matriz de información, correlación serial que varía, versión 3, orden 1. Análogamente se puede hacer un test t del coeficiente de $r^*_t \, y_{t-1} \, r^*_{t-1}$ en la regresión anterior.

Bajo homoscedasticidad condicional, se puede computar un contraste asintóticamente equivalente haciendo la regresión

(4.13) $$r_t \text{ sobre } \quad y_{t-1} \, r_{t-1}, \quad 1, \quad y_{t-1}, \quad x_t$$

Entonces, el contraste $IMVSC4(1) \equiv nR^2$ de la regresión anterior. Este es el contraste usual de los multiplicadores de Lagrange para la alternativa (4.11). Un contraste asintóticamente equivalente sería el contraste t sobre el coeficiente de $y_{t-1} \, r_{t-1}$ en la regresión anterior. Podríamos presentar también versiones asintóticamente equivalentes de los contrastes anteriores de la misma forma que los hicimos en el caso de autocorrelación de primer orden.

Es conveniente resaltar que si ocurren rechazos basados en (4.12) o (4.13), uno puede decidir estimar el modelo alternativo (4.11), si hay razones económicas para creer que el modelo (4.11) es un modelo sensato.

Si uno estimase (4.11), sería en general apropiado contrastar las restricciones de factores comunes implícitas. Volveremos a este punto en la sección 6. Incluso si los contrastes apropiados de restricciones de factores comunes no rechazasen la hipótesis nula, sería necesario continuar los contrastes de especificación del modelo (4.11). Sería apropiado usar contrastes de la Matriz de Información y de los multiplicadores de Lagrange.

Sería sensato esperar que los contrates IMVSC (Matriz de Información, correlación serial que varía) tuviesen poder contra errores en la especificación de la media condicional, en particular contra forma funcional incorrecta y/o una estructura dinámica insuficientemente rica.

Un ejemplo del tipo de no linealidades que los contrastes de la Matriz de Información Dinámica pueden detectar es el siguiente. Consideremos de nuevo el modelo (4.0). Un contraste de la matriz de información dinámica basado en el elemento (1,3) de R_1^0 puede ser considerado como un contraste LM para una alternativa tal como $u_t = \Upsilon x_{t-1} u_{t-1} + \xi_t$ o escrita con más detalle.

$$(4.14)\quad y_t - \rho\, y_{t-1} = \alpha + \beta\, x_t + \gamma\, y_{t-1}\, x_{t-1} - \alpha\,\gamma\, x_{t-1} - \beta\gamma\, x_{t-1}^2 - \Upsilon\,\rho\, x_{t-1}\, y_{t-2} + \xi_t$$

Lo bien que esta forma aproxima formas funcionales desconocidas es un tema abierto. Sin embargo, esta parece una extensión natural de las restricciones de factores comunes implícitamente implicadas por un término de error que sigue un proceso autorregresivo de primer orden. En nuestro caso, este sería:

$$(4.15)\qquad y_t-\rho\, y_{t-1} = \alpha+\beta\, x_t+\gamma\, y_{t-1}-\alpha\,\gamma-\gamma\,\rho\, y_{t-2}-\gamma\,\beta\, x_{t-1}+\xi_t$$

Observamos que los únicos términos en (4.15) que no están en (4.14) son los términos en y_{t-2} y y_{t-1}, mientras que hay tres términos en (4.14) que no están en (4.15), estos son los términos no lineales en:

$$y_{t-1}\, x_{t-1},\quad y_{t-2}\, x_{t-1},\quad x_{t-1}^2$$

La ecuación (4.14) puede ser considerada como una aproximación a una forma funcional no lineal tal como

$$(4.16)\qquad \frac{y_t-\rho\, y_{t-1}}{x_{t-1}} = \gamma\, y_{t-1}-\gamma\,\rho\, y_{t-2}-\beta\,\gamma\, x_{t-1}+u_t$$

donde tendríamos

$$u_t = \beta\,\frac{x_t}{x_{t-1}} + \frac{\alpha}{x_{t-1}} + \frac{\xi_t}{x_{t-1}}$$

así, podríamos tomar a x_{t-1} como y_{t-1} y tendríamos una tasa de aumento en y_t (con ρ=1). Granger y Andersen (1978, pags. 14-17) dan varios ejemplos de modelos de series temporales univariantes que pueden ser aproximados por modelos bilineales de series temporales.

Estos ejemplos sugieren generalizaciones al contexto de regresión en el espíritu del que acabamos de exponer.

La posibilidad de estimar dos modelos anidados sugiere homónimos de los contrastes m presentados, que estarían en el espíritu de los contrastes de Wald y del Cociente de las Verosimilitudes. La utilidad relativa de los diferentes enfoques debería poderse establecer usando experimentos de Monte Carlo.

Las verosimilitudes artificiales de la Definición D.4 sugieren alternativas gaussianas que se pueden construir a partir de ellas básicamente completando el cuadrado del exponente de la verosimilitud. Este procedimiento se puede generalizar a otras densidades de la familia exponencial.

Los contrastes de la matriz de información dinámica podrían también ser útiles para aislar direcciones en las que queríamos centrar nuestros esfuerzos de reespecificación. Unos resultados recientes de Newey (1985, pags. 1058 - 1059) nos sugieren precisamente esto. Los resultados de Newey implican que si hacemos varios tests m unidireccionales, éstos pueden ser usados para distinguir entre diferentes formas de errores de especificción para pequeñas desviaciones de la hipótesis nula. En particular el error de especificación que corresponde el contraste m con mayor valor del coeficiente t se puede suponer que es el que tiene más probabilidades de ocurrir. Este procedimiento será localmente válido asintóticamente.

Consideremos ahora una alternativa más general que (4.11)

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad y$$

$$(4.17) \qquad u_t = \gamma_1\, y_{t-1}\, u_{t-1} + \gamma_2\, y_{t-2}\, u_{t-2} + \ldots + \gamma_p\, y_{t-p}\, u_{t-p} + \xi_t \qquad y$$

$$\xi_t / F_{t-1} \sim N(0,\, \sigma^2_\xi)$$

Bajo heteroscedasticidad condicional un contraste de la matriz de información dinámica basado en los p elementos (2,1) de las matrices R^0_λ ($\lambda = 1,\ldots,p$) de indicadores de la matriz de información dinámica se puede computar regresando

$$(4.18) \quad 1 \text{ en } \quad r^*_t\, y_{t-1}\, r^*_{t-1},\; r^*_t\, y_{t-2}\, r^*_{t-2}, \ldots, r^*_t\, y_{t-p}\, r^*_{t-p},\; r^*_t\, y_{t-1}, r^*_t\, x_t,\; r^*_t$$

De nuevo, bajo la hipótesis nula de ausencia de especificación, el contraste

$$IMVSC3(P)' \equiv nR^2 \overset{A}{\sim} \chi^2_p$$

Bajo homoscedasticidad condicional, podemos computar un contraste asintóticamnte equivalente usando la regresión

(4.19) $\quad r_t$ en $y_{t-1}\, r_{t-1},\ y_{t-2}\, r_{t-2}, \ldots, y_{t-p}\, r_{t-p},\ 1,\ y_{t-1},\ x_t$

y $IMVSC4(p) \equiv nR^2 \overset{A}{\sim} \chi^2_p$ bajo especificación correcta.

De nuevo, podríamos estimar (4.17) si ocurriese un rechazo, pero esta no es necesariamente la forma adecuada de proceder.

Resumiendo, los contrastes de la matriz de información dinámica basados en la submatriz (3x3) superior izquierda de las matrices R^0_λ, $\lambda = 1, \ldots, p$ son fáciles de computar y de interpretar. Pueden ser considerados como contrastes de especificación de la media condicional de y_t, y serán sensibles a una estructura dinámica insuficientemente rica y/o una forma funcional incorrecta. Pueden ser considerados como contrastes de autocorrelación serial generalizada. Estos contrastes son sensibles a autocorrelación serial que varía sistemáticamente con los regresores y regresores retrasados. Esta es una similitud interesante con los contrastes de la matriz de información estática.

Los tests de la matriz de información basados en el subvector 3x1 de la última columna de R^0_1 pueden también ser considerados como contrastes de especificación de la media condicional de y. Sin embargo, su interpretación es ligeramente diferente.

Un contraste de la matriz de información basado en el elemento (1,4) de R^0_1 se puede computar como la regresión de

$$(4.20) \qquad 1 \text{ sobre } \quad r^*_t\,(r^{*2}_{t-1} - \hat{\sigma}^2), \quad r^*_t, \quad r^*_t\, y_{t-1}, \quad r^*_t\, x_t$$

Entonces un contraste de la matriz de información dinámica para efectos ARCH en media versión 2, orden 1 será IM ARCH-M2(1)' $\equiv nR_0^2 \overset{A}{\sim} \chi_1^2$, bajo la hipótesis nula de especificación correcta. Un test t del coeficiente de $r^*_t\,(r^{*2}_{t-1} - \hat{\sigma}^2)$ será asintóticamente apropiado.

Este contraste puede ser considerado como un contraste LM robusto a heteroscedasticidad condicional (heteroscedasticidad de otros tipos diferentes al ARCH) que contrasta la posible existencia de efectos del tipo de ARCH en media. Véase Engle, Lilien y Robins (1985) para una definición de este modelo.

Una versión inconsistente a heteroscedasticidad del test anterior se puede computar haciendo la regresión

$$(4.21) \qquad r^*_t \text{ sobre } \quad (r^{*2}_{t-1} - \hat{\sigma}^2), \quad 1, \quad y_{t-1}, \quad x_t$$

De nuevo nR^2 (ajustado por el uso de una constante) de la anterior regresión se comportará asintóticamente como una variable aleatoria χ_1^2 bajo la hipótesis nula de especificación correcta. Un contraste t del coeficiente de $(r^{*2}_{t-1} - \sigma^2)$ será también apropiado asintóticamente.

Es fácil demostrar que los contrastes anteriores pueden ser considerados contrastes LM para la alternativa siguiente

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad \text{y}$$

$$(4.22) \qquad u_t = \gamma(u^2_{t-1} - \sigma^2_\xi) + \xi_t \qquad \text{y}$$

$$\xi_t / F_{t-1} \sim N(0, \sigma^2_\xi)$$

La generalización a efectos ARCH-M más largos es inmediata.

Si uno se fija en el elemento (2,4) de R_1^0, entonces es fácil ver que una alternativa implícita contra la cual el contraste tendrá poder máximo local asintóticamente va a ser:

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad y$$

$$u_t = \gamma_1\, y_{t-1}(u_{t-1}^2 - \sigma_u^2) + \xi_t \qquad y \tag{4.23}$$

$$\xi_t / F_{t-1} \sim N(0,\, \sigma_\xi^2)$$

El modelo anterior puede ser considerado como un modelo del tipo ARCH-M generalizado, en el cual el parámetro varía sistemáticamente con uno de los regresores. Un contraste de la matriz de información dinámica basado en el elemento (2,4) de la matriz R_1^0, se puede computar a partir de la regresión

$$1 \quad \text{sobre} \quad r_t^* \, y_{t-1}(r_{t-1}^{*2} - \hat{\sigma}^2), \quad r_t^*, \quad r_t^*\, y_{t-1}, \quad r_t^*\, x_t \tag{4.24}$$

Se puede computar una versión inconsistente a heteroscedasticidad del anterior contraste haciendo la regresión

$$r_t^* \quad \text{sobre} \quad y_{t-1}(r_{t-1}^{*2} - \hat{\sigma}^2), \quad 1, \quad y_t, \quad x_t \tag{4.25}$$

El estadístico nR^2 de la regresión anterior será asintóticamente χ_1^2 bajo la hipótesis nula de especificación correcta. También será apropiado asintóticamente el contraste t de significatividad del coeficiente de $y_{t-1}(r_{t-1}^{*2} - \hat{\sigma}^2)$.

La extensión para contrastar la posible existencia de efectos del tipo ARCH-M puede realizarse incluyendo términos adicionales del tipo $r_t^*\, y_{t-\lambda}(r_{t-\lambda}^{*2} - \hat{\sigma}^2)$, $\lambda = 2, \ldots, p$ en

(4.24) o $y_{t-\lambda}(r^{*2}_{t-\lambda} - \hat{\sigma}^2)$ en (4.25)

Los contrastes de la matriz de información dinámica basados en los indicadores de la última fila de R^0_1 (y/o la última fila de cada una de las matrices R^0_λ, $\lambda = 1, \ldots,$ p) serán sensibles a errores en la especificación de la varianza condicional de y_t (y/o de la media condicional).

En particular estos contrastes pueden considerarse como contrastes LM para diferentes formas de efectos del tipo ARCH, pero también pueden ser sensibles a errores bilineales. De nuevo cabe resaltar lo inapropiado de la práctica de estimar un modelo contra el que el contraste tenga poder local máximo cuando se rechace el modelo considerado.

Un contraste basado en el elemento (4,1) de R^0_1 se puede computar, como fue propuesto originalmente por White (1985), con la regresión

$$(4.26)\ 1 \text{ sobre } \frac{(r_t^2 - \hat{\sigma}^2)\, r_{t-1}}{\hat{\sigma}^5},\ \frac{r_t}{\hat{\sigma}^4},\ \frac{r_t\, y_{t-1}}{\hat{\sigma}^4},\ \frac{r_t\, x_t}{\hat{\sigma}^4},\ \frac{r_t^2 - \hat{\sigma}^2}{\hat{\sigma}^5}$$

y entonces nR^2_0 de esta regresión se comportaría asintóticamente como una χ^2_1 bajo la hipótesis nula de especificación correcta. Un contraste equivalente consistiría en mirar el coeficiente t del primer regresor en la regresión artificial anterior.

El contraste m de la regresión anterior (4.26) se puede simplificar. Podemos multiplicar cada regresor por potencias apropiadas de $\hat{\sigma}$ y obtendríamos un contraste equivalente. Además, a causa de la diagonalidad por bloques de la matriz de información bajo la hipótesis nula, el Teorema 2.1 nos permite computar un contraste asintóticamente equivalente regresando

(4.27) $\quad 1$ sobre $(r_t^2 - \hat{\sigma}^2)\, r_{t-1}, \quad (r_t^2 - \hat{\sigma}^2)$

Bajo la hipótesis nula, especialmente normalidad, se puede computar un contraste asintóticamente equivalente haciendo la regresión:

(4.28) $\quad (r_t^2 - \hat{\sigma}^2)$ sobre $r_{t-1}, \; 1$

O de manera análoga regresando

(4.29) $\quad r_t^2$ sobre $r_{t-1}, \; 1$ o bien

(4.30) $\quad r{*}_t^2$ sobre $r{*}_{t-1}, \; 1$

Un contraste asintóticamente válido consistiría en el test t del coeficiente de r_{t-1} o de $r{*}_{t-1}$ en las regresiones (4.29) o (4.30). Los estadísticos nR^2 de estas regresiones serán asintóticamente χ_1^2 bajo la hipótesis nula.

Si basamos un contraste de la matriz de información en los elementos (4,1) de las matrices $R_1^0, \ldots, R_p^0$, podemos computar este contraste con la siguiente regresión artificial

(4.31) $\quad r_t^2$ sobre $r_{t-1}, \ldots, r_{t-p}, \; 1$

y nR^2 (el R^2 ajustado por el uso de una constante) de esta regresión se comportará asintóticamente como una variable aleatoria χ_p^2 bajo la hipótesis nula de especificación correcta.

Se puede demostrar que estos contrastes pueden ser considerados como test LM contra alternativas ARCH no simétricas; véase Engle (1982, p. 993) para una definición de este tipo de proceso. Una alternativa podría ser, en este caso

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad y$$

$$(4.32) \qquad u_t/F_{t-1} \sim N(0,\, h_t) \qquad y$$

$$h_t = \alpha_0 + \alpha_1\, u_{t-1} + \alpha_2\, u_{t-2} + \ldots + \alpha_p\, u_{t-p}$$

Si basamos los contrastes de la matriz de información dinámica en los elementos (4,2) o (3,3) de R_1^0, $R_2^0, \ldots,$ R_p^0 estos contrastes serán de nuevo sensibles a alternativas del tipo de ARCH no simétrico. En particular los contrastes basados en el elemento (4,3) pueden ser considerados como contrastes LM contra la alternativa.

$$u_t \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t \qquad y$$

$$(4.33) \qquad u_t/F_{t-1} \sim N(0,\, h_t) \qquad y$$

$$h_t = \alpha_0 + \alpha_1\, x_{t-1}\, u_{t-1} + \alpha_2\, x_{t-2}\, u_{t-2} + \ldots \alpha_p\, x_{t-p}\, u_{t-p}$$

(con las restricciones apropiadas sobre las α para garantizar que h_t sea siempre positiva para cualquier realización de las u).

La computación de los contrastes basados en los elementos (4,2) y (4,3) es similar a la de los contrastes basados en los elementos (4,1).

Los contrastes de la matriz de información dinámica basados en los elementos (4,4) de las matrices de indicadores R_1^0, $R_2^0, \ldots,$ R_p^0, serán sensibles a efectos del tipo ARCH simétrico y a bilinealidades. Se pueden computar como fue originalmente propuesto por White (1985) regresando

$$(4.34) \qquad 1 \text{ en } (r_t^2 - \hat{\sigma}^2)\,(r_{t-1}^2 - \hat{\sigma}^2)/\hat{\sigma}^6, \quad r_t/\hat{\sigma}^2, \quad r_t\, y_{t-1}/\hat{\sigma}^2$$

$$r_t\, x_t/\hat{\sigma}^2, \qquad (r_t^2 - \hat{\sigma}^2)/\hat{\sigma}^3$$

y nR_0^2 de esta regresión se comporta asintóticamente como una variable aleatoria χ_1^2 bajo la hipótesis nula; análogamente se podría hacer un test t del coeficiente del primer regresor en la regresión (4.34). Podemos multiplicar los regresores de (4.34) por potencias de σ sin cambiar el R_0^2. Gracias a la diagonalidad por bloques de la matriz de información, podemos computar un contraste asintóticamente equivalente usando la regresión auxiliar

$$(4.35) \qquad 1 \text{ sobre } (r_t^2 - \hat{\sigma}^2)\,(r_{t-1}^2 - \hat{\sigma}^2), \qquad (r_t^2 - \hat{\sigma}^2)$$

Análogamente podemos realizar el contraste regresando

$$(4.36) \qquad r_t^2 \text{ sobre } r_{t-1}^2,\ 1$$

y tendríamos que nR^2 (R^2 ajustado por el uso de constante) se comportaría asintóticamente como una variable aleatoria χ_1^2 bajo la hipótesis nula de especificación correcta; también se puede usar como contraste el estadístico t del coeficiente de r_{t-1}^2. Esta es la forma del contraste de residuos ARCH tal como fue propuesto originalmente por Engle (1982), aunque en ese trabajo se da una versión más general de (4.36) que contrasta varios efectos ARCH al mismo tiempo

$$r_t^2 \text{ sobre } r_{t-1}^2,\ r_{t-2}^2, \ldots,\ r_{t-p}^2,\ 1$$

Un contraste esencialmente análogo ha sido propuesto por Granger y Andersen (1978, pp. 43, 55, 63) para detectar la presencia de bilinealidades en modelos univariantes de series temporales.

5. Una alternativa general para los contrastes de la Matriz de Información dinámica en el caso de regresión lineal

En esta sección identificamos una alternativa general contra la cual los contrastes de la matriz de información dinámica pueden ser considerados contrastes LM. Como subproducto proponemos una generalización del modelo de regresión lineal que incluye los modelos de regresión bilineales, modelos ARCH, ARCH en media, modelos que dependen del estado (state dependent), algunos casos de mínimos cuadrados generalizados y algunos procesos nuevos. Algunos de estos procesos son: autocorrelación que varía sistemáticamente y procesos de ARCH en media que varían sistemáticamente. Las propiedades y utilidad de estos nuevos procesos no son conocidos todavía en general.

Algunos casos especiales de estos procesos han recibido considerable atención por parte de investigadores teóricos y aplicados. En particular los procesos bilineales y ARCH están siendo usados ampliamente en la profesión.

Esta alternativa general la podemos obtener fijándonos en que los contrastes de la Matriz de Información dinámica tienen una estructura común en el contexto de regresión lineal. Supongamos que tenemos un conjunto de k regresores (incluyendo una constante, y posiblemente variables dependientes retrasadas). Entonces los contrastes de la matriz de información dinámica basados en los elementos de las matrices de indicadores de la Matriz de Información dinámica $R_1^0, \ldots, R_p^0$, pueden ser considerados como contrastes LM (en sus versiones consistentes e inconsistentes a heteroscedasticidad) contra una alternativa general.

Consideremos de nuevo el modelo de regresión lineal $y_t = X_t\,\beta + u_t$ y $u_t/F_{t-1} \sim N(0, \sigma_u^2)$ donde X_t es una matriz nxk de observaciones y la primera variable $x_{1t} = 1$, para todo $t = 1, \ldots, n$.

Podemos considerar la siguiente alternativa como la alternativa general que los contrastes de la Matriz de Información dinámica están considerando

$$(5.1) \qquad \begin{pmatrix} u_t \\ \sigma_t^2 \end{pmatrix} = A\ \underline{\gamma}\ B + \xi_t \qquad \xi_t/F_{t-1} \sim N(0,\ \Omega)$$

$$\text{donde } A = \begin{pmatrix} 1, & x_{2t}, & x_{3t}, \ldots, & x_{kt} \\ 1, & 0, & 0\ , & ,\ 0 \end{pmatrix}$$

es una matriz 2xk, $\underline{\gamma}$ es la matriz kx2k de constantes con elementos.

$$\underline{\gamma} = \{\gamma_{ij}\} \qquad i = 1, \ldots, k, \quad j = 1, \ldots, 2k$$

y B es el vector columna (2kx1) tal que

$$B' = \{u_{t-1}, \ldots, u_{t-p}, x_{2t-1}u_{t-1}, \ldots, x_{2t-p}u_{t-p}, \ldots, x_{kt-1}u_{t-1}, \ldots, x_{kt-p}u_{t-p},$$

$$(u_{t-1}^2-\sigma_u^2), (u_{t-2}^2-\sigma_u^2), \ldots, (u_{t-p}^2-\sigma_u^2),\ x_{2t-1}(u_{t-1}^2-\sigma_u^2), \ldots, x_{kt-1}(u_{t-1}^2-\sigma_u^2), \ldots$$

$$x_{2t-p}(u_{t-p}^2-\sigma_u^2), \ldots, x_{kt-p}(u_{t-p}^2-\sigma_u^2)\}$$

Este modelo alternativo es muy general y engloba como casos particulares la autocorrelación de orden p, la autocorrelación con parámetros variables, modelos del tipo ARCH en media, con parámetros constantes y variables, ARCH y modelos bilinelaes. El modelo bilineal aparecería cuando algunos de los regresores son ruidos blancos retrasados y algunos de los elementos de $\underline{\gamma}$ se restringen apropiadamente. Conviene hacer notar que si incluimos regresores contemporáneos en B, tendríamos la mayoría de las alternativas que los contrastes de la Matriz de Información estática consideran.

El model (5.1) puede servir como punto de partida para formular alternativas no lineales simples al modelo de regresión lineal. Algunos de los modelos que esta alternativa general engloba han generado ya un interés considerable entre los investigadores teóricos y aplicados, aunque el análisis del comportamiento de estos modelos simples está lejos de ser trivial en muchos casos. En Engle y Bollerslev (1986) y Cartwright (1985) se encuentran listas de referencias de estos estudios para algunos de estos casos.

El estudio de la flexibilidad del modelo (5.1) para aproximar nolinealidades omitidas merece más atención, y estará relacionado con el poder de los contrastes de la matriz de información dinámica.

Una posible generalización del modelo (5.1) sería permitir que algunos elementos de $\underline{Y}$ variasen sistemáticamente con algunos regresores, errores u otras variables . Otra posibilidad sería permitir que los elementos de B no estuviesen restringidos a los valores pasados de los regresores y/o errores, aunque la utilidad de los sistemas anticipativos en economía puede no estar muy clara. Otra posibilidad sería generalizar la alternativa al caso multivariante para incluir el modelo multivariante bilineal de Subba Rao (1986), caso (2.1) como un caso especial de esta alternativa general. Este modelo de Subba Rao sería un caso particular de nuestro modelo ya que no incluye los productos cruzados no lineales que están incluídos en el modelo (5.1).

Los modelos dependientes de los estados (state dependent) de Priestley (1985) pueden considerarse como casos particulares de (5.1) con parámetros variables. Priestley usa desarrollos de Taylor para justificar la generalidad de su modelo dependiente de los estados para aproximar formas funcionales no lineales desconocidas en el contexto de series temporales univariantes. Puede ser posible utilizar argumentos similares para justificar el modelo (5.1) con parámetros que varían sistemáticamente como una generalización del modelo (4.2) de Priestley al caso de regresión. Cartwright (1985) da una lista de referencias para los modelos dependientes de los estados.

Adicionalmente, podemos usar el modelo (5.1) para computar contrastes m de variables omitidas incluyendo a las posibles variables omitidas en el vector B y restringiendo $\underline{Y}$ apropiadamente.

6. Otras aplicaciones de los contrates m

6.1. Restricciones de factores comunes

Otra aplicación natural del marco de los contrastes m dinámicos es el caso de contrastar restricciones de factores comunes en presencia de heteroscedasticidad condicional de forma desconocida.

La hipótesis nula será la existencia de restricciones de factores comunes:

$$(5.1) \qquad y_t = (1 - \alpha(L))\, y_t + \alpha(L)\, X_t\, \beta + \xi_t$$

donde $\xi_t/F_{t-1} \sim N(0, \alpha^2)$ y además y_t es un vector nx1 de observaciones de una variable dependiente, X es una matriz nxk con rango completo por columnas de observaciones de los regresores; β es un vector fijo kx1 de coeficientes desconocidos y

$$\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_e L^e \qquad y\ L^i x_t \equiv x_{t-i}$$

Engle destaca que la alternativa de ausencia de restricciones de factores comunes se puede reparametrizar como:

$$(6.2) \qquad \alpha(L)\, y_t = \alpha(L)\, x_t\, b + \Upsilon(L)\, x_t' + \xi_t$$

donde

$$\Upsilon(L) = \Upsilon_0(L) + \Upsilon_1(L) + \ldots + \Upsilon_\lambda(L)$$

$$\Upsilon_i(L) = \Upsilon_{i1} L + \Upsilon_{i2} L^2 + \ldots + \Upsilon_{ip} L^p \quad i=1, \ldots, k$$

$$\Upsilon_i(0) = 0$$

El modelo alternativo se puede escribir como

(6.3) $y^*_t = x^*_t \beta + \Upsilon(L)\, x'_t + \xi_t$

donde $y^*_t = \alpha(L)\, y_t$ y $x^*_t = \alpha(L)\, x_t$

Nosotros seguimos aquí la notación de Domowitz y Hakkio (1984) y llamamos X a la matriz de dimensión (n-p)xpk de los regresores retrasados $X = (X_{-1}, \ldots, X_{-p})$. Sea X^* la matriz (n-p)xk de regresores transformados $X^* = \alpha(L)\ X$ (téngase en cuenta que ajustamos el tamaño de la muestra según el número de retrasos que usamos), sea i el vector de unos de dimensión (n-p)x1. Sea R la matriz diagonal (n-p)x(n-p) tal que $R = \text{diag}\ (r_{p+1}, \ldots, r_n)$, donde r_t son los residuos estimados bajo la hipótesis nula de factores comunes, y $r \equiv (r_{p+1}, r_{p+2}, \ldots, r_n)'$

$$r = (I - X^*(X^{*\prime}\, X^*)^{-1}\, X^{*\prime})\ y$$

Un contraste asintóticamente válido de estas restricciones de factores comunes se puede llevar a cabo usando la regresión artificial

(6.4) i sobre $R\ X'$, $R\ X^*$

y $(n-p)\ R^2_0$ de la anterior regresión se comportaría asintóticamente como una variable aleatoria χ^2 con pk grados de libertad bajo la hipótesis nula de existencia de factores comunes. Este contraste es una versión computacionalmente simplificada del contraste de restricciones de factores comunes robusto a heteroscedasticidad de Domowitz y Hakkio (1984).

Un contraste asintóticamente equivalente con, posiblemente, mejores resultados en muestras finitas se podría computar sustituyendo en las anteriores expresiones r_t por $r^*_t = r_t/\surd(1 - k_{tt})$.

En el caso de autocorrelación que varía sistemáticamente, podemos encontrar variables transformadas análogas a y^*_t y x^*_t y contrastar restricciones de factores comunes más generales de forma similar a la que hemos presentado aquí.

7. Conclusiones y sugerencias para investigaciones futuras

En este trabajo hemos obtenido resultados generales que simplifican la computación de los test m y los hacen fácilmente interpretables. También se ha demostrado que los test m pueden ser considerados en general como test de los multiplicadores de Lagrange.

En la sección 3 discutimos la computación e interpretación de los test m en el contexto de la regresión lineal. Presentamos versiones de los test consistentes a heteroscedasticidad e inconsistentes a heteroscedasticidad con propiedades de muestras finitas posiblemente mejoradas. Un caso especial es la computación de los contrastes LM bajo heteroscedasticidad condicional de forma desconocida.

En la sección 4 centrábamos la discusión en el caso de los test de la matriz de información dinámica en el contexto de regresión lineal. Demostramos que muchos contrastes tienen interpretaciones familiares y se proponen nuevos test para mala especificación dinámica.

Un subproducto de esta investigación es un test para correlación serial en presencia de heteroscedasticidad de forma desconocida. Este test tendría buen comportamiento en muestras finitas y puede ser computado con una regresión de mínimos cuadrados ordinarios.

También damos contrastes para efectos del tipo ARCH en media que son robustos a otros tipos de heteroscedasticidad.

En la sección 5 identificamos una alternativa general en contra de la cual los test de la matriz de información dinámica pueden ser considerados test de los Multiplicadores de Lagrange. También proponemos una generalización del modelo de regresión lineal que incluye procesos bilineales, ARCH y ARCH en media entre otros. Este marco también incluye nuevos modelos tales como procesos con correlación serial que varía sistemáticamente y procesos del tipo ARCH en media que varían sistemáticamente.

En la sección 6 aplicamos el marco de los test m para obtener un test para restricciones de factores comunes en presencia de heteroscedasticidad de forma desconocida. Este contraste puede ser computado con una regresión adicional de mínimos cuadrados ordinarios.

Los resultados que hemos obtenido sugieren varias áreas de investigación futura. Los teoremas de la sección 2 pueden servir como guía para llevar a cabo test m y test de los multiplicadores de Lagrange en contextos más generales que el considerado aquí. En particular pueden ser aplicados a otros modelos estimados por máxima verosimilitud. También pueden ser de utilidad para identificar alternativas que se hallan en la misma familia paramétrica que la hipótesis nula, contra la cual los tests m tendrán máximo poder local. El Teorema 2.2 puede ser válido también para sistemas de ecuaciones.

Los resultados de las secciones 3 y 4 sugieren que los estudios de simulación serán necesarios para comparar el comportamiento en pequeñas muestras de diferentes contrastes m.

La alternativa general presentada en la sección 5 claramente necesita un mayor estudio teórico y práctico en el futuro.

Otra área de interés sería estudiar hasta qué punto los resultados presentes se aplican a modelos estimados por el método generalizado de momentos.

APENDICE MATEMATICO

Todos los símbolos y definiciones son los dados en el texto.

PRUEBA DEL TEOREMA 2.1:

$$n\ R_0^2 \equiv i'(\hat{m}\ \hat{l}_\beta\ \hat{l}_\alpha) \begin{Bmatrix} \hat{m}'\hat{m} & \hat{m}'\hat{l}_\beta & \hat{m}'\hat{l}_\alpha \\ \hat{l}_\beta'\hat{m} & \hat{l}_\beta'\hat{l}_\beta & \hat{l}_\beta'\hat{l}_\alpha \\ \hat{l}_\alpha'\hat{m} & \hat{l}_\alpha'\hat{l}_\beta & \hat{l}_\alpha'\hat{l}_\alpha \end{Bmatrix}^{-1} \begin{Bmatrix} \hat{m}' \\ \hat{l}_\beta' \\ \hat{l}_\alpha' \end{Bmatrix} i$$

escribiendo

$$\begin{Bmatrix} \hat{C}_{mm} & \hat{C}_{m\beta} & \hat{C}_{m\alpha} \\ \hat{C}_{\beta m} & \hat{C}_{\beta\beta} & \hat{C}_{\beta\alpha} \\ \hat{C}_{\alpha m} & \hat{C}_{\alpha\beta} & \hat{C}_{\alpha\alpha} \end{Bmatrix} \equiv \begin{Bmatrix} \hat{m}'\hat{m} & \hat{m}'\hat{l}_\beta & \hat{m}'\hat{l}_\alpha \\ \hat{l}_\beta'\hat{m} & \hat{l}_\beta'\hat{l}_\beta & \hat{l}_\beta'\hat{l}_\alpha \\ \hat{l}_\alpha\hat{m} & \hat{l}_\alpha'\hat{l}_\beta & \hat{l}_\alpha'\hat{l}_\alpha \end{Bmatrix}^{-1}$$

$$n\ R_0^2 = i'\ \hat{m}\ \hat{C}_{mm}\ \hat{m}'\ i + i'\ \hat{m}\ \hat{C}_{m\beta}\ \hat{l}_\beta'\ i + i'\ \hat{m}\ \hat{C}_{m\alpha}\ \hat{l}_\alpha'\ i$$

$$+ i'\ \hat{l}_\beta\ \hat{C}_{\beta m}\ \hat{m}'\ i + i'\ \hat{l}_\beta\ \hat{C}_{\beta\beta}\ \hat{l}_\beta'\ i + i'\ \hat{l}_\beta\ \hat{C}_{\beta\alpha}\ \hat{l}_\alpha'\ i$$

$$+ i'\ \hat{l}_\alpha\ \hat{C}_{\alpha m}\ \hat{m}'\ i + i'\ \hat{l}_\alpha\ \hat{C}_{\alpha\beta}\ \hat{l}_\beta'\ i + i'\ \hat{l}_\alpha\ \hat{C}_{\alpha\alpha}\ \hat{l}_\alpha'\ i$$

$= i'\ \hat{m}\ \hat{C}_{mm}\ \hat{m}'\ i + o_p(1)$ por el Teorema 3.3 de White(1985).

Podemos escribir la expresión anterior como

$$n\ R_0^2 = (n^{-1/2}\ i'\ \hat{m})\ n\ \hat{C}_{mm}\ (n^{-1/2}\ \hat{m}'\ i) + o_p(1)$$

y $n\ \hat{C}_{mm} = \hat{V}_n^{-1}$ donde

$$\hat{V}_n = \hat{m}'\,\hat{m}/n - (\hat{m}'\,\hat{l}_\beta\;/n,\;\;\hat{m}'\,\hat{l}_\alpha/n)\begin{pmatrix}\hat{l}_\beta'\,\hat{l}_\beta/n & \hat{l}_\beta'\,\hat{l}_\alpha/n\\ \hat{l}_\alpha'\,\hat{l}_\beta/n & \hat{l}_\alpha'\,\hat{l}_\alpha/n\end{pmatrix}^{-1}\begin{pmatrix}\hat{l}_\beta'\,\hat{m}/n\\ \hat{l}_\alpha'\,\hat{m}/n\end{pmatrix}$$

Por el Supuesto S.5 tenemos una Matriz de Información que es diagonal por bloques, por tanto los bloques fuera de la diagonal $\hat{l}_\alpha\,\hat{l}_\beta/n$ y $\hat{l}_\beta'\hat{l}_\alpha/n$ son matrices $o_p(1)$ y $\hat{l}_\alpha'\hat{l}_\alpha/n$ es $O_p(1)$ por los supuestos S.2 y S.4'(V).

Bajo correcta especificación $\hat{m}'\,\hat{l}_\alpha/n$ y $\hat{l}_\alpha'\,\hat{m}/n$ son $o_p(1)$ por el supuesto S.5; por consiguiente por una modificación del Corolario 2.36 de White(1984a):

$$\hat{V}_n = \hat{m}'\,\hat{m}/n - \hat{m}'\,\hat{l}_\beta/n\,(\hat{l}_\beta'\,\hat{l}_\beta/n)^{-1}\,\hat{l}_\beta'\,\hat{m}/n + o_p\,(1) \qquad y$$

$$(1)\;\; nR_0^2=(m^{-1/2}\;i'\hat{m})(\hat{m}'\,\hat{m}/n-\,\hat{m}'\hat{l}_\beta/n(\hat{l}_\beta'\,\hat{l}_\beta/n\,)^{-1}\hat{l}_\beta'\,\hat{m}/n)(n^{-1/2}\,\hat{m}'\,i)+\,o_p(1)$$

Todo lo que necesitamos probar ahora es que la expresión anterior es asintóticamente equivalente a n R_0^2 de la regresión de mínimos cuadrados ordinarios de

$$i \quad en \quad \hat{m},\,\hat{l}_\beta$$

Para la regresión anterior:

$$n\,R_0^2 = i'(\hat{m},\,\hat{l}_\beta)\begin{pmatrix}\hat{m}'\,\hat{m} & \hat{m}'\,\hat{l}_\beta\\ \hat{l}_\beta'\,\hat{m} & \hat{l}_\beta'\,\hat{l}_\beta\end{pmatrix}^{-1}\begin{pmatrix}\hat{m}'\\ \hat{l}_\beta'\end{pmatrix} i =$$

$$= (n^{-1/2}i'\,\hat{m},\;n^{-1/2}\;i'\,\hat{l}_\beta)\begin{pmatrix}\hat{m}'\,\hat{m}/n & \hat{m}'\,\hat{l}_\beta/n\\ \hat{l}_\beta'\,\hat{m}/n & \hat{l}_\beta'\,\hat{l}_\beta/n\end{pmatrix}\begin{pmatrix}n^{-1/2}\,\hat{m}'i\\ n^{-1/2}\,\hat{l}_\beta'i\end{pmatrix}$$

y $n^{1/2}\, i'\,\hat{l}_\beta = o_p(1)$, por el Supuesto S.5 (vi); $\hat{l}_\beta'\,\hat{l}_\beta/n$ es $O_p(1)$ por el Supuesto S.2 y S.5(v), por lo tanto

$$nR_0^2=[n^{-1/2}i'\hat{m}][\hat{m}'\hat{m}/n-\hat{m}'\hat{l}_\beta/n(\hat{l}_\beta'\hat{l}_\beta/n)^{-1}\,\hat{l}_\beta'\,\hat{m}/n]^{-1}[n^{-1/2}\hat{m}'i] + o_p(1)$$

Que es asintóticamente equivalente a la expresión (1).

PRUEBA DEL TEOREMA 2.2: Por los Supuestos S.1 - S.4' y el Teorema 3.3 de White(1985), un test m puede ser computado como $n\,R_0^2$ de la regresión de

$$i \quad \text{sobre} \quad \hat{m},\ \hat{l}_\beta$$

Por los Supuestos B.1 y B.2 la regresión anterior puede escribirse como

$$i \quad \text{sobre} \quad RW,\ RX$$

y nR_0^2 de la regresión anterior es

$$n\,R_0^2 = (i'RW/\sqrt{n},\ i'RX/\sqrt{n}) \begin{pmatrix} W'RRW/n & W'RRX/n \\ X'RRW/n & X'RRX/n \end{pmatrix}^{-1} \begin{pmatrix} W'Ri/\sqrt{n} \\ X'Ri/\sqrt{n} \end{pmatrix}$$

donde $R = R'$, porque es una matriz diagonal. $n^{-1/2}\, i'\, RX = o_p(1)$ por el Supuesto S.4' (vi). $X'RRW/n$ y $X'RRX/n$ son $O_p(1)$ por los Supuestos S.4' y S.2, así pues por una modificación del Corolario 2.36 de White(1984a):

$$n\,R_0^2 = (i'RW/\sqrt{n})\ (W'RRW/n - W'RRX/n(X'RRX/n)^{-1}$$

$$X'RRW/n)^{-1}\ (W'Ri/\sqrt{n}) + o_p(1).$$

Por los Supuestos B.3 (ii), S.4'(vi) y el Corolario 2.36 de White(1984a):

$$n\ R_0^2 = (i'\ R\ W/\sqrt{n})\ (\hat{\sigma}_n^2\ W'W/n - \hat{\sigma}_n^2$$

$$W'X/n(\hat{\sigma}_n^2\ X'\ X/n)^{-1}\ \hat{\sigma}_n^2\ X'\ W/n)^{-1}$$

$$(W'\ R\ i/\sqrt{n}) + o_p(1)$$

$$=(i'R\ W/\sqrt{n})(W'W/n-W'X/n(X'X/n)^{-1}X'W/n)^{-1}(W'R\ i/\sqrt{n})/\hat{\sigma}_n^2+o_p\ (1)$$

que es asintóticamente equivalente al nR^2 de la regresión de mínimos cuadrados ordinarios de

r sobre W, X,

dado que $r = Ri$ y

$$n\ R^2= (i'R\ W/\sqrt{n},\quad i'RX/\sqrt{n})\begin{pmatrix} W'W/n & W'X/n \\ X'W/n & X'X/n \end{pmatrix}^{-1}\begin{pmatrix} W'Ri/\sqrt{n} \\ X'Ri/\sqrt{n} \end{pmatrix}/(i'RRi/n)$$

y $iRX/\sqrt{n} = o_p(1)$ por el Supuesto S.4'(vi) $X'X/n = O_p(1)$ por los Supuestos B.3, S.4' (v), S.2, y $i'RRi/n = n^{-1} \sum_{t=1}^{n} r_t(\omega^t, \hat{\theta}_n) = \hat{\sigma}_n^2$.

$$n\ R^2 = (i'\ R\ W/\sqrt{n})(W'X/n - W'X/n(X'X/n)^{-1}X'\ W/n)^{-1}(W'\ R\ i/\sqrt{n})/\hat{\sigma}_n^2+o_p(1)$$

por una modificación del Corolario 2.36 de White(1984a), lo cual prueba el Teorema.

Nótese que dado que estamos usando una constante como uno de los regresores, la media de los errores estimados es cero y por consiguiente el R^2 ajustado por la constante y el R^2 no ajustado por la constante de la regresión r sobre W, X son numéricamente equivalentes. Ver Theil(1971) ecuaciones (1.4), página 164 y (4.2), página 176.

PRUEBA DEL TEOREMA 2.3:

(i) Para probar esto es suficiente demostrar que

a) $\int h_n(\omega^n, \Theta, \lambda)\, d\mu^n = 1$ y

b) $h_n(\omega^n, \Theta, \lambda) >= 0$

c) $h_n(\omega^n, \Theta, \lambda)_{|\lambda=0} = f_n(\omega^n, \Theta)$

(i) a) $\int h_n(\omega^n; \Theta, \lambda)\, d\,\mu^n =$

$\int \exp[\lambda' m_n(\omega^n; \Theta) - \Psi(\Theta,\lambda)] f_n(\omega^n;\Theta)\, d\,\mu^n =$

$[\int \exp[\lambda' m_n(\omega^n; \Theta)] f_n(\omega^n; \Theta)\, d\,\mu^n]/$

$[\int\{\int \exp[\lambda' m_n(\omega^n; \Theta)] f_n(\omega^n, \Theta)\, d\,\mu^n\} f_n(\omega^n, \Theta)\, d\,\mu^n] =$

$= \int f_n(\omega^n; \Theta)\, d\,\mu^n = 1$

(i) b): Como $f_n(\omega^n; \Theta)$ es una densidad propiamente dicha, $f_n(\omega^n, \Theta) >= 0$.

Además, para todo ω^n, Θ y λ.

$\exp[\lambda' m_n(\omega^n, \Theta) - \Psi(\Theta, \lambda)] >= 0$

Por lo tanto, $h_n(\omega^n, \Theta, \lambda) >= 0$

(i) c) $h_n(\omega^n, \Theta, \lambda)_{|\lambda=0} = \exp[0 - \Psi_n(\Theta, \lambda)_{|\lambda=0}]\, f_n(\omega^n, \Theta)$

$= f_n(\omega^n, \Theta)$, puesto que $\Psi_n(\Theta, \lambda)_{|\lambda=0} =$

$$\log \int \exp [0]\, f_n (\omega^n, \theta)\, d\, \mu^n = 0$$

(ii) $$[\nabla_\theta \ln h_n(\omega^n;\theta,\lambda)]_{|\lambda=0} = \nabla_\theta[\ln h_n(\omega^n;\theta,\lambda)_{|\lambda=0}] = \nabla_\theta \ln f_n(\omega^n;\theta)$$

La primera igualdad anterior se sigue de la definición del vector de derivadas parciales que existe por supuesto. La segunda igualdad se sigue de la definición de $h_n(\omega^n;\, \theta,\, \lambda)$

(iii) $$h_n(\omega^n;\, \theta,\, \lambda) = \exp[\lambda' m_n(\omega^n,\, \theta) - \Psi(\theta,\, \lambda)]\, f_n(\omega^n,\, \theta)$$

$$\ln h_n(\omega^n;\, \theta,\, \lambda) = \lambda' m_n(\omega^n,\, \theta) - \Psi(\theta,\, \lambda) + \ln f_n(\omega^n,\, \theta)$$

$$\nabla_\lambda \ln h_n(\omega^n;\, \theta,\, \lambda) = m_n(\omega^n,\, \theta)' - \nabla_\lambda \Psi(\theta,\, \lambda)$$

$$\nabla_\lambda \ln h_n(\omega^n;\, \theta,\, \lambda)_{|\lambda=0} = m_n(\omega^n,\, \theta)' - \nabla_\lambda \Psi(\theta,\, \lambda)_{|\lambda=0} \quad y$$

$$\nabla_\lambda \Psi(\theta,\, \lambda)_{|\lambda=0} = \nabla_\lambda [\Psi(\theta,\, \lambda)_{|\lambda=0}] = 0$$

Por lo tanto

$$\nabla_\lambda \ln h_n(\omega^n;\, \theta,\, \lambda)_{|\lambda=0} = m_n(\omega^n,\, \theta)'.$$

NOTAS

1. El autor está en deuda con Halbert White por la dirección y el apoyo prestados. Han sido muy útiles varias conversaciones con Robert F. Engle. Alvaro Escribano, Jeffrey Wooldridge, Jeffrey Zabel, Tim Bollerslev y Joel Sobel. La responsabilidad de los errores que todavía queden es del autor.

2. Los resultados de Chesher y Jewitt para el caso en que haya heteroscedasticidad implican que si se usa la matriz de varianzas y covarianzas usual de mínimos cuadrados ordinarios, (V), en vez de HC1 o bien HC2, el sesgo proporcional de V puede ser positivo o negativo. Los sesgos de usar HC0 o HC2 serán en general más pequeños que los de usar V.

REFERENCIAS

Belsley, Kuh y Welsch (1980), Regression Diagnostics. New York: Wiley.

Breusch, T.S. (1978), "Testing for Autocorrelation in Dynamic Linear Models," Australian Economic Papers, 17, pp. 334-355.

Cartwright, P., (1985), "Forecasting Time Series: A comparative Analysis of Alternative Classes of Time Series Models," Journal of Time Series Analysis, Vol. 6, pp. 202-212.

Cavanagh, C., (1985), "Second Order Admissibility of Likelihood Based Tests," Harvard Institute of Economic Research, Discussion Paper 1148.

Chesher y Jewitt, (1984), "Finite Sample Properties of Least Squares Covariance Matrix Estimators," manuscrito.

Cook y Weisberg, (1982), Residuals and Influence in Regression. New York: Chapman and Hall.

Davidson y MacKinnon, (1985), "Heteroskedasticity-Robust Tests in Regression Directions," Working Paper, Department of Economics, Queen's University

Domowitz y Hakkio, (1984), "Testing for Serial Correlation and Common Factor Dynamics in the Presence of Heteroskedasticity," Manuscrito, Northwestern University.

Durbin, J., (1970), "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica, 38, pp. 410-421.

Eicker, F., (1963), "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," Annals of Mathematical Statistics, 34, pp. 447-456.

Efron, B., (1982), The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

Engle, R.F., (1982a), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica, 50, pp. 987-1007.

Engle, R.F., (1982b), "A General Approach to Lagrange Multiplier Model Diagnostics," Journal of Econometrics, 20, pp. 83-104.

Engle, R.F., (1984), "Wald, Likelihood Ratio and Lagrange Multipliers Tests in Econometrics" en: Z. Griliches y M. Intriligator, eds., Handbook of Econometrics: North Holland, Amsterdam.

Engle y Bollerslev, (1986), "Modelling the Persistence of Conditional Variances," manuscrito, University of California, San Diego.

Engle, Lilien y Robins, (1985), "Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model," Econometrica 55, nº 2, p. 391-408.

Godfrey, (1978), "Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables," Econometrica, 46, pp. 1293-1302.

Gourieroux, Monfort, Renault y Trognon, (1984), "Residus generalisés ou interpretations Linéaires de l'Econométrie non linéaire," INSEE Documents de Travail, No. 8410.

Granger y Andersen, (1978), An Introduction to Bilinear Time Series Models. Vandenhoech & Ruprecht.

Hall, Alastair (1985), "Heteroskedasticity Robust Lagrange Multiplier Tests for Serial Correlation and a Simplified Method of their Calculation," Mimeo.

Horn, Horn y Duncan, (1975), "Estimating Heteroskedastic Variances in Linear Models," Journal of the American Statistical Association, 70, pp. 380-385.

Kiviet, J., (1981), "On the Rigour of Some Misspecification Tests for Modelling Dynamic Relationships," Manuscrito, University of Amsterdam.

MacKinnon y White, (1985), "Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties," Journal of Econometrics, Vol. 29, No. 3, pp. 305-326.

Newey, W., (1985), "Maximum Likelihood Specification Testing and Conditional Moment Tests," Econometrica, 53, pp. 1047-1070.

Priestley, M., (1980), "State-Dependent Models: A General Approach to Non-linear Time Series Analysis," Journal of Time Series Analysis, Vol. 1, pp. 47-72.

Rao, Subba, (1986), "Statistical Analysis of Bivariate Bilinear Time Series Models," University of Manchester Institute of Science and Technology, Technical Report 177.

Tauchen, G., (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," Journal of Econometrics, 30, 1/2, pp. 415-444.

Theil, H., (1971). Principles of Econometrics. New York: Wiley.

White, H., (1980), "A Heteroskedasticity-consistent Covariance Matrix and a Direct Test for Heteroskedasticity," Econometrica, 48, pp. 817–838.

White, H., (1984a), Asymptotic Theory for Econometricians. New York: Academic Press.

White, H., (1984b), "Comments to tests of Specification in Econometrics," Econometric Reviews, 3, pp. 261–268.

White, H., (1985), "Specification Testing in Dynamic Models," UCSD discussion paper 85-30R.