Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE
FACULTAD DE ECONOMICAS
Campus de Somosaguas
28223 MADRID
Teléfono 3942611 - FAX 3942613

Documento de Trabajo

Contrastes de momentos y de la matriz de información

Teodosio Pérez Amaral

No. 9401

Junio 1994

Instituto Complutense de Análisis Económico
UNIVERSIDAD COMPLUTENSE

CONTRASTES DE MOMENTOS Y DE LA MATRIZ DE INFORMACION

Teodosio Pérez Amaral
Instituto Complutense de Análisis Económico
Universidad Complutense
Campus de Somosaguas
28223 Madrid

RESUMEN

En este trabajo se presentan los principales resultados de la literatura reciente sobre contrastes de momentos (m) y de la matriz de información. Los contrastes de momentos son un marco general para obtener diagnósticos de la especificación de modelos estimados bien por máxima verosimilitud o por el método de momentos. Los contrastes m pueden considerarse, bajo condiciones generales, contrastes de los multiplicadores de lagrange.

Una fuente de condiciones de momentos en los que basar la construcción de los diagnósticos m es la igualdad de la matriz de información. Se ilustra cómo, en el caso de regresión lineal, los contrastes basados en la igualdad de la matriz de información generan diagnósticos tanto conocidos como más novedosos. El comportamiento en muestras finitas de los contrastes es una consideración importante a la hora de su utilización, debiendo elegirse en cada circunstancia la versión más apropiada.

Finalmente se señala el gran potencial de la igualdad de la matriz de información para generar baterías de diagnósticos para modelos para los cuales se dispone actualmente de una menor variedad de diagnósticos que para el caso de regresión lineal.

ABSTRACT

This paper presents the main results of the recent literature on moment (m) tests and information matrix tests. The m tests provide a general framework for deriving specification diagnostics for models estimated by maximum likelihood or the method of moments. Under general conditions, m tests can be considered LM tests.

A source of moment conditions in which one could base the construction of m tests is the information matrix equality. It is shown that in the case of linear regression, tests based on the information matrix equality generate a variety of diagnostics, some of them new while others are already familiar. Small sample considerations are important for the practical application of the tests; it is suggested that appropriate choice of the particular test will be an important practical decision.

One of the most important implications of this framework is that it can generate batteries of diagnostics for models for which few diagnostics are presently available.

1. Introducción

En los últimos años ha habido un gran interés por el estudio de las consecuencias de los errores de especificación sobre la estimación e inferencia realizada con modelos econométricos. Por ello se han desarrollado gran variedad de métodos estadísticos para detectar la presencia de errores de especificación. En general, resulta apropiado pensar en estos procedimientos como diagnósticos que intentan detectar síntomas que revelen la posible existencia de mala especificación.

Con frecuencia, el que un modelo esté correctamente especificado tiene una serie, a menudo infinita, de implicaciones (condiciones necesarias de correcta especificación). Una estrategia fructífera para obtener diagnósticos consiste en comprobar si alguna o algunas de esas implicaciones (condiciones necesarias) se cumplen en la muestra considerada.

Un ejemplo sería un modelo cuya especificación correcta implicase ausencia de correlación serial en los errores (u_t). En ese caso tendríamos un conjunto infinito de implicaciones (condiciones necesarias) $E(u_t u_{t-j}) = 0$ para $j = 1, 2, \ldots$ en las que basar contrastes de especificación. Esta sería la base de los contrastes de autocorrelación.

Otro ejemplo sería un modelo para el cual hubiese más de un estimador consistente de los parámetros de interés, (p.e.: mínimos cuadrados y variables instrumentales). Bajo especificación correcta los dos estimadores deberían tener valores asintóticamente similares. Si los dos estimadores tienen valores estadísticamente diferentes entonces el modelo estaría mal especificado. Este sería el principio de los contrastes de Hausman (1978).

Otro ejemplo sería una estimación por máxima verosimilitud, en la que la especificación correcta implica la validez de la igualdad de la matriz de información, esto es, que la matriz de información se puede estimar tanto por el negativo del hesiano como por la forma de producto exterior del gradiente. Si estos dos estimadores no están suficientemente próximos entre sí,

entonces tendremos evidencia en contra de la hipótesis de especificación correcta. Esta es la base de los contrastes de la matriz de información de White (1982, 1987 y 1994).

Hay una gran variedad de procedimientos para derivar diagnósticos de especificación basados en contrastar la validez de alguna(s) implicación(es) (condiciones necesarias) de especificación correcta. En este trabajo nos vamos a centrar en el estudio de un marco concreto de análisis de especificación: los contrastes de momentos (m), propuestos independientemente por Newey (1985) y Tauchen (1985), y desarrollados posteriormente por White (1987 y 1994) y Pérez (1993a y c).

Muchos diagnósticos previamente disponibles pueden ser considerados como casos particulares de este marco general de los constrastes de momentos (m), por ejemplo los contrastes de la matriz de información de White. En este trabajo se expondrán de manera sintética los principales resultados de esta literatura.

El resto del trabajo está organizado como sigue: en la sección 2 se explicitan y se comentan los resultados generales sobre contrastes m. En la 3 se obtiene la relación entre contrastes m y otros contrastes más conocidos, se justifica que los contrastes m pueden ser considerados contrastes de los multiplicadores de lagrange de alternativas paramétricas interpretables, y se detalla la construcción de contrastes de la razón de las verosimilitudes y de Wald equivalentes a un contraste m dado. En la 4 se presentan los contrastes de la matriz de información y en la 5 se particularizan los contrastes de la matriz de información de primer orden (dinámicos) al caso de regresión lineal. En la 6 se considera el comportamiento en muestras finitas de los contrastes m y en la 7 se encuentran las conclusiones.

2. Contrastes de momentos

Newey (1985) y Tauchen (1985) han propuesto un marco general para la detección de los errores de especificación en los modelos econométricos basado en el hecho de que un modelo dado, **P**, con un vector de parámetros asociado $\theta \in \Theta \subset R^k$ generalmente implica la validez de un número, quizá infinito, de condiciones de momentos de la forma

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

donde $m_t : \Omega \times \Theta \rightarrow R^p$, *con* $\omega \in \Omega$.

Si para todo θ en Θ

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

resultaría que $\mathbf{P}_0$, el mecanismo que genera los datos, con vector de parámetros asociado θ_0 no está en **P**, esto es, que el modelo no está bien especificado. Esto sugeriría que se puede basar un contraste de especificación en las medias muestrales[1]

$$\hat{m}_n \equiv n^{-1} \sum_{t=1}^{n} \hat{m}_t \equiv n^{-1} \sum_{t=1}^{n} m_t(\omega, \hat{\theta}_n) \tag{3}$$

donde $\hat{\theta}_n$ sería un estimador consistente de θ_0 cuando $P_0 \in P$ y n sería el tamaño de la muestra. Estas medias muestrales (3) estarían próximas a cero cuando P estuviese bien especificado, pero de lo contrario, serían distintas de cero. Un ejemplo sería el caso en que la

[1] La notación sub t sugiere dependencia de la información en el momento t; sub n sugiere dependencia de la información de toda la muestra. El superíndice t sugiere dependencia de la información de la muestra hasta el momento t. $\hat{m}_t$ va a depender, en general, de toda la muestra, debido a que depende de $\hat{\theta}_n$; para simplificar la notación, lo denotaremos simplemente $\hat{m}_t$ en adelante.

validez del modelo implicase que los errores no estuviesen autocorrelacionados. En ese caso, bajo especificación correcta

$$E_0(u_t u_{t-j}) = 0 \qquad \begin{matrix} t=1,\ 2,\ \ldots \\ j=1,\ 2,\ \ldots \end{matrix} \tag{4}$$

Esto sugiere que podríamos basar un contraste de momentos en las medias de los análogos muestrales de los errores, los residuos estimados

$$n^{-1} \sum_{t=1}^{n} r_t r_{t-j} \qquad j=1,\ 2,\ \ldots\ . \tag{5}$$

que es precisamente un ejemplo típico de contrastes de momentos.

Cuando el modelo está correctamente especificado, se puede demostrar que, bajo condiciones generales, $\sqrt{n}\ \hat{m}_n$ se distribuyen asintóticamente como normal multivariante, y en el caso de que se dispusiese de un estimador consistente de su matriz de varianzas y covarianzas, por ejemplo $\hat{V}_n$, se podría formar un estadístico χ^2 de la siguiente forma:

$$\mathrm{M}_n \equiv n\ \hat{m}_n{}'\ \hat{V}_n^{-1}\ \hat{m}_n \sim \chi^2_p \tag{6}$$

Donde p es el número de condiciones de momentos contenidas en $\hat{m}_n$. Este enfoque es bastante general e incluye como casos particulares los contrastes de los multiplicadores de lagrange (Rao(1947), Aitchison y Silvey(1958)), de Hausman (1978), de Cox (1961, 1962) y de la matriz de información de White (1982, 1987, 1994). Véase Godfrey (1988) para una panorámica de gran parte de esta literatura. Hay que señalar que este enfoque de condiciones de momentos es aplicable a modelos estimados tanto por máxima verosimilitud como por el método de momentos.

La expresión (6) requiere un conjunto de condiciones de regularidad habituales en esta literatura, que se encuentran en White (1994). El contexto de máxima verosimilitud es especialmente conveniente para presentar los resultados sobre contrastes de momentos, y por eso lo utilizaremos en el resto de este trabajo. También nos limitaremos a una situación en la que las variables consideradas van a ser no explosivas (quizá después de alguna transformación) aunque hay evidencia de que los resultados serían aplicables a ciertos tipos de variables no estacionarias; véase Wooldridge (1986).

En la aplicación de la expresión (6) surgen dos cuestiones importantes: la primera es la elección de $\hat{m}_n$, de la que nos ocuparemos más adelante en la sección 4, y la segunda la elección de un estimador consistente de la varianza: $\hat{V}_n$.

Bajo condiciones generales se podría elegir el estimador

$$\hat{V}_n \equiv n^{-1} \sum_{t=1}^{n} \hat{m}_t \, \hat{m}_t' \tag{7}$$

que sería consistente si los elementos de

$$V_n \equiv E\,[n^{-1} \sum_{t=1}^{n} m_t \, m_t'] \tag{8}$$

son finitos y continuos y obedecen una ley débil de los grandes números, uniformemente. En esas condiciones el estadístico (6) se podría computar como nR^2 donde R^2 es el R^2 no centrado (no ajustado por el uso de una constante), de Theil[2] (1971, p. 164) de una regresión artificial donde la variable dependiente es la unidad y los regresores son las condiciones de momentos $\hat{m}_t$ (los indicadores estimados).

[2] En un modelo tal como $y = X b + e$, se define $R^2 = b'X'Xb / y'y$.

$$1 \;\; sobre \;\; \hat{m}_t' \tag{9}$$

(Véase White (1994), corolario 9.9). Cuando $\mathbf{P}_0 \ni \mathbf{P}$, entonces $M_n \equiv n\,R^2 \sim \chi^2_p$ de forma que un contraste m del tamaño α se llevaría a cabo rechazando H_0: $\mathbf{P}_0 \ni \mathbf{P}$ cuando nR^2 fuese mayor que el valor crítico de la $\chi^2_{p,\,1-\alpha}$.

Si se desea usar un estimador $\hat{V}_n$ que sea consistente cuando $\hat{\theta}_n$ sea asintóticamente eficiente pero no lo sea necesariamente cuando $\hat{\theta}_n$ no sea eficiente, entonces se podría utilizar como estimador de la matriz de varianzas y covarianzas

$$\hat{J}_n = n^{-1}\sum_{t=1}^{n} \hat{m}_t \, \hat{m}_t' - (n^{-1} \sum_{t=1}^{n} \hat{m}_t \hat{s}_t')[n^{-1}\sum_{t=1}^{n} \hat{s}_t \hat{s}_t']^{-1}(n^{-1}\sum_{t=1}^{n} \hat{s}_t \hat{m}_t') \tag{10}$$

donde el segundo término después de la igualdad está teniendo en cuenta que los $\hat{\theta}_n$ que se emplean para evaluar $\hat{m}_t$ han sido estimados, siendo su matriz de varianzas y covarianzas $[n^{-1} \sum_{t=1}^{n} \hat{s}_t \, \hat{s}_t']^{-1}$, donde $\hat{s}_t$ es el gradiente del logaritmo de la verosimilitud para la observación t evaluado en el estimador de máxima verosimilitud.

Usar este estimador de varianza en (6) equivaldría a utilizar el estadístico $M_n' = nR^2$, donde R^2 es el R^2 no ajustado por el uso de una constante de la regresión auxiliar en la que la variable dependiente es la constante unidad y las variables explicativas son $\hat{m}_t'$ y $\hat{s}_t'$, esto es,

$$1 \;\; sobre \;\; \hat{m}_t', \; \hat{s}_t' \qquad t = 1,\, 2,\, \ldots,\, n. \tag{11}$$

En este caso se incluye en la regresión $\hat{s}_t$ para tomar en cuenta que los $\hat{m}_t$ han sido estimados. La anterior regresión auxiliar, como se ve, no es más que un artificio para computar de forma simplificada el estadístico (6) usando (10) como estimador de la matriz de varianzas y covarianzas. Este estadístico fue propuesto por Newey(1985) para el caso de observaciones independientes e idénticamente distribuidas y extendido al caso de observaciones dependientes por White (1987). La interpretación intuitiva de la regresión auxiliar (11) sería la siguiente: la regresión intenta "explicar" la variable dependiente, que es una constante; sólo lo podrá hacer si alguno de los regresores tiene media distinta de cero. Cada uno de los componentes del gradiente $\hat{s}_t^{\prime}$ tiene media cero puesto que el procedimiento de estimación (máxima verosimilitud) lo garantiza; así, lo único que podría explicar la constante es que alguna/s de las condiciones de momentos $\hat{m}_t^{\prime}$ tengan media distinta de cero, que es precisamente lo que intentamos comprobar con el estadístico.

La inferencia basada en (11) tendrá un tamaño asintótico apropiado cuando $\hat{\theta}_n$ sea un estimador asintóticamente eficiente, debido a que se usa la matriz de varianzas y covarianzas de $\hat{m}_t$ que resultaría en el caso de que $\hat{\theta}_n$ fuese asintóticamente eficiente (producto exterior del gradiente). El contraste anterior utiliza como estimador de la matriz de varianzas y covarianzas de $\hat{\theta}_n$ la inversa de $\hat{B}_n = n^{-1}\sum_{t=1}^{n} \hat{s}_t \hat{s}_t^{\prime}$ que es la forma de producto exterior de la matriz de información. El comportamiento en muestras finitas de este estimador puede ser deficiente, lo que, en ocasiones nos llevaría a querer modificar el contraste en alguna de las direcciones que se sugieren más adelante.

La importancia del resultado (11) radica en que permite llevar a cabo contrastes en la dirección elegida por los indicadores $\hat{m}_t$ para cualquier verosimilitud, con el único requisito de disponer del gradiente $\hat{s}_t$, que en general está disponible cuando se ha estimado por máxima verosimilitud.

Además, si podemos descomponer los indicadores $\hat{m}_t$ y el gradiente $\hat{s}_t$ como producto de residuos generalizados (ν_t) por funciones de variables predeterminadas, (j_t, w_t), esto es, si podemos hacer $m_t = j_t \cdot \nu_t$ y $s_t = w_t \cdot \nu_t$, entonces bajo homoscedasticidad condicional (quizá después de estandarizar ν_t), los contrastes m se podrían computar utilizando la regresión auxiliar de los residuos generalizados $\hat{\nu}_t$ sobre la variables explicativas $\hat{j}_t^{\prime}$ y $\hat{w}_t^{\prime}$, esto es:

$$\hat{\nu}_t \ \ sobre \ \ \hat{j}_t^{\prime}, \ \hat{w}_t^{\prime} \tag{12}$$

y $M_n^{\prime\prime} = nR^2$, de la regresión auxiliar anterior. Véase una prueba simplificada en el Apéndice 1. Expresado en la forma de (12), el contraste m resultaría idéntico al contraste de los multiplicadores de lagrange de variables omitidas $\hat{j}_t^{\prime}$ en el que ya se han incluido como regresores las variables $\hat{w}_t^{\prime}$.

Los resultados (9), (11) y (12) usan progresivamente más estructura del modelo para conseguir una computación más sencilla y un mejor comportamiento en muestras finitas. Los estadísticos (9) y (11) pueden requerir muestras especialmente grandes para tener una distribución que se aproxime a su distribución asintótica. Además, Taylor (1987), los resultados analíticos de Cavanagh(1985) y las simulaciones de Chesher y Spady (1991) implican que los estadísticos (9) y (11) pueden tener una distribución que se aproxima mal por una χ_p^2. Esto sugiere utilizar algunas modificaciones de los estadísticos (9) y (11) como la de Wooldridge (1990), que explota la existencia de residuos generalizados y consiste en depurar los elementos de j_t de los componentes correlacionados con ν_t, para obtener un mejor comportamiento de los contrastes en muestras finitas. Este punto se retoma en la sección 6.

3. La relación de los contrastes m con otros contrastes

La similitud de los contrastes m con los contrastes de multiplicadores de lagrange (LM) de variables omitidas sugerida por la expresión (12) resulta acertada. Pérez (1993a) y White (1994) explicitan las condiciones de regularidad bajo las cuales esencialmente cualquier contraste m puede ser considerado un contraste LM. Se demuestra por construcción que los contrastes m son contrastes LM de una alternativa paramétrica en la que las condiciones m podrían ser consideradas como una parte del gradiente de una verosimilitud artificial, de la cual la verosimilitud original sería un caso particular. La verosimilitud artificial se construye para conseguir que los contrastes m se puedan derivar directamente como contrastes de los multiplicadores de Lagrange. Para ello, es necesario construir, partiendo de una función de verosimilitud original $f_n(\omega^t, \theta)$, una verosimilitud artificial o ampliada, $h_n(\omega^n, \theta, \lambda)$, que cumple las siguientes condiciones:

1. La función de verosimilitud original $f_n(\omega^t, \theta)$ es un caso particular de la verosimilitud artificial cuando las condiciones m son cero.
2. La integral de la función de verosimilitud artificial es uno.
3. Las condiciones m se obtienen como parte del gradiente de la función de verosimilitud artificial.

El considerar a los contrastes m como contrastes LM nos permite interpretarlos dentro del marco de los contrastes de hipótesis clásicos, y de esta forma, gozarían de las propiedades de optimalidad de los contrastes LM. La construcción de la verosimilitud artificial podría llevarse a cabo como en Pérez (1989, 1993a), utilizando la siguiente verosimilitud artificial.

Definición de verosimilitud artificial

Si partimos de una familia generatriz de verosimilitudes $f_n(\omega^t, \theta)$ y un vector de dimensión p x 1 de condiciones de momentos $m_{nt}(\omega^t, \theta)$ tal que bajo especificación correcta ($P_0 \ni P$) se cumple que

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

definimos la función de verosimilitud artificial $h_n(\omega^n, \theta, \lambda)$ donde $\lambda \in \Lambda \subset R^p$ y Λ es 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 θ en Θ y todo n en N. Además, para todo θ, λ en $\Theta \times \Lambda$ definimos la verosimilitud artificial

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

donde

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

El papel que cumple el primer término dentro del corchete es introducir las condiciones m dentro de la verosimilitud y conseguir que al diferenciar con respecto a sus parámetros, las condiciones m resulten ser parte del gradiente. El papel del segundo término dentro del corchete es normalizar para que la integral de la verosimilitud artificial sea uno.

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

Esta verosimilitud artificial cumple los requisitos para ser una función de verosimilitud propiamente dicha, y con su ayuda los contrastes m pueden ser considerados contrastes LM de las restricciones implicadas en la función de verosimilitud artificial por la hipótesis nula de que $f_n(\omega, \theta)$ es la verosimilitud correcta en vez de $h_n(\omega, \theta, \lambda)$. Estas funciones de verosimilitud artificiales tendrán mayor o menor interpretabilidad dependiendo de cuál sea la verosimilitud original y las condiciones de momentos que se elijan. Si la verosimilitud de partida pertenece a la familia lineal exponencial, entonces la verosimilitud artificial pertenecerá a la misma familia. En el caso de que se parta de una verosimilitud gaussiana (normal), en general, es posible construir para cada conjunto de condiciones de momentos, una verosimilitud artificial también gaussiana de la cual la verosimilitud original sea un caso particular.

Si se dan las condiciones para poder construir la verosimilitud artificial, los contrastes m serían contrastes LM de variables omitidas en la verosimilitud artificial $h_n(\omega, \theta, \lambda)$. Debido a ello, se plantea de forma inmediata la existencia de contrastes de Wald equivalentes a contrastes m dados. Dadas un conjunto de condiciones m en las que basar un contraste de momentos, para llevar a cabo el contraste de Wald equivalente bastarían los siguientes pasos:

1. Partiendo de la función de verosimilitud considerada, $f_n(\omega, \theta)$ construir una función de verosimilitud artificial $h_n(\omega, \theta, \lambda)$, que en general no será única.

2. Estimar por máxima verosimilitud el modelo implicado por la función de verosimilitud artificial $h_n(\omega, \theta, \lambda)$ elegida. De entre las funciones de verosimilitud artificiales posibles, el investigador elegirá la que le parezca más plausible desde un punto de vista teórico o estadístico.

3. El contraste de Wald consistiría en contrastar, por el método apropiado en cada caso, la significatividad conjunta del vector de parámetros λ en la anterior estimación. Si se rechaza la no significatividad conjunta de los parámetros λ, entonces tendremos evidencia en contra de la verosimilitud $f_n(\omega, \theta)$. Porque al ser significativos los parámetros λ, las condiciones de momentos tomarían valores distintos de cero, y la verosimilitud artificial no se reduciría al caso particular $f_n(\omega, \theta)$.

De forma análoga, un contraste de la razón de las verosimilitudes (LR) correspondiente a un contraste m dado se llevaría a cabo de la siguiente forma:

1. Estimar el modelo con la verosimilitud $f_n(\omega, \theta)$ y obtener el valor del logaritmo de función de verosimilitud, digamos l_f.

2. Estimar el modelo con la verosimilitud artificial $h_n(\omega, \theta, \lambda)$ y obtener el valor del logaritmo de la función de verosimilitud, digamos l_h.

3. Formar el estadístico del cociente de las verosimilitudes $LR_n = -2(l_f - l_h)$, que se distribuirá como una χ^2_p si el modelo $f_n(\omega, \theta)$ es correcto. Si $LR_n > \chi^2_p$, entonces se rechazará la hipótesis nula de especificación correcta, y tendremos evidencia en contra de la verosimilitud $f_n(\omega, \theta)$.

4. Un caso particular de los contrastes m: los contrastes de la matriz de información.

En las secciones anteriores hemos visto cómo se pueden computar e interpretar los contrastes m bajo circunstancias diversas. Sin embargo, quedaría por resolver la importante pregunta de cómo obtener condiciones de momentos útiles, esto es, implicadas por el supuesto de especificación correcta de $f_n(\omega, \theta)$ y, además informativas, en cuanto a posibles errores de especificación. Necesitaríamos, pues, una fuente de condiciones de momentos relevantes.

Este problema tendría una importancia menor en los casos en que usemos verosimilitudes $f_n(\omega, \theta)$ familiares como la gaussiana, pues en ese caso hay una serie de diagnósticos conocidos que se podrían aplicar. Sin embargo el problema se acentuaría cuando utilicemos

verosimilitudes menos frecuentes en la práctica (p. ej.: la poisson) para las cuales no hay una batería de contrastes estándar.

White (1982, 1987, 1994) ha introducido un marco general para derivar diagnósticos en el caso de estimación por máxima verosimilitud. Pérez (1989, 1993a) lo ha particularizado para el caso de regresión lineal. Este marco parte de la igualdad de la matriz de información, esto es, que bajo especificación correcta, ($P_0 \ni P$), hay dos formas equivalentes de la matriz de información de Fisher (véase el Apéndice 2), por el negativo del hessiano y por el producto exterior del gradiente, esto es:

$$A_n = -E(\frac{\partial s}{\partial \theta}) = -E(\frac{\partial^2 \log f(\omega, \theta)}{\partial \theta\, \partial \theta'}) \quad (hesiano)$$

$$B_n \equiv s\, s' = var\ (\ n^{-1/2}\ \nabla\ \log f(\omega, \theta)) \quad (producto\ exterior) \tag{13}$$

Sustituyendo θ por un estimador asintóticamente eficiente, $\hat{\theta}_n$ y las esperanzas por medias, tendríamos dos estimadores distintos de la matriz de información

$$\hat{A}_n = n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \log f_t(\omega, \hat{\theta}_n)}{\partial \theta\, \partial \theta'}$$

$$\hat{B}_n = n^{-1} \sum_{t=1}^{n} \frac{\partial \log f_t(\omega, \hat{\theta}_n)}{\partial \theta} \frac{\partial \log f_t(\omega, \hat{\theta}_n)'}{\partial \theta} \tag{14}$$

Así, diremos que la igualdad de la matriz de información se cumple para una especificación dada si $\hat{A}_n + \hat{B}_n = 0$ o bien $\hat{A}_n + \hat{B}_n \rightarrow 0$. Se podría basar un conjunto de contrastes m en las diferencias entre esas dos formas de estimar la matriz de información:

$$\hat{m}_t = vec\ (\hat{A}_n + \hat{B}_n)$$

donde vec es el operador que transforma una matriz en un vector columna colocando las columnas de la matriz original una a continuación de la otra.

Las especificaciones para las cuales la igualdad de la matriz de información no se cumple no consiguen, en general, la eficiencia asintótica, y ello indicaría que la precisión de la estimación y la potencia de los contrastes podrían mejorarse por el uso de otro estimador; por ejemplo, si se detecta autocorrelación en los errores, quizá habría que utilizar mínimos cuadrados generalizados. El incumplimiento de la igualdad de la matriz de información puede también indicar un error de especificación en $f_n(\omega, \theta)$; por ejemplo, la existencia de una variable omitida. Este incumplimiento sugiere la utilidad de continuar la investigación de la posible mala especificación de $f_n(\omega, \theta)$ usando contrastes direccionales; por ejemplo, un contraste LM de variable omitida.

4.1 Contrastes de la matriz de información de primer orden (dinámica)

Hay un conjunto de contrastes de la matriz de información que reciben el nombre de contrastes de la matriz de información de primer orden debido a que están basados en las primeras derivadas del logaritmo de la verosimilitud f_t, esto es, el gradiente, s_t. Una de las implicaciones de la igualdad de la matriz de información (véase el Apéndice 2) es que

$$var\,(n^{-1/2} \sum_{t=1}^{n} s_t^{*}) = n^{-1} \sum_{t=1}^{n} var\; s_t^{*} + o(1) \qquad (15)$$

donde la estrella sugiere que se evalúan en el óptimo verdadero de la función de verosimilitud elegida. Condiciones suficientes, pero no necesarias para esto son que $E(s_t^{*}) = 0$ y que $E(s_t^{*}\; s_{t-j}^{*}) = 0,\; t = 1, 2, \ldots;\; j = 1, , 2, \ldots, t.$ Esta segunda condición, la incorrelación del gradiente, es una condición útil de examinar, y nos centraremos, por simplicidad, en el análisis de si un conjunto finito de elementos de $E(s_t^{*}\; s_{t-j}^{*})$ son cero. De forma que basaremos los contrastes de especificación en los siguientes indicadores de la matriz de información dinámica:

$$m_t(\omega, \theta) = S\; vec\; s_t(\omega, \theta)\; [s_{t-1}(\omega^{t-1}, \theta)', \ldots, s_{t-\lambda}(\omega^{t-\lambda}, \theta)'] \qquad (16)$$

donde, para mantener la notación original, $\lambda \geq 1$ es ahora un entero que determina el máximo orden de correlación del gradiente que se va a analizar. S es una matriz de selección (u otra matriz no estocástica) de orden $q \times \lambda\, p^2$ que se usa para fijarnos en elementos específicos, o combinaciones lineales de elementos, de la matriz de indicadores.

La computación de estos contrastes se haría aplicando los resultados (9), (11) y (12) bajo las condiciones de regularidad apropiadas. Por ejemplo, un análogo del resultado (11) sería aquí

$$1 \;\; sobre \;\; \hat{m}_t', \; \hat{s}_t' \qquad (17)$$

donde $\hat{m}_t = S\; vec\, \hat{s}_t\, [\hat{s}_{t-1}', \ldots, \hat{s}_{t-\lambda}']$. El contraste de la matriz de información dinámica sería aquí $M_n^{III} = nR^2$ de la anterior regresión auxiliar y bajo especificación correcta M_n^{III} se distribuiría asintóticamente como una χ_p^2 donde p es el número de condiciones de momentos que utiliza el contraste.

Aquí el énfasis de la detección de errores de especificación se halla en la detección de correlación en el gradiente. El hallazgo de correlación en el gradiente puede suponer un error en la especificación dinámica o bien un error más fundamenteal (como la existencia de variables omitidas) en $f_n(\omega, \theta)$. Es útil pensar en los contrastes de la matriz de información dinámica como una generalización de los contrastes de autocorrelación.

4.2 Contrastes de la matriz de información de segundo orden.

Los contrastes de la matriz de información de segundo orden reciben ese nombre porque están basados en la segunda derivada del logaritmo de la verosimilitud. Otra de las implicaciones de la igualdad de la matriz de información (véase el Apéndice 2) es que

$$n^{-1} \sum_{t=1}^{n} E\,(\nabla s_t^*) + n^{-1} \sum_{t=1}^{n} E(s_t^* \; s_t^{*\prime}) = 0 \qquad (18)$$

Así, bajo especificación correcta, la matriz de información podría ser estimada por el negativo del hesiano o por el producto exterior del gradiente. En este caso, podemos considerar contrastes m basados en los indicadores

$$m_t(x^t, \theta) = S \; vec \; [\nabla s_t(x^t, \theta) + s_t(x^t, \theta)\; s_t(x^t, \theta)'] \qquad (19)$$

donde S es una matriz fija, de dimensión $q \times p^2$. Este contraste es sensible ante heteroscedasticidad condicional, y asimetría y curtosis condicionales, aunque puede no ser sensible a mala especificación dinámica. Estos contrastes se presentan en White (1982) y son más conocidos que los de la matriz de información dinámica, en los que ponemos énfasis en este trabajo. Su computación, bajo condiciones de regularidad apropiadas, es similar a la de los contrastes de la matriz de información dinámica de (17).

En este epígrafe hemos visto que una fuente potencialmente útil para obtener contrastes de momentos con los que analizar la especificación de cualquier modelo estimado por máxima verosimilitud es la igualdad de la matriz de información, tanto de primer orden, como de segundo orden. En el epígrafe siguiente veremos cómo en el caso particular de regresión lineal los contrastes de la matriz de información de primer orden dan lugar a contrastes conocidos, así como a otros nuevos que pueden ser interpretados como contrastes de especificación dinámica.

5. Contrastes de la matriz de información de primer orden (dinámicos) en el caso de regresión lineal.

En este epígrafe presentamos los contrastes de la matriz de información de primer orden (dinámicos) para el caso especial, pero particularmente relevante e intuitivo, de regresión lineal. Para simplificar la presentación usaremos como ejemplo la siguiente densidad

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

donde F_{t-1} representa la información contenida en las variables predeterminadas. El logaritmo de la verosimilitud es

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

siendo $\theta \equiv (\alpha, \rho, \beta, \sigma)$; definimos $u_t(\theta) \equiv y_t - \alpha - \rho\, y_{t-1} - \beta\, x_t$ y los residuos

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

donde $\hat{\theta}_n$ es el estimador de máxima verosimilitud de θ, (mínimos cuadrados ordinarios). En este caso, el gradiente condicional de la observación t es:

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

y el gradiente condicional con j retrasos sería la expresión anterior con t-j en vez de t. Si nos centramos en la autocorrelación del gradiente de orden 1, tendremos la siguiente matriz de indicadores, sacando factor común, omitiendo la dependencia entre u y θ y definiendo

$$\eta_t \equiv (u_t^2 - \sigma^2)$$

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

El valor esperado de cada uno de los elementos de esta matriz es cero bajo especificación correcta. Asimismo, el valor esperado de cada elemento de las matrices correspondientes a retardos del gradiente superiores, R_j^0, $j = 2, \ldots, \lambda$, sería cero. Si basamos un contraste m en el elemento 1,1 de R_1^0 tenemos un contraste de correlación serial de orden 1, que se puede computar como nR^2 (no centrado) de la regresión auxiliar

$$1 \quad sobre \quad r_t r_{t-1},\ r_t,\ r_t y_{t-1},\ r_t x_t,\ (r_t^2 - \hat{\sigma}^2) \tag{20}$$

esta regresión sería análoga a la del resultado (11). Aquí se han eliminado las constantes σ^2 sin que por ello cambie el R^2 de la regresión. Por la diagonalidad por bloques de la matriz de información podemos eliminar el último término de la anterior regresión.

Debido a que ambos, el gradiente y el indicador, se pueden escribir como producto de residuos por variables predeterminadas, podríamos aplicar el resultado (12) y computar el contraste de autocorrelación de orden 1 como nR^2 de la regresión auxiliar

$$r_t \quad sobre \quad r_{t-1},\ 1,\ y_{t-1},\ x_t \tag{21}$$

idéntica a la que se utilizaría para computar el contraste LM de autocorrelación de orden 1. El resto de los elementos de la submatriz 3 x 3 superior izquierda darían lugar a contrastes de especificación dinámica y posibles no linealidades omitidas. Los tres primeros elementos de la última columna darían lugar a contrastes de ARCH en media y los de la última fila darían lugar a contrastes sensibles a ARCH tanto no simétricos como simétricos.

Eligiendo los correspondientes elementos de las matrices de indicadores R_j^0 se pueden construir contrastes m sensibles a errores en la especificación dinámica para retardos de orden

superior. Por ejemplo, un contraste de autocorrelación de órdenes 1 y 2 se puede computar como nR^2 no centrado de la regresión auxiliar de

$$1 \quad sobre \quad r_t r_{t-1},\ r_t r_{t-2},\ r_t,\ r_t y_{t-1},\ r_t x_t$$

o bien si utilizamos la versión de (12), sería el nR^2 de la regresión auxiliar de

$$r_t \quad sobre \quad r_{t-1},\ r_{t-2},\ 1,\ y_{t-1},\ x_t$$

que, de nuevo, sería idéntico al contraste LM de autocorrelación de órdenes 1 y 2. Eligiendo los elementos apropiados de las matrices R_j^0 se pueden diseñar contrastes sensibles a los errores de especificación que sean de interés para cada modelo considerado.

Como se ha visto en esta sección, en el caso de regresión lineal, los contrastes de la matriz de información dinámica dan lugar a contrastes de la especificación dinámica tanto conocidos como más novedosos y tanto de la media condicional como de la varianza condicional (tipo ARCH).

Estos resultados sugieren que los contrastes de la matriz de información dinámica van a ser una fuente de diagnósticos relevantes y reveladores en modelos estimados por máxima verosimilitud. Un ejemplo de esto serían los diagnósticos propuestos por Pérez (1993c) para modelos estimados por ARCH.

6. El comportamiento en muestras finitas de los contrastes m

Recientemente se han obtenido resultados que indican un posible mal comportamiento en muestras finitas de los contrastes m. De hecho, el comportamiento en muestras finitas de algunas versiones de los estadísticos puede no ser satisfactorio, tal y como ilustran Cavanagh (1985), Kennan y Neumann (1988), Orme (1990a) y Chesher y Spady (1991). En general, estos autores encuentran una tendencia al sobrerrechazo de algunas versiones de los contrastes. Esta tendencia se puede achacar a que la χ_p^2 no es una buena aproximación al comportamiento de algunas versiones de los contrastes, particularmente de aquellas que utilizan el producto exterior del gradiente como estimador de la matriz de varianzas y covarianzas.

Estos trabajos han motivado una líneas de investigación centrada en obtener modificaciones de los contrastes o de la distribución que los aproxima que puede proporcionar un mejor control sobre el tamaño de los contrastes en muestras de tamaños habituales en econometría. Trabajos que avanzan en esta dirección son los de Kennan y Neumann (1988), Davidson y Mackinnon (1988), Orme (1990b), Chesher y Spady (1991) y Wooldridge (1990).

De cualquier forma, cabe resaltar que lo que define un contraste m es su hipótesis nula $E_0\ (m_t\ (\omega,\ \theta)) = 0$, y no cuál sea el estadístico concreto que se utilice para su computación. Si tomamos como ejemplo un contraste de autocorrelación de orden uno para el caso de regresión lineal, la hipótesis nula sería $E_0\ (u_t\ u_{t-1}) = 0$, y habría distintos estadísticos para comprobar si esa hipótesis nula se da en la muestra, por ejemplo:

$$1 \quad sobre \quad r_t\, r_{t-1} \tag{22}$$

que corresponde a (9) o bien (20) y (21) que acabamos de presentar.

Los resultados de los trabajos que acabamos de citar sugieren que si utilizamos muchos indicadores conjuntamente y, además, hay residuos no simétricos o atípicos, entonces la distribución de (22) y, probablemente (21), estará mal aproximada por una χ^2 .

Cuando se utilicen contrastes m sería conveniente seleccionar un número razonablemente pequeño de indicadores, que se traten previamente los atípicos, y que se utilicen versiones de los contrastes, en nuestro caso (21), que tengan un mejor comportamiento en muestras finitas; véase Pérez (1993b).

No se debe descalificar a los contrastes m basándose en que algunas versiones de la computación de los mismos son insatisfactorias. La mayoría de los diagnósticos habituales son casos particulares de los contrastes m (y de la matriz de información) y, en general, hay versiones que tienen un comportamiento satisfactorio en muestras finitas.

7. Conclusiones

En este trabajo se han sintetizado los principales resultados de la literatura reciente sobre contrastes de momentos (m) y de la matriz de información. Los contrastes m fueron propuestos independientemente por Newey(1985) y Tauchen(1985), y generalizados por White (1987) para el caso de observaciones dependientes. Los contrastes de la matriz de información fueron propuestos por White(1982, 1987, 1994) y desarrollados para algunos casos particulares por Pérez (1989, 1993 a y c).

Los contrastes m son un marco general para llevar a cabo diagnósticos de modelos, basándose en que bajo especificación correcta habrá algunos momentos muestrales que deben ser cero. Además se demuestra que bajo condiciones generales los contrastes de momentos pueden ser considerados contrastes de los multiplicadores de lagrange (LM).

Los contrastes de la matriz de información pueden ser considerados una fuente para obtener condiciones de momentos. En este trabajo nos hemos centrado en los contrastes de la matriz de información de primer orden (dinámicos), por ser más recientes y menos conocidos que los de segundo orden.

Hemos visto cómo los contrastes de la matriz de información de primer orden (dinámicos) dan lugar, en el caso de regresión lineal, a un conjunto de contrastes interpretables y en algunos casos ya conocidos, así como fácilmente computables.

Estos contrastes han sido criticados por su, en ocasiones, deficiente comportamiento en muestras finitas. Habría que señalar que la crítica sería aplicable a algunas formas de computar los contrastes más que a los contrastes en sí. Ya hemos visto que, en general, hay diferentes maneras de computar un contraste dado, y habría que elegir en cada caso la más apropiada. Asimismo, cabría señalar que ya se ha abordado el estudio de modificaciones de la computación de los contrastes para conseguir un mejor comportamiento en muestras finitas. Resulta crucial y es responsabilidad del investigador aplicado, elegir tanto un número de indicadores razonablemente pequeño como una versión del contraste que sea apropiada en cada situación y aproveche la información disponible de la manera más eficiente.

El verdadero valor de la literatura sobre contrastes m y de la matriz de información reside en que aporta los resultados generales para desarrollar baterías de diagnósticos para modelos tanto estimados por método de momentos como por máxima verosimilitud, especialmente si se usan verosimilitudes menos habituales que la gaussiana, como podrían ser ARCH, GARCH, poisson, gamma y binomial. El desarrollo de estos diagnósticos es un reto para la investigación futura.

APENDICE 1

Prueba simplificada de (12).

Sea R una matriz diagonal cuyos elementos son los residuos $\hat{\nu}_t$

X: matriz de $\hat{w}_t$ (para respetar la notación original).

W: matriz de $\hat{j}_t$.

i: vector columna de unos.

Entonces, $\hat{m}' \equiv R\ W$ y $\hat{s}' \equiv R\ X$. Utilizando la notación anterior podemos expresar (11) como nR_0^2 de i sobre RW, RX, que será:

$$nR_0^2 = \left(i'RW/\sqrt{n},\ i'RX/\sqrt{n}\right) \tag{A.1}$$

$$\begin{bmatrix} W'RRW/n & W'RRX/n \\ X'RRW/n & X'RRX/n \end{bmatrix}^{-1} \begin{bmatrix} W'Ri/\sqrt{n} \\ X'Ri/\sqrt{n} \end{bmatrix}$$

y como $i'RX/\sqrt{n}$ es $o_p(1)$, (por ser los residuos ortogonales a los regresores), entonces

$$nR_0^2 = (i'RW/\sqrt{n})\ (\ (W'RRW/n)\ -\ (W'RRX/n)\ (X'RRX/n)^{-1}\ (X'RRW/N))^{-1}$$
$$(W'Ri/\sqrt{n})\ +\ o_p(1)$$

y bajo homoscedasticidad condicional quedaría

$$nR_0^2 = (i'RW/\sqrt{n})\ (\ (W'W/n)\ -\ (W'X/n)\ (X'X/n)^{-1}\ (X'W/N))^{-1}$$
$$(W'Ri/\sqrt{n})\ /\ \hat{\sigma}_n^2\ +\ o_p(1)$$

por tanto, nR_0^2 de i sobre RW, RX sería asintóticamente equivalente a esta expresión.

Por otro lado, nR^2 de la regresión mco de los residuos $\hat{\nu}$ sobre W, X sería, (dado que $\hat{\nu}$ = Ri)

$$nR^2 = (i'RW/\sqrt{n},\ i'RX/\sqrt{n})$$

$$\begin{pmatrix} W'W/n & W'X/n \\ X'W/n & X'X/n \end{pmatrix}^{-1} \begin{pmatrix} W'Ri/\sqrt{n} \\ X'Ri/\sqrt{n} \end{pmatrix} \ /\ (\ i'RRi/n)$$

y como $iRX/\sqrt{n} = o_p(1)$ y $i'RRi\ /\ n = \hat{\sigma}_n^2,$ tendríamos que

$$nR_0^2 = (i'RW/\sqrt{n}\,)\ (\ (W'W/n) - (W'X/n)\ (X'X/n)^{-1}\ (X'W/N))^{-1}$$
$$(W'Ri/\sqrt{n})\ /\ \hat{\sigma}_n^2 + o_p(1)$$

con lo que se obtiene lo que se quería demostrar, esto es, que las expresiones resultantes de (11) y (12) son asintóticamente equivalentes.

APENDICE 2

Bajo condiciones de regularidad generales

$$\hat{\theta}_n - \theta_n^* \to 0 \quad prob.\text{-} \; P_0,$$

siendo θ_n^* el verdadero óptimo de la función de verosimilitud utilizada, y, además,

$$\sqrt{n}\ (\hat{\theta}_n - \theta_n^*) \sim N(0,\ A_n^{*\,-1}\ B_n^*\ A_n^{*\,-1})$$

donde

$$A_n^* \equiv n^{-1}\sum_{t=1}^{n} E\ (\nabla^2 \log f_t\ (\theta_n^*)) \quad (hesiano)$$

donde ∇^2 denota la matriz de derivadas parciales segundas con respecto a los parámetros θ.

$$B_n^* \equiv var\ (\sqrt{n}\ \nabla\ n^{-1}\sum_{t=1}^{n} \log f_t\ (\theta_n^*)) \quad (producto\ \ exterior)$$

y bajo especificación correcta, $(\theta_n^* = \theta_0)$ la varianza de los parámetros estimados sería $avar\ \hat{\theta}_n = B_n^{0^{-1}}$, porque, debido a la igualdad de la matriz de información, que se deriva a continuación, $A_n^* = - B_n^*$; (donde B_n^* es la matriz de información de Fisher).

A.2.1. Derivación de la igualdad de la matriz de información dinámica (de primer orden)

La igualdad de la matriz de información se obtiene partiendo de que para toda función de densidad

$$\int f_t\ (\theta)\ d\nu_t = 1, \quad -P_0$$

y además, diferenciando con respecto a los parámetros θ tendríamos

$$\nabla \int f_t\ (\theta)\ d\nu_t = 0, \quad -P_0 \tag{A.2.1}$$

y por ello:

$$\nabla \int f_t\ (\theta)\ d\nu_t = \int \nabla f_t\ (\theta)\ d\nu_t = \int \nabla \log f_t\ (\theta)\ f_t(\theta)\ d\nu_t = 0, \quad -P_0 \tag{A.2.2}$$

y suponiendo especificación correcta, esto es, $\theta = \theta_0$:

$$\int \nabla \log f_t\ (\theta_0)\ \ f_t\ (\theta_0)\ d\nu_t = E\ (\nabla \log f_t\ (\theta_0) / \tilde{X}^{t-1}) = 0, \quad -P_0$$

donde $\tilde{X}^{t-1}$ es el conjunto de variables predeterminadas. Esto significa que el gradiente $s_t^0 \equiv \nabla \log f_t\ (\theta_0)'$ es una secuencia de diferencias de martingalas, esto es,

$$E\ (s_t^0\ /\ \tilde{X}^{t-1}) = 0 \Rightarrow E\ (\ s_t^0\ s_\tau^{0'}) = 0, \quad \forall\ t \neq \tau$$

y eso implicará que

$$B_n^* \equiv var\ (n^{-1/2}\sum_{t=1}^{n} s_t^0)) = n^{-1} \sum_{t=1}^{n} var\ s_t^0 \equiv \bar{B}_n^*$$

esto es, la igualdad de la matriz de información dinámica, de primer orden,

$$B_n^* = \bar{B}_n^*$$

La condición suficiente para esta igualdad es que el gradiente sea una secuencia de diferencias de martingalas.

A.2.2. Derivación de la igualdad de la matriz de información de segundo orden.

Si derivamos de nuevo la expresión (A.1) tendremos

$$\nabla^2 \int f_t(\theta)\, d\nu_t = 0, \quad -P_0$$

$$\nabla^2 \int f_t(\theta)\, d\nu_t = \int \nabla^2 f_t(\theta)\, d\nu_t = \int [\nabla \log f_t(\theta)' \nabla \log f_t(\theta) + \nabla^2 \log f_t(\theta)]\, f_t(\theta)\, d\nu_t = 0$$

y bajo especificación correcta:

$$\int [\nabla s_t^0 + s_t^0 s_t^{0\prime}]\, f_t(\theta_0)\, d\nu_t =$$

$$E[\nabla s_t^0 + s_t^0 s_t^{0\prime} / \tilde{X}^{t-1}] = 0$$

lo que implica que $\nabla s_t^0 + s_t^0 s_t^{0\prime}$ es una secuencia de diferencias de martingalas, de forma que se daría

$$A_n^* \equiv n^{-1} \sum_{t=1}^{n} E(\nabla s_t^0) = -n^{-1} \sum_{t=1}^{n} E(s_t^0 s_t^{0\prime}) = -\bar{B}_n^*$$

que es la igualdad de la matriz de información de segundo orden

$$A_n^* = -\bar{B}_n^*.$$

Combinando las dos igualdades, de primero y segundo orden, tenemos:

$$A_n^* = -B_n^*,$$

que es la igualdad de la matriz de información.

REFERENCIAS

Aitchison, J. y S. D. Silvey (1958) «Estimation subject to restraints,» Annals of Mathematical Statistics 29, 813-828.

Cavanagh, C. L. (1985) «Second order admissibility of likelihood based tests,» Harvard Institute of Economic Research Discussion paper 1148.

Chesher, A. y Spady, R. (1991) «Asymptotic expansions of the information matrix test statistic» Econometrica 59, 787-816.

Cox, D. R. (1961) «Tests of separate families of hypotheses,» en Proceedings of the Fourth Berkeley Simposium on Mathematical Statistics and Probability, Vol. 1 Berkeley: University of California Press, 105-123.

Cox, D. R. (1962) «Further results on tests of separate families of hypotheses,» Journal of the Royal Statistical Society B, 406-424.

Davidson, R. y J. G. MacKinnon (1992) «A new form of the information matrix test», Econometrica 60, 145-158.

Godfrey, L. G. (1988) Misspecification tests in econometrics, Econometric Society monograhs 16, Cambridge University Press.

Hausman, J. (1978) «Specification tests in econometrics», Econometrica 46, 1251-1272.

Kennan, J. y Neumann, G. R. (1988) «Why does the information matrix test reject too often? a diagnosis of some Monte Carlo symptoms», Working Papers in Economics E-88-10, Hoover Institution, Stanford University.

Newey, W. (1985) «Maximum likelihood specification testing and conditional moment tests», Econometrica 53, 1047-1070.

Orme, C. (1990a) «The small sample performance of the information matrix test», Journal of Econometrics 46, 309-331.

Orme, C. (1990b) «Double and triple length regressions for the information matrix test and other conditional moment tests», University of New York, mimeo.

Pérez, T. (1989) «Dynamic information matrix tests and m tests: new results, simulations and an application», University of California San Diego Department of Economics Doctoral Dissertation.

Pérez, T. (1993a) «Contrastes m y de la matriz de información dinámica con una aplicación a regresión lineal», Estadística Española, Vol 35, 133, 291-317.

Pérez, T. (1993b) «El comportamiento en muestras finitas de los contrastes m y de la matriz de información dinámica: simulaciones», Estadística Española, Vol. 35, 132, 123-139.

Pérez, T. (1993c) «Comentario a "Guía para la estimación de contrastes ARCH"», Estadística Española, Vol 35, 132, 64-67.

Rao, C. R. (1947) «Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation,» Proceedings of the Cambridge Philosophical Society 44, 50-57.

Taylor, L. W. (1987) «The size bias of the information matrix test», Economics Letters 24, 63-67.

Tauchen, G. (1985) «Diagnostic testing and evaluation of maximum likelihood models», Journal of Econometrics 30, 515-444.

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

White, H. (1982) «Maximum likelihood estimation of misspecified nodels», Econometrica 50, 1-25.

White, H. (1987) «Specification testing in dynamic models», en Advances in Econometrics, Fifth World Congress, Vol. 1, Bewley, T. (editor), New York, Cambridge University Press, 1-58.

White, H. (1994) Estimation, Inference and Specification Analysis, Cambridge University Press.

Wooldridge, J. (1986) «Asymptotic Properties of Econometrics Estimators» University of California San Diego Department of Economics Doctoral Dissertation.

Wooldridge, J. (1990) «A unified approach to robust, regression-based specification tests» Econometric Theory 6, 17-43.

SERIE DE DOCUMENTOS DE TRABAJO DEL *ICAE*

9301 *"Análisis del Comportamiento de las Cotizaciones Reales en la Bolsa de Madrid bajo la Hipótesis de Eficiencia"*. Rafael Flores de Frutos. Diciembre 1992. (Versión final aceptada para publicación en Estadística Española)

9302 *"Sobre la Estimación de Primas por Plazo dentro de la Estructura Temporal de Tipos de Interés"*. Rafael Flores de Frutos. Diciembre 1992.

9303 *"Cambios de Estructuras de Gasto y de Consumo en el Cálculo del IPC"*. Antonio Abadía. Febrero 1993. (Versión final publicada en Revista de Economía Aplicada, Vol.1, N°1)

9304 *"Tax Analysis in a Limit Pricing Model"*. Félix Marcos. Febrero 1993.

9305 *"El Tipo de Cambio Propio: Reformulación del Concepto y Estimación para el Caso Español"*. José de Hevia Payá. Junio 1993. (Versión final aceptada para publicación en Revista Española de Economía)

9306 *"Price Volatility Under Alternative Monetary Instruments"*. Alfonso Novales. Abril 1992.

9307 *"Teorías del Tipo de Cambio: Una Panorámica"*. Oscar Bajo Rubio. Simón Sosvilla Rivero. Junio 1993. (Versión final publicada en Revista de Economía Aplicada, Vol.1, N°2).

9308 *"Testing Theories of Economic Fluctuations and Growth in Early Development (the case of the Chesapeake tobacco economy)"*. Rafael Flores de Frutos. Alfredo M. Pereira. Diciembre 1992.

9309 *"Maastricht Convergence Conditions: A Lower Bound for Inflation?"*. Jorge Blázquez. Miguel Sebastián. Marzo 1992.

9310 *"Recursive Identification, Estimation and Forecasting of Nonstationary Economic Time Series with Applications to GNP International Data"*. A. García-Ferrer. J. del Hoyo. A. Novales. P.C. Young. Marzo 1993.

9311 *"General Dynamics in Overlapping Generations Models"*. Carmen Carrera. Manuel Morán. Enero 1993. (Versión final aceptada para publicación en Journal of Economic Dynamics and Control)

9312 *"Further Evidence on Forecasting International GNP Growth Rates Using Unobserved Components Transfer Function Models"*. A. García-Ferrer, J. del Hoyo, A. Novales, P.C. Young. Septiembre 1993. (De próxima aparición en un volumen de homenaje a A. Zellner)

9313 *"Public Capital and Aggregate Growth in the United States: Is Public Capital Productive?"*. Rafael Flores de Frutos. Alfredo M. Pereira. Julio 1993.

9314 *"Central Bank Structure and Monetary Policy Uncertainty"*. José I. García de Paso. Abril 1993.

9315 *"Monetary Policy with Private Information: A Role for Monetary Targets"*. José I. García de Paso. Julio 1993.

9316 *"Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models"*. José Alberto Mauricio. Julio 1993. (Versión final aceptada para publicación en Journal of the American Statistical Association)

9317 *"The Exact Likelihood Function of a Vector ARMA Model"*. José Alberto Mauricio. Julio 1993.

9318 *"Business Telephone Traffic Demand in Spain: 1980-1991, An Econometric Approach"*. Teodosio Pérez Amaral. Francisco Alvarez González. Bernardo Moreno Jiménez. Septiembre 1993. (Versión final aceptada para publicación en Information Economics and Policy)

9401 *"Contrastes de momentos y de la matriz de información"*. Teodosio Pérez Amaral. Junio 1994. (Versión final aceptada para publicación en Cuadernos Económicos del ICE)

9402 *"A partisan explanation of political monetary cycles"*. José I. García de Paso. Junio 1994.

9403 *"Estadísticos para la detección de observaciones anómalas en modelos de elección binaria: Una aplicación con datos reales"*. Gregorio R. Serrano. Junio 1994.

9404 *"Effects of public investment in infraestructure on the spanish economy"*. Rafael Flores de Frutos. Mercedes Gracia Díez. Teodosio Pérez Amaral. Junio 1994.

9405 *" Observaciones anómalas en modelos de elección binaria"*. Mercedes Gracia. Gregorio R. Serrano. Junio 1994. (Versión final aceptada para publicación en Estadística Española)

9406 *" Permanent components in seasonal variables"*. Rafael Flores. Alfonso Novales. Junio 1994.

9407 *" Money demand instability and the performance of the monetary model of exchange rates"*. Rodrigo Peruga. Junio 1994.

9408 *" Una nota sobre la estimación eficiente de modelos con parámetros cambiantes"*. Sonia Sotoca. Junio 1994.

9409 *" Distribución de la renta y redistribución a través del IRPF en España"*. Rafael Salas. Junio 1994.

9410 *" Trade balances: Do exchange rates matter?"*. Rodrigo Peruga. Junio 1994.