LA ECONOMETRÍA ESTRUCTURAL: UNA NOTA METODOLÓGICA

Hernando C. L. Sabau

I. INTRODUCCIÓN

La incapacidad que mostraron los modelos econométricos para dar cuenta de los hechos económicos que sucedieron al embargo petrolero de 1973 generó un amplio debate cuya riqueza ha transformado, particularmente en los diez años recién pasados, la teoría y la práctica econométricas. En el núcleo del debate ha estado el intento de cerrar la brecha entre teoría y práctica, proponiendo metodologías cuya coherencia lógica no dependa del "axioma de especificación correcta" (Leamer, 1978), cuyo papel es central en los textos de econometría clásicos de la década de los años setenta (por ejemplo, Dhrymes, 1970; Theil, 1971; Johnston, 1972). Aunque aún persisten muchos problemas, consideramos que el avance ha sido importante en la última década, y ha elevado considerablemente el nivel del trabajo aplicado. Con la presentación de este número de *Economía Mexicana*, al tiempo que exponemos el resultado de nuestras investigaciones de un año y las opiniones que de ellas derivamos en relación con las perspectivas de nuestra economía, pretendemos sentar una base metodológica más sólida para nuestro análisis econométrico, y aspiramos a que nuestro punto de vista contribuya al debate metodológico en búsqueda de mejorar la calidad del trabajo aplicado que se desarrolla en nuestro país.

En la segunda sección de esta nota presentamos nuestras consideraciones en relación con el papel de los argumentos teóricos y los argumentos empíricos para la construcción de modelos econométricos, y, a partir de ellas, fundamentamos nuestra opción metodológica. En la tercera sección se hace la distinción entre el proceso generador de información y el modelo que pretende aproximarlo, estableciendo la relación entre ambos. Los mecanismos para evaluar el modelo desde el punto de vista econométrico se sustentan en la cuarta sección, que describe así las pruebas de diagnóstico reportadas en los distintos artículos de corte econométrico. Finalmente, en la quinta sección revisamos algunos de los problemas que consideramos siguen vigentes.

Con esta nota reforzamos algunas de las ideas planteadas en Ruprah y Sabau (1984) y modificamos algunas posiciones, reflejando las implicaciones de nuestra reflexión metodológica en los años recientes.

II. ARGUMENTOS TEÓRICOS Y EMPÍRICOS

Partimos de la base de que el objeto de la econometría es el estudio de uno o un conjunto de aspectos de una economía específica, mediante el uso sistemático de la reflexión teórica y de la información empírica. Es así que el trabajo econométrico utiliza dos fuentes de información y, por ende, dos tipos de argumentos: el teórico y el empírico. Nuestra posición es que se debe otorgar el mismo nivel de validez a ambos argumentos. Desde un punto de vista práctico, de conocimiento de una economía en particular, resulta de poco interés una teoría que no sea coherente con el conocimiento empírico. De igual forma, la mera descripción de los datos con base en sus propiedades estocásticas es de escaso valor para el análisis si no existe una racionalización y explicación normativa de las relaciones entre variables. En contraste, la fuerza y el poder de convencimiento de un argumento teórico se ven enriquecidos si éste es capaz de explicar la realidad.

Hay que aclarar que nuestra propuesta no es en forma alguna opuesta a la investigación teórica pura, la que constituye un ejercicio básico del que podemos aprender y establecer proposiciones para nuevos estudios aplicados. Reconocemos, igualmente, que muchos fenómenos no son susceptibles de estudio econométrico con el estado actual del arte, dado que su medición puede resultar en un problema colosal, al tiempo que la complejidad de sus argumentos puede ser un obstáculo difícil de superar. También reconocemos que en muchos casos los conceptos teóricos y los conceptos empíricos no concuerdan. Este problema puede resolverse en muchas instancias si partimos de reconocer la diferencia entre un *modelo teórico* y un *modelo estimable* (Spanos, 1986). El primero de ellos plantea las relaciones entre conceptos teóricos, mientras que el segundo traduce dichas relaciones en términos de conceptos observables y *medibles*, tomando en cuenta las bases de construcción de la información empírica disponible. Tampoco negamos la utilidad de modelos puramente empíricos, tales como los propuestos por Box y Jenkins (1971), pero confinamos esta utilidad al terreno de la predicción.

De las tres principales propuestas metodológicas surgidas a partir del debate econométrico de los años recientes, encontramos que sólo una incorpora este balance entre los argumentos teóricos y empíricos (para una discusión de las tres metodologías véase Pagan, 1985). La metodología propuesta por Leamer (1978, 1983, 1985a) sobre argumentos de corte bayesiano otorga un peso específico a la información *a priori* que, en algunos casos, no puede ser revertida por ningún tipo de evidencia empírica (ver Mc Aleer *et al.* [1985, 1986]). En contraste, la metodología propuesta por Sims (1980) (véase también Doan *et al.* [1984]) basada en las autorregresiones vectoriales otorga un peso desusado a la información empírica, sobre el argumento de que el problema de identificación es insoluble a nivel empírico en economía y por lo tanto carece de sentido el estudio de relaciones estructurales. Esta escuela ha tenido en nuestro país cierta influencia, y a nivel metodológico ha sido criticada por Leamer (1985b) y Cooley y Leroy (1985), entre otros.

La tercera escuela, que llamaremos la econometría estructural, es la que a nuestro juicio hace un mejor balance de los argumentos de distinto tipo y, si bien no está exenta de problemas técnicos y de interpretación, nos parece la más adecuada y por ello es la que adoptamos y discutimos en las siguientes secciones de esta nota. Esta metodología tiene sus raíces en el trabajo de Sargan (1964) y es propuesta de una manera más acabada en Hendry y Richard (1982, 1983), encontrando finalmente una expresión pedagógica en el texto de Spanos (1986).

La econometría estructural asume la validez de aprender tanto de la reflexión teórica como de la información empírica, reconociendo que la primera debe aportar los argumentos para sustentar la causalidad entre variables, mientras que la segunda contribuye con los aspectos que competen a las relaciones de dependencia que surgen del procedimiento de observación. Así, el argumento teórico se refiere a la distribución *incondicional* conjunta de la(s) variable(s) estudiada(s), digamos y, y las variables explicativas, digamos z, proporcionando elementos para definir la distribución condicional de y dada z. Lo anterior implica un proceso de marginalización de las variables no relevantes e introduce el primer componente para definir el conjunto de información condicionante del modelo: las variables explicativas z. El argumento empírico introduce de forma explícita la dimensión temporal, muestra la evolución de la información disponible en cada momento y con ello centra el interés en el estudio dinámico de y. Con esto, la conjunción de ambos argumentos se refiere a la distribución *condicional* de y_t, la(s) variable(s) explicada(s) en el periodo t dada la información disponible al momento t. Así, el conjunto de información condicionante F_t, se compone por el presente z_t de z en vista del argumento teórico, y por toda la información pasada en vista del argumento empírico. Es decir

$$F_t = \sigma \left\{ y_{t-j-1} \ , z_{t-j} \right\}_{j=0}^{\infty}$$

donde σ (.) indica que al conjunto de observaciones se le da la estructura de un conjunto de información.

Existe un problema de identificación entre los modelos teóricos –infinitos en principio– y la realidad observada, que es única. Es decir, existen infinitas explicaciones compatibles con la observación empírica. Este problema es irresoluble, en principio, pero al respecto podemos hacer dos proposiciones: *i*) el espacio de explicaciones se verá sustancialmente reducido al eliminar las proposiciones sin coherencia lógica, *ii*) es posible reducir aún más el número de explicaciones posibles si la coherencia teoría-dato se califica de forma estricta. Así, la econometría estructural propone que los modelos deben evaluarse de forma exhaustiva y que solamente los que satisfagan un alto nivel de coherencia teórico-empírica pueden ser calificados como razonables, aun cuando esto no asegura que sean correctos.

III. PROCESO GENERADOR DE INFORMACIÓN Y MODELO

Para romper con el axioma de especificación correcta es necesario distinguir al *proceso generador de información* (PGI) (el mecanismo real que subyace en la generación de la información empírica) del modelo (el esfuerzo de investigador por aproximar el PGI), y así adquirir conocimiento acerca del objeto de estudio utilizando teoría y datos.

El PGI está dado por la distribución de probabilidad de y_t dado el conjunto de información F_t. Nótese que la existencia de tal mecanismo probabilístico es poco restrictiva por referirse a una distribución *condicional* y no a una *incondicional*, ya que no se requiere introducir supuesto alguno en relación con la homogeneidad y memoria del proceso en el tiempo (v. gr., no requiere suponer aspectos tales como ergodicidad, estacionariedad, etc.). En correspondencia, el modelo propone una forma específica y parametrizada para dicho mecanismo. Para ilustrar la relación entre ambos, consideramos el primer momento (la media), que es usualmente el foco de atención de los investigadores. Supongamos que

$$\bar{\mu}_t = E\left[y_t \mid F_t\right]$$

es decir, $\bar{\mu}_t$ es la media del PGI para y_t dado F_t. Definimos las innovaciones

$$\bar{u}_t = y_t - E\left[y_t \mid F_t\right] = y_t - \bar{\mu}_t \,, \tag{1}$$

lo que permite descomponer a y_t en un componente sistemático dado por $\bar{\mu}_t$ y uno no sistemático $\bar{u}_t$, pues de (1) obtenemos

$$y_t = \bar{\mu}_t + \bar{u}_t \,. \tag{2}$$

La no sistematicidad de $\bar{u}_t$ se sigue de que $\bar{\mu}_t \in F_t$ (por ser momento condicional en F_t) y, por lo tanto,

$$E\left[\bar{u}_t \mid F_t\right] = E\left[y_t \mid F_t\right] - \bar{\mu}_t = 0, \tag{3}$$

mientras que, para $s < t$,

$$E\left[\bar{u}_t \ \bar{u}_s \mid F_t\right] = E\left[\bar{u}_t \mid F_t\right] \ \bar{u}_s = 0, \tag{4}$$

pues $\bar{u}_s = y_s - \bar{\mu}_s \in F_t$ para $s < t$.

El modelo propone que la media de y_t condicional a F_t toma la forma μ_t (β), donde μ_t es función del conjunto de información (i.e., del pasado de y_t y del presente y pasado de z_t), $\mu_t \in F_t$, y el vector de parámetros β pertenece a un cierto espacio parametral B, $\beta \in B$. Si suponemos que la especificación es correcta, entonces existe $\beta_0 \in B$ tal que $\mu_t = \mu_t(\beta_0)$. Definimos las innovaciones del modelo

$$u_t = u_t(\beta) = y_t - \mu_t(\beta), \tag{5}$$

a las que corresponde la descomposición sistemática/no sistemática modelada

$$y_t = \mu_t(\beta) + u_t, \tag{6}$$

que se vincula con la descomposición del PGI en (2) mediante

$$y_t = \bar{\mu}_t + \bar{u}_t + \mu_t(\beta) - \mu_t(\beta) = \mu_t(\beta) + u_t,$$

donde

$$u_t = \bar{u}_t + \left[\bar{\mu}_t - \mu_t(\beta)\right] \tag{7}$$

de forma que u_t tiene dos componentes: la innovación real $\bar{u}_t$, y un componente de error dado por la diferencia entre el momento real $\bar{\mu}_t$ y el modelado μ_t. Así, las propiedades de no sistematicidad $\bar{u}_t$ se reproducen en u_t sólo si $\bar{\mu}_t = \mu_t$, es decir, si el modelo está bien especificado y u_t se evalúa en β_0, el valor real de los parámetros.

La relación entre modelo y PGI resumida en (7) –en la esfera de la media de la distribución– tendrá una función relevante en los mecanismos de evaluación que discutiremos en la siguiente sección. Como ejercicio de modelística nos interesa empequeñecer lo más posible el componente de error con el fin de que el modelo sea una buena aproximación del PGI. Cabe mencionar que uno de los aportes teóricos importantes de esta escuela estructural es que ha permitido darle sentido a la inferencia en situaciones de especificación incorrecta, destacando los trabajos de White (1982), Domowitz y White (1982), y Gourieroux *et al.* (1985a, 1985b).

La media de la distribución produce información con respecto a la localización de la masa de probabilidad, pero, salvo para densidades muy específicas (v. gr. Posson), esto es insuficiente para conocer toda la distribución. Supongamos que los momentos de orden 2 y mayores del PGI alrededor de la media están dados por

$$\bar{h}_t^{(r)} = E\left[(y_t - \mu_t)^r \mid F_t\right] = E\left[\bar{u}_t^r \mid F_t\right], \quad r = 2, 3, \ldots, \tag{8}$$

para los que definimos las correspondientes innovaciones

$$\bar{\epsilon}_t^{(r)} = \bar{u}_t^r - E\left[\bar{u}_r^r \mid F_t\right] = \bar{u}_t^r - \bar{h}_t^{(r)}, \quad r = 2, 3, \ldots, \tag{9}$$

que poseen las propiedades (3) y (4) de no sistematicidad por definición, y que, por lo tanto, permiten descomponer

$$\bar{u}_t^r = \bar{h}_t^{(r)} + \bar{\epsilon}_t^{(r)}, \tag{10}$$

en un componente sistemático y uno no sistemático. Para caracterizar el PGI requerimos todos sus momentos ($\bar{\mu}_t$ y $\bar{h}_t^{(r)}$, $r = 2, 3, \ldots$), o alternativamente su función característica o su función de distribución y los momentos que la determinan. El modelo busca aproximar el PGI mediante la transformación del modelo estimable en un modelo probabilístico, que incluye una proposición para la forma de la distribución y la parametrización de los momentos, digamos N en número, que determinen completamente a la distribución propuesta. Así el modelo probabilístico toma la forma

$$y_t \mid F_t \sim D\left[\mu_t(\beta), h_t^{(2)}(\theta_2), h_t^{(9)}(\theta_9), \ldots, h_t^{(N)}(\theta_N)\right], \tag{11}$$

donde $D(.)$ es una distribución caracterizada por sus N primeros momentos, $h_t^{(r)}(\theta_r)$ es la parametrización del r-ésimo momento alrededor de la media, y $\theta = (\beta', \theta_2', \ldots, \theta_N')$, es el vector de parámetros. Las innovaciones del modelo para $r \geq 2$ son

$$\epsilon_t^{(r)} = \epsilon_t^{(r)}(\beta, \theta_r) = u_t^r - h_t^{(r)}(\theta_r), \tag{12}$$

que indican la descomposición

$$u_t^r = h_t^{(r)}(\theta_r) + \epsilon_t^{(r)} \tag{13}$$

Lo anterior define para cada momento una ecuación de regresión similar a la obtenida en (6) para la media, y se vincula con la descomposición del correspondiente momento del PGI en (10) usando (7) y (12).

$$\begin{aligned}
\epsilon_t^{(r)} &= u_t^r - h_t^{(r)}(\theta_r) \\
&= \left[\bar{u}_t + \left\{\bar{\mu}_t - \mu_t(\beta)\right\}\right]^r - h_t^{(r)}(\theta_r) \\
&= \sum_{j=0}^{r} \begin{bmatrix} r \\ j \end{bmatrix} \bar{u}_t^{(r-j)} \left\{\bar{\mu}_t - \mu_t(\beta)\right\}^j - h_t^{(r)}(\theta_r) \\
&= \bar{u}_t^r - h_t^{(r)}(\theta_r) + \sum_{j=1}^{r} \begin{bmatrix} r \\ j \end{bmatrix} \bar{u}_t^{(r-j)} \left\{\bar{\mu}_t - \mu_t(\beta)\right\}^j
\end{aligned}$$

y usando (10) obtenemos

$$\epsilon_t^{(r)} = \bar{\epsilon}_t^{(r)} + \left[\bar{h}_t^{(r)} - h_t^{(r)}(\theta_r)\right] + \sum_{J=1}^{r} \begin{bmatrix} r \\ j \end{bmatrix} \bar{u}_t^{(r-j)} \left\{\bar{\mu}_t - \mu_t(\beta)\right\}^j, \tag{14}$$

de forma que $\epsilon_t^{(r)}$ tiene tres componentes: la innovación del PGI, y dos componentes de error. El primer componente de error se refiere a la diferencia entre el momento real y el modelado. El segundo componente de error refleja los errores inducidos por el error de especificación en el primer momento y aparece en vista de la parametrización alrededor de la media de momentos de orden dos y superiores. Al igual que (7) para la media, (14) será de decisiva importancia para los mecanismos de evaluación en momentos superiores. El supuesto de especificación correcta se traduce ahora en que existe $\theta_0 = \beta_0', \theta_2^{0\,\prime}, \ldots, \theta_N^{0\,\prime}$ en el espacio parametral correspondiente θ, tal que $\bar{\mu}_t = \mu_t(\beta_0)$ y $\bar{h}_t^{(r)} = h_t^{(r)}(\theta_r^0)$, $r = 2, \ldots, N$, y $D(.)$ sea la distribución correcta, de manera que se cumpla también que

$$h_t^{(r)} = h_t^{(r)}(\theta_0) \text{ para } r = N+1, N+2, \ldots$$

Para completar el modelo estadístico que permita conjuntar información teórica y empírica es necesario introducir un modelo muestral que se refiera a los datos, tanto en sus propiedades probabilísticas como a la forma como se obtienen las observaciones en el periodo muestral que incluye observaciones acerca de los periodos $t = 1, \ldots, T$. Conocido el PGI, estaríamos en condiciones de hacer proposiciones probabilísticas con respecto a la muestra. Vista a partir del modelo, la densidad conjunta proporciona la función de verosimilitud $L(\theta)$, que es función de θ dada la información empírica G_T, donde $G_t = \left\{ y_{t-j}; z_{t-j} \right\}_{j=0}^{\infty}$. Con información de corte transversal, $L(\theta)$ se construye tomando en cuenta el proceso de muestreo utilizado para obtener los datos. Con información de serie de tiempo, la dependencia debe estar en principio incorporada en la especificación de los momentos condicionales, aunque hay que reconocer que en el proceso de inferencia se utilizará la información empírica para darle forma (por ejemplo, Geweke y Meese, 1981; Hendry *et al.*, 1984). Mediante un proceso de factorización condicional/marginal se puede escribir:

$$L(\theta) \propto \prod_{t=1}^{T} f\left[y_t \mid F_t, \theta \right], \tag{15}$$

con la constante de proporcionalidad arbitraria.[1] Sin embargo, si la distribución de z_t condicional en el pasado (G_{t-1}) depende de un vector de parámetros φ entonces la verosimilitud conjunta para θ y φ es

$$L_c(\theta, \varphi) = L(\theta) L_z(\varphi),$$

donde: $L_z(\varphi) \propto \prod_{t=1}^{T} f(z_t \mid G_{t-1}, \varphi)$

[1] Como está formulada en (15), la función de verosimilitud es condicional a la información inicial Go. La verosimilitud incondicional sería $L^*(\theta) = L(\theta) f(Go \mid \theta)$, donde $f(Go \mid \theta)$ representa la densidad incondicional de las condiciones iniciales. Aunque asintóticamente es indistinto usar $L(\theta)$ o $L^*(\theta)$, en pequeña muestra la pérdida de información por utilizar la verosimilitud condicional puede ser importante.

y para estar en capacidad de hacer inferencia respecto de los parámetros de interés θ a partir de la verosimilitud condicional $L(\theta)$ sin perder información, requerimos que las z_t sean débilmente exógenas para θ en el sentido de Engle *et al.* (1983). Esto requiere dos cosas: en primer lugar, que los momentos de y_t dado F_t estén bien especificados de forma que

$$E\left[u_t \mid F_t \right] = E\left[\epsilon_t^{(r)} \mid F_t \right] = 0$$

(i. e., que z_t sea exógena en el sentido usual), lo que es indispensable para congruencia; y que θ y φ no estén relacionados, de manera que $L_z(\varphi)$ no contenga información acerca de θ. De igual forma, se necesita que $L(\theta)$ esté captando información suficiente con respecto a todos los elementos de θ para que éstos puedan ser analizados en forma estadística. En otras palabras, requerimos que θ sea identificable, lo que se plasma en que la matriz de información

$$I(\theta) = -\lim_{T \to \infty} T^{-1} \; E \; \frac{\vartheta^2 L(\theta)}{\vartheta\theta \, \vartheta\theta'}$$

donde $L(\theta) = \ln L(\theta)$, sea positiva definida en una vecindad de θ^*. El valor θ^* es el valor pseudo-cierto de θ, es decir, donde converge el estimador máximo verosímil $\hat{\theta}$ $\left[\text{i. e., } L(\hat{\theta}) \geq L(\theta) \;\; \forall \, \theta \in \Theta \right]$ (ver White, 1982; Dowowitz y White, 1982) dentro de ciertas condiciones de regularidad que imponemos al proceso. Estas condiciones implican restricciones en cuanto a memoria y homogeneidad, así como condiciones de suavidad y acotamiento respecto de los momentos condicionales del modelo. Las condiciones de regularidad aseguran, además, la normalidad asintótica de $\hat{\theta}$ $\left\{ \text{i. e., de } T^{1/2}(\hat{\theta} - \theta^*) \right\}$. Aunque es posible realizar inferencias sensatas a partir de modelos incorrectamente especificados, el interés desde el punto de vista de evaluación se centra en calificar la bondad de la especificación –qué tan bien se puede pensar que el modelo aproxima al PGI– y con ello interesa calificar si $\theta^* = \theta_0$, el valor real de los parámetros. Esto dependerá de las relaciones entre el modelo y el PGI dadas en (7) y (14). Si los componentes de error en estas ecuaciones se anulan, entonces $\theta^* = \theta_0$. Cuando la media está correctamente especificada se tiene $u_t = \bar{u}_t$ y, con ello, β_0 cobra sentido como "valor real" de los parámetros. Esto elimina, al mismo tiempo, uno de los componentes de error de la relación (14) (el inducido por mala especificación de la media). Para los momentos de orden $2, 3, \ldots, N$ requerimos entonces que estén bien especificacios para que desaparezca el restante componente de error y con ello cada vector θ_r^0 tenga sentido como "valor real". Dado lo anterior, todo el vector θ_0 puede ser interpretado como "valor real" de los parámetros. El que la distribución D (.) sea incorrecta no afecta el sentido de θ_0, pero induciría errores en los momentos de orden más alto ($N+1, N+2, \ldots$). Aun cuando estos errores no tendrán efecto en la congruencia del estimador máximo verosímil $\hat{\theta}$, sí pueden afectar su eficiencia, y, lo que es más importante, pueden llevarnos a estimar incorrectamente la matriz de covarianzas de $\hat{\theta}$. Cuando esto sucede, estamos en riesgo de hacer inferencias incorrectas y llegar a conclusiones erróneas. Por ello es importante evaluar la forma de la distribución propuesta, y, en caso de hallar signos de su impropiedad, buscar un supuesto más adecuado o bien utilizar un estimador de la matriz de covarianzas que sea robusto a errores de especificación en la forma de D (.). Esto puede hacerse siguiendo el principio propuesto por Eicker (1967) y White (1980b).

IV. EVALUACIÓN ECONOMÉTRICA

Para evaluar el modelo desde un punto de vista econométrico son de central importancia las relaciones entre las innovaciones del modelo y del PGI dadas en (7) y (14). A su análisis se le ha dado el nombre de análisis de residuos (Pagan y Hall, 1983) y adición de variables (Pagan, 1984).

Consideremos la media en primer lugar. Una relación como (7) plantea una función de regresión en la que el componente sistemático está dado por el error de especificación $\bar{\mu}_t - \mu_t(\beta)$, y el no sistemático por las innovaciones $\bar{u}_t$ del PGI. Existen, sin embargo, dos problemas: *i*) las innovaciones del modelo (la variable dependiente de la regresión en (7)) no son observables, y *ii*) el error de especificación puede tomar muchas formas, paramétricas o no. El primer problema se resuelve utilizando residuos de la estimación de la media, es decir, $\hat{u}_t = y_t - \mu_t(\hat{\beta})$, a partir de (5). Para simplificar la exposición suponemos que la media ha sido especificada linealmente, $\mu_t(\beta) = x_t'\beta$ para $x_t \in F_t$, pero aclaremos que la extensión a funciones no lineales es trivial. Si a ambos lados de la igualdad en (7) sumamos $\hat{u}_t - u_t$ se tiene

$$\hat{u}_t = (\hat{u}_t - u_t) + \left[\bar{\mu}_t - \mu_t(\beta)\right] + \bar{u}_t,$$

Pero $\hat{u}_t - u_t = x_t'\left[\beta - \hat{\beta}\right] = x_t' \; b$, donde $b = \beta - \hat{\beta}$,

con lo que

$$\hat{u}_t = x'b + \left[\bar{\mu}_t - \mu_t(\beta)\right] + \bar{u}_t \,. \qquad (16)$$

Para atacar el segundo problema se plantean dos tipos de estrategias: *específica* y *general*, que se analizan a continuación revisando algunas de las principales pruebas en la bibliografía especializada.

1. *Evaluación de errores específicos en la media condicional*

Esta primera estrategia propone formas específicas para el componente de error, lo que presupone que existe un conocimiento *a priori* de direcciones en las que el modelo puede estar incorrectamente especificado, y, a la vez, implica que los mecanismos de evaluación posean una alta potencia en la dirección escogida y posiblemente (pero no necesariamente) baja potencia en otras direcciones.

Para seguir esta estrategia, el componente de error es parametrizado. En lo sucesivo suponemos linealidad para mantener el argumento simple, pero los mecanismos de evaluación aquí propuestos generalizan fácilmente el caso no lineal. La parametrización del componente de error es

$$\bar{\mu}_t - \mu_t(\beta) = s_t'\,\alpha \,, \qquad (17)$$

donde α es un vector de parámetros que, desde la hipótesis de que el modelo está correctamente especificado, sería tal que $\bar{\mu}_t - \mu_t(\beta_0) = 0$ (i. e., $\alpha = 0$). Sustituyendo (17) en (16) obtenemos

$$\hat{u}_t = x_t' b + s_t'\alpha + u_t \qquad (18)$$

es decir, una regresión de los residuos $\hat{u}_t$ sobre x_t (los regresores de la media) y s_t (las variables ocasionan el error de especificación). La prueba de $\alpha = 0$ es una prueba de error de especificación en la dirección prescrita. La inclusión de x_t entre los regresores en (18) es indispensable, pues de otra forma esta ecuación estaría incorrectamente especificada y esto implicaría que la matriz de covarianzas de α se estimase con error. La prueba de $\alpha = 0$ estaría sesgada en consecuencia (véase Pagan, 1984). La prueba de $\alpha = 0$ se hace con base en el estadístico usual, que se distribuye asintóticamente como x^2, con grados de libertad igual al número de elementos en α. Engle (1982b, 1984) ha mostrado que esta prueba es equivalente a la de multiplicadores de Lagrange.

Esta estrategia permite realizar pruebas mediante la selección del vector s_t, lo que equivale a hacer propuestas específicas acerca de errores de especificación.

Si el error de especificación consiste en que se han excluido variables relevantes de la regresión, dadas por el vector $w_t = (z_{tn+1}, \ldots, z_{tn+m})'$, se tendría $E[y_t' \mid F_t] = x_t'\beta + w_t\;\alpha$, de donde $u_t = y_t - x_t'\beta = w_t'\alpha + \mu_t$. Para evaluar la posibilidad de este error hacemos $s_t = w_t$. Es decir, las variables que se deben agregar en la regresión en (18) serían las que se piensa pudieran haber sido excluidas en forma incorrecta, y el número de grados de libertad para la prueba es la dimensión del vector w_t (el número de variables excluidas).

Supongamos que las variables relevantes han sido incluidas pero la dinámica de la ecuación no está captada adecuadamente. Esto sucederá si la regresión para la media del modelo es

$$y_t = \beta_1 y_{t-1} + \ldots + \beta_r y_{t-r} + z_t'\beta_{r+1} + z_{t-1}' \beta_{r+2} + \ldots + z_{t-s}' \beta_{r+s+1} + u_t,$$

mientras que la media del PGI está dada por

$$E\left[y_t \mid F_t\right] = \beta_1 y_{t-1} + \ldots + \beta_r y_{t-r} + z_t' \beta_{r+1} + \ldots + z_{t-s}' \beta_{r+s+1}$$
$$+ \alpha_1 y_{t-r-1} + \ldots + \alpha_m y_{t-r-m}$$
$$+ z'_{t-s-1} \alpha_{m+1} + \ldots + z_{t-s-q}' + \bar{u}_t,$$

Entonces,

$$u_t = \alpha_1 y_{t-r-1} + \ldots + \alpha_m y_{t-r-m}$$
$$+ z'_{t-s-1} \alpha_{m+1} + \ldots + z_{t-s-q}' + \bar{u}_t,$$

y las variables que se deben agregar en (18) para evaluar el error en la especificación dinámica son

$$s_t = \left[y_{t-r-1}, \ldots, y_{t-r-m}, z'_{t-s-1}, \ldots, z'_{t-s-q}\right]',$$

resultando en una prueba con $m + qn$ grados de libertad donde n es la dimensión de z_t.

Un caso particular de incorrecta especificación dinámica se presenta cuando las innovaciones del modelo están autocorrelacionadas, de suerte que

$$u_t = \beta_1 u_{t-1} + \ldots + \beta_m u_{t-m} + \bar{u}_t .$$

Evidentemente, las variables que se deben agregar en (18) serían en este caso $s_t = \left[u_{t-1}, \ldots, u_{t-m}\right]'$. Sin embargo, esta especificación de s_t es no observable y para hacerla operativa resulta necesario hacer uso de nueva cuenta de los residuos de la regresión. Al utilizar dichos residuos con la selección $s_t = \left[\hat{u}_{t-1}, \ldots, \hat{u}_{t-m}\right]'$ produce una prueba de autocorrelación de orden "m" y tiene "m" grados de libertad. La sustitución de residuos no afecta a la distribución y, más aún, para esta prueba se pueden omitir de la regresión en (18) los regresores x_t de la ecuación original (Breusch y Pagan, 1980). Así, la prueba se puede reducir a probar $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$ en la regresión

$$\hat{u}_t = \alpha_1 \hat{u}_{t-1} + \ldots + \alpha_m \hat{u}_{t-m} + \text{error}.$$

Otro aspecto importante de esta prueba es que si la autocorrelación en los errores se presenta en forma de media móvil (MA) en lugar de en forma autorregresiva, es decir,

$$u_t = \alpha_1 \bar{u}_{t-1} + \ldots + \alpha_m \bar{u}_{t-m} + \bar{u}_t,$$

la misma selección de s_t produce el mecanismo de evaluación adecuado y abarca un número de posibilidades de modelos compuestos (ARMA) entre el AR(m) y el MA(m) (Godfrey, 1978). En otras palabras, la contraparte "observable" de las innovaciones reales $\bar{u}_t$ está dada también por los residuos $\hat{u}_t$.

La prueba de cambio estructural propuesta por Chow (1960) también se puede considerar en esta perspectiva. En este caso la media del PGI es, efectivamente, una función lineal de x_t, pero los coeficientes pueden haber cambiado durante el periodo muestral. Así, la media del PGI es

$$E\left[y_t \mid F_t\right] = x_t' \beta, \qquad t = 1, \ldots, T_1,$$

mientras que

$$E\left[y_t \mid F_t\right] = x_t' \gamma, \qquad t = T_1 + 1, \ldots, T,$$

donde $1 < T_1 < T$. Para el primer subperiodo (sin pérdida de generalidad), el modelo es correcto y $u_t = \bar{u}_t$. Sin embargo, para el segundo periodo tenemos que $u_t = x_t' \gamma - x_t' \beta + \bar{u}_t = x_t' \alpha + \bar{u}_t$, donde $\alpha = \gamma - \beta$. Si ha habido cambio estructural $\gamma \ddagger \beta$ y, por lo tanto, $\alpha \ddagger 0$, mientras que, si no ha habido cambio estructural, $\gamma = \beta$ y en consecuencia $\alpha = 0$. Para hacer esta prueba, entonces, las variables s_t que se deben agregar en (18) son $s_t = 0$ para $t = 1, \ldots, T_1$, y $s_t = x_t$ para $t = T_1 + 1, \ldots, T$. El estadístico posee k grados de libertad (k es la dimensión de x_t) y se generaliza fácilmente a cambios estructurales parciales (agregando solamente un subvector de x_t al segundo subperiodo). La prueba requiere submuestras completas, de suerte que $T_1 > k$ y $T - T_1 > k$.

La evaluación del cambio estructural puede también hacerse desde una perspectiva predictiva estimando el modelo utilizando únicamente el primer subperiodo, y probando a partir de los errores de predicción resultantes para el segundo subperiodo. Salkever (1976) y Nicholls y Pagan (1984) han mostrado que esta prueba puede interpretarse con base en la relación modelo-PGI dada en (7), y, por lo tanto, se puede obtener de una regresión auxiliar como la planteada en (18). Para el segundo subperiodo tenemos

$$u_t = x_t'(\gamma - \beta) + \bar{u}_t = \alpha_t + \bar{u}_t,$$

donde $\alpha_t = x_t' (\gamma - \beta)$, de suerte que $\alpha_t = 0$ implica que $u_t = \bar{u}_t$. Así, se agrega un parámetro por cada observación en el segundo subperiodo resultando en s_t de dimensión $(T - T_1) \times 1$, formado por ceros excepto la posición $t - T$, que es igual a la unidad. Para el primer subperiodo $s_t = 0$ de dimensión $(T - T_1) \times 1$. La prueba tiene $T - T_1$ grados de libertad y no es necesario que $T - T_1 > k$.

Los errores de especificación considerados hasta ahora presentan la característica de que el modelo está anidado en el PGI (i. e., es un caso particular de éste). Es también posible que el modelo sea totalmente ajeno al PGI y no esté anidado en él. Si

$$E\left[y_t \mid F_t\right] = w_t' \alpha,$$

donde w_t es un conjunto de variables distinto de x_t, entonces

$$u_t = w_t' \gamma - x_t' \beta + \bar{u}_t,$$

o bien, procediendo como en (16).

$$\hat{u}_t = x_t' C + w_t' \alpha + \bar{u}_t,$$

donde $c = b - \beta = - \hat{\beta}$. La prueba de $\alpha = 0$ está vinculada con la prueba de hipótesis no anidadas propuesta por Atkinson (1971), y a la prueba N de Davidson y MacKinnon (1981). La prueba tiene tantos grados de libertad como la dimensión de w_t, se generaliza fácilmente al caso en que existan variables comunes en w_t y x_t, y tiene la ventaja de que no requiere la estimación del modelo alternativo. En términos de (18), la selección es $s_t = w_t$, pero nótese que el vector de coeficientes asociados con x_t es $- \hat{\beta}$, en lugar de $\beta - \hat{\beta}$. Esto es consecuencia de la estructura no anidada de las hipótesis, y es necesario incluir a x_t entre los regresores para que la prueba sea del tamaño adecuado. Davidson y MacKinnon (1981) también propusieron la prueba "J", que es similar a la anterior pero requiere que el modelo alternativo sea ajustado primero, se obtengan los valores ajustados $\hat{y}_t^* = w_t' \alpha$, y se considere.

$$\hat{u}_t = x_t' c + \alpha_0 \hat{y}_t^* + \left\{ \bar{u}_t + w_t' (\alpha - \alpha_0 \hat{\alpha}) \right\},$$

probando que $\alpha_0 = 0$. Es decir, se selecciona $s_t = \hat{y}_t^*$. Esta prueba tiene un grado de libertad únicamente y es, por lo tanto, más potente que la mencionada anteriormente (DasGupta y Perlman, 1974). Es equivalente a la prueba de Cox (1962) y Pesaran (1974) basada en la razón de verosimilitud centrada. El modelo alternativo puede ser no lineal y esto no altera la forma de la prueba pues solamente utiliza los valores ajustados. Estas ventajas son suficientes para evaluar hipótesis no anidadas con base en la prueba "J", aun al costo de requerir el ajuste separado (lineal o no lineal) del modelo alternativo.

Es posible que la selección de las variables x_t sea correcta, pero que la forma lineal dada al modelo sea inadecuada y que la media del PGI es en realidad.

$$E\left[y_t \mid F_t \right] = g(x_t; \alpha),$$

para alguna función $g(.)$ continua. La evaluación del supuesto de linealidad contra la alternativa específica de la forma funcional $g(x_t; \alpha)$ no es sino un caso particular de la prueba "J" en la que s_t es el valor ajustado del modelo no lineal alternativo.

2. *Evaluación general de errores en la media condicional*

La segunda estrategia consiste en no dar una forma específica al tipo de error, sino considerar alternativas muy generales que así puedan actuar como aproximadores para variadas formas del error de especificación. Estas pruebas pueden tener potencia en una gama más amplia de direcciones que las pruebas específicas, pero por regla general tendrán relativamente baja potencia en cada dirección particular. La hipótesis alternativa es simplemente la negación de la nula (i. e., H_0: el modelo es correcto *vs.* H_1: el modelo es incorrecto) y, en consecuencia, este tipo de pruebas es no constructiva. El objetivo de estas pruebas consiste en ampliar el espectro de la evaluación. En esta estrategia, el error de especificación no es parametrizado.

Probablemente la prueba más conocida de este estilo sea RESET (Ramsey, 1969; Ramsey y Schmidt, 1976) en la que se propone que un polimonio en los valores ajustados ($\hat{y}_t = x_t' \hat{\beta}$) puede actuar como aproximante de diversos errores de especificación. Así, aunque no se propone una alternativa, la evaluación se puede hacer a partir de una regresión auxiliar como (18), escogiendo $s_t = \hat{y}_t^2$ (RESET $-$ 2), $s_t = \left[\hat{y}_t^2, \hat{y}_t^9 \right]'$ (RESET $-$ 3), ... , $s_t = \left[\hat{y}_t^2, \ldots, \hat{y}_t^L \right]$ (RESET $-L$).

Un principio general que permite construir fácilmente pruebas sin utilizar información acerca de la hipótesis alternativa fue propuesto por Hausman (1978), y consiste en comparar dos estimadores distintos del vector de parámetros, desde la premisa de que ambos sean congruentes cuando el modelo es correcto. Para la media condicional, sean $\hat{\beta}$ y $\bar{\beta}$ los dos estimadores (v. gr., uno puede ser el de mínimos cuadrados y el otro un estimador instrumental). Como ambos son congruentes, pueden expresarse (asintóticamente) como

$$\hat{\beta} = \beta_0 + g_1 \left[\bar{\mu}_t - \mu_t (\beta_0); t = 1, \ldots, T \right],$$

$$\bar{\beta} = \beta_0 + g_2 \left[\bar{\mu}_t - \mu_t (\beta_0); t = 1, \ldots, T \right],$$

donde $g_1(0) = g_2(0) = 0$, y $g_1(x) \ddagger g_2(x)$ cuando $x \ddagger 0$ al menos en parte del rango de definición de las funciones (de los contrarios los estimadores serían iguales). Así, si se considera la diferencia entre los dos estimadores

$$q = \hat{\beta} - \bar{\beta} = g_1 - g_2,$$

tenemos, cuando el modelo es correcto ($\bar{\mu}_t = \mu_t$), $q = 0$ (asintóticamente). Pero si hay error de especificación, $q \ddagger 0$. Existe la posibilidad de que en la presencia de algunos errores se tenga que $g_1 = g_2$, y por ello la prueba puede ser incongruente (Holly, 1982). Para evaluar si q es o no distinto de cero, se requiere su matriz de covarianzas $V(q)$. Si $\hat{\beta}$ es el estimador máximo verosímil en la hipótesis nula, Hausman ha establecido que

$$V(q) = V(\bar{\beta}) - V(\hat{\beta}),$$

pero el principio no requiere que uno de los estimadores sea el máximo verosímil (Ruud, 1984). Así, el estadístico se calcula como $T\,q'\,V(q)^{-1}\,q$, y tiene k grados de libertad.

White (1980) propuso una prueba de este estilo, en la que el primer estimador es el de la regresión de y_t en x_t, y el segundo proviene de la regresión de $y_t/\hat{y}_t$ en $x_t/\hat{y}_t$, donde $\hat{y}_t$ son los valores ajustados de la primera regresión. Pruebas similares han sido propuestas por Utts (1982) y Plosser *et al.* (1982).

Hausman propuso comparar el estimador mínimo cuadrático con el mínimo cuadrático en dos etapas, lo que en realidad constituye una prueba que tiene una alternativa clara en mente: exogeneidad. La prueba se relaciona con la propuesta originalmente por Wu (1973) y, de hecho, Engle (1982b, 1984) ha mostrado que es la prueba de multiplicadores de Lagrange para exogeneidad.

Breusch y Godfrey (1986) han demostrado que las pruebas de transformación de variables se pueden llevar a cabo como pruebas de adición de variables. En particular la prueba de White descrita arriba se puede realizar seleccionando $s_t = x_t/\hat{y}_t^2$, mientras que la Hausman se puede llevar a cabo seleccionando $s_t = \hat{x}_t - x_t$, donde $\hat{x}_t$ son los valores ajustados de la regresión de x_t en los instrumentos (i. e., s_t serían los residuos de esta regresión). Esta relación está ya presente en el trabajo original de Hausman, quien presenta otras aplicaciones del principio.

Cabe resaltar que las regresiones en (18) producen resultados eficientes solo si las innovaciones son homoscedásticas $\bar{h}_t^{(2)}$ es constante $\forall t$. De otra forma hay que proponer un modelo para la heteroscedasticidad, ajustarlo (véase IV.4), y, posteriormente, utilizar los valores ajustados de la función de heteroscedasticidad para estimar (18) por el método de mínimos cuadrados ponderados. El coeficiente de determinación no centrado debe provenir de la regresión ponderada en este caso. Si sospechamos que existe heteroscedasticidad pero no hemos logrado un modelo satisfactorio para ella, nos queda como alternativa calcular el estadístico de prueba para $\alpha = 0$ en (18) utilizando la matriz de covarianzas robusta de White (1980b). Con esto se pierde potencia pero se asegura la propiedad de la inferencia.

Es importante notar también que el principio planteado en (18) es igualmente útil para efectuar pruebas multidireccionales como las propuestas por Bera y Jarque (1982). Basta agregar en forma simultánea los regresores correspondientes a distintos tipos de errores para emprender una evaluación conjunta.

3. *Evaluación de la varianza condicional*

Para diagnosticar los momentos de orden 2 y mayores se puede seguir exactamente el mismo principio, con base en la ecuación (14). Para simplificar el argumento, supóngase que la media ha sido evaluada en forma exhaustiva, de forma que se está dispuesto a ignorar el componente de error que proviene de la media del proceso. Con esto se renuncia a la posibilidad de detectar errores en la media que aparezca en momentos de orden superior (Pagan y Sabau, 1988; y Sabau, 1987 consideran esta posibilidad). Asimismo, hay que considerar que la regresión (13) no es factible puesto que u_t no es observable. La regresión operativa correspondiente utiliza residuos para quedar como

$$\hat{u}_t^{\,r} = h_t^{(r)}\,(\theta_r) + e_t^{(r)}\,,$$

donde $e_t^{(r)} = \epsilon_t^{(r)} + \left\{\hat{u}_t^{\,r} - u_t^{\,r}\right\}$. Si se supone linealidad, $h_t^{(r)}(\theta_r) = x_t^{(r)\,\prime}\,\theta_r, x_t^{(r)} \in F_t$, y si el error $\bar{h}_t^{(r)} - h_t^{(r)}$ es parametrizado también linealmente $\bar{h}_t^{(r)} - \bar{h}_t^{\,r}\,(\theta) = s_t^{(r)\prime}\,\alpha, s_t^{(r)} \in F_t$ entonces se obtiene de (14)

$$\hat{e}_t^{(r)} = x_t^{(r)\,\prime}\,c_r + s_t^{(r)}\,\alpha_r + e_t^{(r)}\,, \tag{19}$$

donde $\hat{e}_t^{(r)} = \hat{u}_t^r - x^{(r)\,\prime}\,\hat{\theta}_r$, $c_r = \hat{\theta}_r - \theta_r$, y $\bar{e}_t^{(r)} = \bar{\epsilon}_t^{(r)} + \left\{\hat{u}^{\,r} - u_t^r\right\}$, que plantea un criterio para evaluar el r-ésimo momento de carácter similar a (18), pero en el que hay que tomar en cuenta el efecto que $\left\{\hat{u}_t^r - u_t^r\right\}$ pudiese tener en cuenta la distribución de los estadísticos de prueba.

Para mantener el argumento simple, supongamos que se ha supuesto normalidad y homoscedasticidad, de modo que el modelo es $y_t \mid F_t \sim N\,\left[x_t'\,\beta, \sigma^2\right]$.

Para evaluar la varianza condicional —el supuesto de homoscedasticidad—, (19) se convierte para $r = 2$ en

$$\hat{e}_t^{(2)} = c_2 + s_t^{(2)\prime}\,\alpha_2 + \bar{e}_t^{(2)}, \tag{20}$$

donde

$$\hat{e}_t^{(2)} = \hat{u}_t^2 - \hat{\sigma}^2 \text{ y } \bar{e}_t^{(2)} = \epsilon_t^{(2)} + \left\{\hat{u}_t^2 - u_t^2\right\}$$

Pagan y Hall (1983) y Sabau (1987) han mostrado que el término $\left\{\hat{u}_t^2 - u_t^2\right\}$ no tiene efecto en la distribución de estimadores y puede ser, por lo tanto, ignorado. La selección de $s_t^{(2)}$ propone la forma para la posible heteroscedasticidad y se puede aplicar el principio de calcular el estadístico como TR_0^2 de la regresión auxiliar en (20), cuya distribución es asintóticamente x^2 con grados de libertad igual al número de elementos en α_r (Engle, 1982a; Kraft y Engle, 1982; Pagan y Sabau, 1988). Además, la prueba es equivalente a la de multiplicadores de Lagrange (Breusch y Pagan, 1979).

La especificación de la función de heteroscedaticidad es más compleja que la especificación de la media condicional, ya que, en general, se tiene menos intuición acerca del comportamiento de la varianza. Existen, sin embargo, algunas formas típicas que vale la pena explorar. Éstas podemos clasificarlas en estáticas y dinámicas.

Las formas estáticas proponen que la varianza es una función de la media o de las variables que afectan la media. Así aparece el modelo de heteroscedasticidad simple

$$h_t^{(2)} = \sigma^2 + \bar{x}_t' \alpha ,$$

donde $\bar{x}_t$ es el vector de regresores de la media, salvo que se excluye la constante en caso de que la haya. Para este caso $s_t^{(2)} = \bar{x}_t$, y el principio se extiende fácilmente a que $s_t^{(2)}$ incluya sólo un subconjunto de las variables en x_t o bien otras variables que no estén en x_t.

Prais y Houthakker (1955) y Amemiya (1973) han propuesto un modelo en que la varianza es lineal en el cuadrado de la media. Incluyendo una constante para que el modelo tenga como caso particular al de homoscedasticidad, se tiene el modelo de Amemiya generalizado,

$$h_t^{(2)} = \sigma^2 + \alpha \mu_t ,$$

para el que se selecciona $s_t^{(2)} = \bar{e} y_t^2$, utilizando los valores ajustados de la regresión para estimar la media. Una propuesta similar proviene de generalizar el modelo Poisson (Hausman *et al.*, 1984; Gourieroux *et al.* 1985b) para que incluya varianza constante como caso particular (Cameron y Trivedi, 1985, 1986). Así tenemos el modelo Poisson generalizado,

$$h_t^{(2)} = \sigma^2 + \alpha \mu_t,$$

para el que se selecciona $s_t^{(2)} = \hat{y}_t$.

Los grados de libertad con el modelo de heteroscedasticidad simple como alternativa son el número de elementos en $\bar{x}_t$, mientras que el Poisson y Amemiya generalizados generan pruebas con un grado de libertad. Cabe aclarar que estos modelos serán en realidad dinámicos cuando la media lo sea (i. e., si x_t incluye rezagos de y_t).

El modelo dinámico de mayor interés por sus propiedades teóricas y su éxito en aplicaciones empíricas, principalmente con variables financieras, es el modelo ARCH (Engle, 1982a). Este modelo parte de sustentar que la varianza de los pronósticos para la media depende del éxito que tengan los mismos, de suerte que tenderá a reducirse luego de hacer un buen pronóstico, y a incrementarse después de realizar uno malo. En el modelo ARCH (q) la memoria que afecta a la varianza dura "q" periodos, y se tiene:

$$h^{(2)} = \sigma^2 + \sum_{j=1}^{q} \alpha_j u_{t-j}^2 ,$$

con lo que se selecciona $s_t^{(2)} = \left[\hat{u}_{t-1}^2, \ldots, \hat{u}_{t-q}^2\right]'$, y la prueba tiene q grados de libertad. Para lograr una memoria larga manteniendo un número pequeño de parámetros, Bollerslev (1986) ha generalizado el modelo ARCH al GARCH (p, q), dado por

$$h_j^{(2)} = \sigma^2 + \sum_{j=1}^{p} \alpha_j h_{t-j}^{(2)} + \sum_{j=1}^{q} \alpha_{p+j} u_{t-j}^2 ,$$

cuya función para la varianza es similar a la de los procesos ARMA para la media (Diebold, 1986; Sabau, 1988). De igual forma que la prueba de autocorrelación en la media abarca como alternativa a procesos AR, MA y ARMA, la prueba contra la alternativa ARCH incluye a GARCH.

Otro modelo dinámico que ha tenido éxito en aplicaciones con series de tiempo es el modelo regresivo (Weiss, 1984), que está relacionado con el modelo ARCH pero impone un menor grado de simetría en la varianza. El modelo de orden "q" es

$$h_t^{(2)} = \sigma^2 + \sum_{j=1}^{q} \alpha_j y_{t-j}^2 ,$$

para el que seleccionamos $s_t^{(2)} = \left[y_{t-1}^2, \ldots, y_{t-q}^2\right]'$, y la prueba tiene "$q$" grados de libertad.

Desde luego, la heteroscedasticidad puede ser más compleja que las propuestas anteriores. Por ejemplo, podría combinar varios de los efectos descritos. Al igual que en el diagnóstico de la media, podemos abrir esta posibilidad incluyendo en $s_t^{(2)}$ diferentes propuestas en forma simultánea. Con el propósito de diagnosticar homoscedasticidad, sin embargo, parece suficiente el considerar cada propuesta por separado. Es pertinente hacer notar, igualmente, que en los modelos Amemiya, Poisson y ARCH la varianza depende del vector β mediante μ_t y u_{t-j}^2. Esto puede ignorarse para el diagnóstico de homoscedasticidad, pero tendría implicaciones si se desea modelar la varianza, ya que existe información adicional que se puede utilizar acerca de β.

Cuando se rechaza de hecho el supuesto de homoscedasticidad hay que considerar los puntos del párrafo anterior al modelar la varianza, y al hacer esto es posible hacer uso de los diagnósticos que anteriormente se presentaron para la media, aplicándolos ahora a (19) (Sabau, 1988).

4. *Evaluación de momentos más altos*

En el caso de normalidad, la media y la varianza determinan la forma de la distribución, y, por ello, no pueden ser utilizadas para evaluar este supuesto. Dadas la media y la varianza, los momentos de orden más alto quedan determinados en función de ellos y pueden utilizarse para diagnosticar normalidad. Cualquier momento de orden 3 y superior puede ser utilizado para este fin, pero nos limitamos al tercero y cuarto momentos que son los de más fácil interpretación y que enmarcan la distribución dentro de la familia Pearson, como en Bera y Jarque (1982). Para el coeficiente de simetría, $h_t^{(9)} = 0$ bajo normalidad, con lo que (19) se convierte en

$$\hat{u}_t^9 = s_t^{(9)\prime} \alpha_9 + \bar{e}_t^{(9)} , \tag{21}$$

donde $\bar{e}_t^{(9)} = \bar{\epsilon}_t^{(9)} + \left\{\hat{u}_t^9 - u_t^9\right\}$. Si la hipótesis de homoscedasticidad se ha mantenido, no parece demasiado restrictivo el pensar que las desviaciones de simetría se mantuviesen constantes en el tiempo. Esto simplifica el modelo en (21) haciendo $s_t^{(9)} = 1 \quad \forall t$, con lo que obtenemos la regresión

$$\hat{u}_t^9 = \alpha_9 + \bar{e}_t^{(9)}. \tag{22}$$

Si aplicamos mínimos cuadrados a esta ecuación, el estimador de α_9 es, evidentemente, $\hat{\alpha}_9 = T^{-1} \Sigma \hat{u}_t^9$. El problema con esta regresión radica en que el término $\{\hat{u}_t^9 - u_t^9\}$ tiene un efecto que no desaparece de la distribución de α^9 al crecer la muestra. Por lo tanto, basarse en la "t" de esta regresión resulta en inferencias incorrectas. Esto puede corregirse fácilmente ya que la varianza de $T^{1/2}\alpha^9$ es estimada de forma congruente por $\hat{V}(\hat{\alpha}_9) = 6\hat{\sigma}^6$, donde $\hat{\sigma}^2$ es el estimador de σ^2 obtenido de la regresión en la media Pagan y Hall, 1983: $\left[\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2\right]$

Con esto, el estadístico

$$s = \frac{T\ \hat{\alpha}_9^2}{6\ \hat{\sigma}^6} \tag{23}$$

es el adecuado para probar la hipótesis de simetría, y tiene una distribución x^2 con un grado de libertad desde la hipótesis nula.

Para el coeficiente de Kurtosis tenemos que $h_t^{(4)} = 3\ \sigma^4$ dentro de normalidad y homoscedasticidad. La ecuación (19) queda como

$$\hat{e}_t^{(4)} = c_4 + s^{(4)\prime}\ \alpha_4 + \bar{e}_t^{(4)} \tag{24}$$

donde $\hat{e}_t^{(4)}$ son los residuos de la regresión de $\hat{u}_t^4$ en una constante, o pueden computarse equivalentemente desde el punto de vista asintótico como

$$\hat{e}_t^{(4)} = \hat{u}_t^4 - 3\hat{\sigma}^4 ; \text{y}\ \bar{e}_t^{(4)} = e_t^{(4)} + \left\{\hat{u}_t^4 - u_t^4\right\}.$$

Nuevamente, dada homoscedasticidad, no parece particularmente restrictivo considerar kurtoris constante, con lo que hacemos $s_t^{(4)} = 1$ y (24) queda en

$$\hat{e}_t^4 = \alpha_4 + \bar{e}_t^{(4)}, \tag{25}$$

donde $\alpha_4 = c_4 + \alpha_4$ son agregadas para evitar multicolinealidad. Como c_4 converge a cero por definición, la hipótesis de mesokurtosis (kurtosis igual a la de la normal) se traduce en $\alpha_4 = 0$. En este caso el término en $\left\{\hat{u}_t^4 - u_t^4\right\}$ no afecta la distribución asintótica y, por lo tanto, el estadístico para $\alpha_4 = 0$ obtenido de (25) es adecuado (Pagan y Hall, 1983; Sabau, 1987). Una forma asintóticamente equivalente de calcular el estadístico se puede obtener utilizando nuevamente el estimador de σ^2 obtenido de la regresión en la media, y es

$$k = \frac{T\hat{\alpha}_4^2}{24\ \hat{\sigma}^8} \tag{26}$$

que es también una x^2 con un grado de libertad en la hipó-tesis nula. Además, es una evidencia que k y s son asintóticamente independientes, con lo que $k + s \overset{a}{\sim} x_2^2$ es un estadístico que evalúa conjuntamente simetría y mesokurtosis. Jarque y Bera (1981) han mostrado que este es el estadístico de multiplicador de Lagrange para probar normalidad contra la alternativa de la familia Pearson, y el resultado ha sido extendido en Sabau (1988) a contextos heteroscedásticos.

V. ALGUNOS PROBLEMAS POR RESOLVER

En la sección anterior se han presentado un conjunto de pruebas de diagnóstico estadístico que incluyen diferentes alternativas y abarcan los cuatro primeros momentos del modelo condicional. La búsqueda de modelos razonables lleva a la aplicación de todo este conjunto de pruebas, lo cual se ha hecho con los modelos presentados en la Revista, para aceptar solamente los que muestren un grado de satisfacción razonables de estos criterios. Este mecanismo de evaluación representa un gran avance sobre el tipo de práctica econométrica de fines de los años setenta y principios de los ochenta, pero deja aún una serie de cuestiones a las que es necesario referirse. En primer lugar, tenemos aspectos de corte técnico tales como la determinación de niveles de significancia y el problema que representa el trabajo con aproximaciones asintóticas en muestras pequeñas. En segundo lugar, está la comparación del modelo con experiencias anteriores de modelística del mismo fenómeno para evaluar propiamente la relevancia de la nueva propuesta en la explicación de la realidad. Finalmente, está el problema más de fondo de interpretación de la prueba de hipótesis –basada en un argumento frecuentista– con datos no experimentales.

La determinación del nivel global de significancia ha recibido la atención de econometristas en general y Savin (1984) resume la literatura, que concluye con el establecimiento de cotas con base en sucesiones Bonferroni. Desde el punto de vista metodológico, el uso de análisis de residuos puede simplificar este hecho y permitir obtener el nivel exacto de significancia asociado a una prueba. Este hecho fue utilizado por Jarque y Bera (1981) y Bera y Jarque (1982) para la construcción de pruebas multidireccionales en el modelo lineal clásico, mientras que Bera y McKenzie (1987) analizan en forma más general las condiciones de independencia de estadísticos de multiplicadores de Lagrange. La obtención de un nivel global de significancia es simple siguiendo la idea de análisis de residuos. Por ejemplo, todas las pruebas deseadas para la media pueden ser especificadas de forma conjunta en (17), con lo que la regresión en (18) produciría la prueba de todas ellas hecha al nivel de significancia deseado. Así, si se va a diagnosticar al modelo en L direcciones distintas representadas por los vectores de variables adicionales s_{t1}, $s_{t2}, \ldots, s_{tL}$, respectivamente, éstas se acumulan en el vector $s_t = (s'_{t1}, \ldots, s'_{tL})'$ que se debe incluir en la regresión (18). Hay que tener cuidado para evitar que la multicolinealidad se vuelva extrema, lo que vendría a ser equivalente a dar cuenta de la dependencia entre las pruebas individuales. El nivel de dependencia entre pruebas se dará no solamente en función de la relación teórica de las proposiciones, sino que estará también condicionado por el tamaño de la muestra, y es posible que la dimensionalidad de s_t tenga que ser reducida. Esto tiene la virtud de mostrar la capacidad misma de la información disponible para diagnosticar. La prueba global no es informativa en caso de rechazo y las pruebas individuales deben aportar en este sentido las direcciones para continuar el ejercicio de especificación.

El mismo principio puede aplicarse a las ecuaciones de los momentos más altos y, finalmente, para obtener una prueba global en todos los momentos deseados, lo único que hay que hacer es considerar a las ecuaciones para los distintos momentos en forma conjunta como un sistema de ecuaciones aparentemente no relacionadas (SURE, Zellner, 1962; Sabau, 1988b).

El uso de aproximaciones asintóticas es sin duda un problema en muestras pequeñas como las que hemos podido utilizar en nuestros trabajos. El principal problema que esto ocasiona es el de la diferencia entre los niveles nominal y real de significancia, generando un riesgo de sobre-rechazo de hipótesis correctas. Como herramienta analítica no existe, desgraciadamente, una teoría alternativa general al uso de aproximaciones asintóticas. La teoría de la inferencia exacta que se ha venido desarrollando en años recientes (e. g., Phillips, 1984), dista todavía de ser aplicable. Para reducir el problema, se han utilizado las formas más robustas de los estadísticos de prueba: las basadas en coeficientes de determinación (Koenker, 1983). Esto no garantiza la cercanía entre los niveles real y nominal de significancia, y la salida que vemos a futuro es la de utilizar técnicas intensivas en el uso de la computadora con el fin de evaluar los resultados. Entre éstas está la simulación y el posible uso de superficies de respuestas (Hendry, 1984), el *bootstrap* (Efron, 1979), y la exploración exhaustiva de la función de verosimilitud a partir de integración Monte-Carlo (Geweke, 1987, 1988). Respecto de la potencia, ésta depende de la cantidad de información utilizada y el uso de pruebas potentes como los multiplicadores de Lagrange representa el mejor esfuerzo.

En el presente número de *Economía Mexicana* no hemos realizado pruebas de "englobamiento" (*encompassing*, Mizon, 1984), que están estrechamente vinculadas con la comparación de modelos no anidados (Mizon y Richard, 1987). Al proponer un nuevo modelo acerca de uno o un conjunto de fenómenos, debemos argumentar a nivel teórico en qué medida el nuevo modelo es diferente y/o representa algún avance sobre esfuerzos modelísticos anteriores. La idea de 'englobamiento' pretende dar un argumento similar a nivel de la coherencia teórico-empírica de los modelos, proporcionando pruebas estadísticas para diferentes formas de avance en el modelo econométrico. Consideramos de importancia que empecemos a hacer ejercicios de este corte en nuestro medio para establecer una base sólida de discusión sobre la cual desarrollar el futuro de la economía aplicada. Esto nos llevaría al mismo tiempo a compatibilizar bases de información empírica y hacerlas disponibles para la profesión en general.

Finalmente, queda el problema más delicado, que es el de la interpretación de la inferencia. En este sentido seguimos considerando que las interpretaciones de corte bayesiano son interesantes, siempre que permitan mantener la operatividad de las técnicas desarrolladas en la escuela estructural (para este punto ver Ruprah y Sabau [1984]).

México, D. F., 1988; revisado en 1990.

REFERENCIAS BIBLIOGRÁFICAS

Amemiya, T. (1973), "Regression Analysis when the variance of the dependent variable is proportional to the square of its expectation", *Journal of the American Statistical Association*, 68, pp. 928-934.

Bera, A. K. y C. M. Jarque (1982), "Model specification tests: a simultaneous approach", *Journal of Econometrics*, núm. 20, pp. 59-82.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, núm. 31, pp. 307-327.

Box, G. E. P., y G. M. Jenkins (1971), *Time Series Analysis, Forecasting and Control*, Holden-Day.

Breusch, T. S., y A. R. Pagan (1979), "A simple test for heteroscedasticity and random coefficient variation", *Econometrica*, núm. 47, pp. 1287-1294.

——— (1980), "The Lagrange multiplier test and its applications to model specification in econometrics", *Review of Economic Studies*, núm. 47, pp. 239-253.

Breusch, T. S., y L. G. Godfrey(1986), "Data transformation tests", *The Economic Journal*, núm. 96, pp. 47-58.

Cameron, A. C., y P. K. Trivedi (1985), "Regression based tests for overdispersion", Informe Técnico núm. 9, *Stanford University Econometric Workshop*.

——— (1986), "Econometric models based on count data: comparisons and applications of some estimators and tests", *Journal of Applied Econometrics*, núm. 1, pp. 29-53.

Chow, G. C. (1960), "Tests of equality between sets of coefficients in two linear regressions", *Econometrica*, núm. 28, pp. 591-605.

Dasgupta, S., y M. D. Perlman (1974), "Power of the noncentral F-test: effect of additional variates on Hotelling's T^2-test", *Journal of the American Statistical Associaton*, núm. 69, pp. 174-180.

Davidson, R. y J. G. MacKinnon (1981), "Several tests for model specification in the presence of alternative hypotheses", *Econometrica*, núm. 49, pp. 781-793.

Dhrymes, P. J. (1970), *Econometrics*, Harper and Row.

Diebold, F. H. (1986). Weiss, A. A. (1984), ARMA models with ARCH errors, *Journal of Time Series Analysis*, 5, pp. 129-143.

Doan, T., R. Litterman, y C. Sims (1984), "Forecasting and conditional projection using realistic prior distributions", *Econometric Reviews*, núm. 3, pp. 1-100.

Domowitz, I. y H. White (1982), "Miespecified models with

dependent observations", *Journal of Econometrics*, núm. 20, pp. 35-58.
Eicker, F. (1967), "Limit theorems for regressions with unequal and dependent error", *Proceedings of the Fifth Berkeley Simposium on Mathematical Statistics and Probability*, vol. I, University of California Press.
Engle, R. F. (1982a), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, núm. 50, pp. 987-1007.
——— (1982b), "A general approach to Lagrange multiplier model diagnostics", *Journal of Econometrics*, núm. 20, pp. 83-104.
——— (1984), "Wald, likelihood ratio, and Lagrange multiplier tests in econometrics", en Z. Griliches y M. D. Intrilligator (comps.), *Handbook of Econometrics*, vol. II Elsevier Science Publishers.
Engle, R. F., D. F. Hendry, y J. F. Richard (1983), "Exogeneity", *Económetrica*, núm. 51, pp. 277-304.
Geweke, J. (1987), "Exact inference in models with autoregressive conditional heteroscedasticity", *Duke University*, mimeografiado.
Geweke, J., y R. Meese (1981), "Estimating regression models of finite but unknown order", *International Economic Review*, núm. 22, pp. 55-70.
Godfrey, L. G. (1978), "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables", *Econometrica*, núm. 46, pp. 1303-1310.
Gourieroux, C., A. Monfort y A. Trognon (1984a), "Pseudo maximum likelihood methods: theory", *Económetrica*, núm. 52, pp. 681-700.
——— (1984b), "Pseudo maximum likelihood methods: applications to Poisson models", *Econometrica*, núm. 52, pp. 701-720.
Hausman, J. A. (1978), "Specification tests in econometrics", *Económetrica*, núm. 46, pp. 1251-1271.
Hausman, J. A., B. H. Hall, y Z. Griliches (1984), "Econometric models for count data with an application to the patents R and D relationship", *Econometrica*, núm. 52, pp. 909-938.
Hendry, D. F., A. R. Pagan y J. D. Sargan (1984), "Dynamic Specification", en Z. Griliches y M. D. Intrilligator (comps.), *Handbook of Econòmetrics*, vol. 2, North Holland.
Hendry, D. F., y J. F. Richard (1982), "On the formulation of empirical models in dynamic econometrics", *Journal of Econometrics*, núm. 20, pp. 3-33.
——— (1983), "The econometric analysis of economic time series", *International Statistical Review*, núm. 51, pp. 111-148.
Holly, A. (1982), "A remark on Hausman's specification test", *Econometrica*, núm. 50, pp. 749-759.
Jarque, C. M., y A. K. Bera (1980), "Efficient tests for normality, homosdedasticity and serial independence off regression residuals", *Economics Letters*, núm. 6, pp. 255-259.
Johnston, J. (1972), *Econometric Methods*, McGraw-Hill.
Kraft, D. F., y R. F. Engle (1982), "Autoregressive conditional heteroscedasticity in multiple time series models", UCSD *Discussion Paper*, núm. 82-23.
Leamer, E. E. (1978), *Specification Searches*, John Wiley.
Pagan, A. R. (1984), "Model evaluation by variable addition", en D. F. Hendry y K. F. Wallis (comps.) *Econometrics and Quantitative Economics*, Basil Blackwell.
Pagan, A. R., y A. L. Hall (1983), "Diagnostic tests as residual analysis", *Econometric Reviews*, núm. 2, pp. 159-218 (Apéndice al ANU *Working Papers in Economics and Econometrics*, núm. 87).
Pagan, A. R., y D. F. Nicholls (1984), "Estimating predictions, prediction errors and their standard deviations using constructed variables", *Journal of Econometrics*, núm. 24, pp. 293-310.
Pagan, A. R. y H. Sabau (1987), "On the inconsistency of the MLE in certain heteroskedastic regression models", *University of Rochester y* ANU, mimeografiado.
——— (1988), "Consistency tests for heteroskedastic and risk models", *University of Rochester y* ANU, mimeografiado.
Plosser, C. I., G. W. Schwert y H. White (1982), "Differencing as a test of specification", *International Economic Review*, núm. 23, pp. 535-552.
Prais, S. J. y H. S. Houthakker (1955), *The Analysis of Family Budgets*, Cambridge University Press.
Ramsey, J. B. (1969), "Tests for especification errors in classical linear least-equares regression analysis", *Journal of the Royal Statistical Society*, Series B, núm. 31, pp. 350-371.
Ramsey, J. B., y P. Schmidt (1976), "Some further results on the use of OLS and BLUS residuals in specification error tests", *Journal of the American Statistical Association*, núm. 71, pp. 389-390.
Ruud, P. A. (1984), "Tests of specification in econometrics", *Econometric Reviews*, núm. 3, pp. 211-242.
Sabau, H. C. (1987), "The structure of GMM and ML estimators in conditionally heteroskedastic models", ANU *Working Papers in Economics and Econometrics*, núm. 153.
——— (1988), "Some Theworetical Aspects of Econometric Inference with Heteroskedastic Models", Tesis Doctoral, Australian National University.
Salkever, D. S. (1976), "The use of dummy variables to compute predictions, prediction errors, and confidence intervals", *Journal of Econometrics*, núm. 4, 393-397.
Sims, C. A. (1980), "Macroeconomics and reality", *Econometrica*, núm. 48, pp. 1-48.
Spanos, A. (1986), *Statistical Foundations of Econometric Modelling*, Cambridge University Press.
Theil, H. (1971), *Principles of Econometrics*, John Wiley.
Weiss, A. A. (1984), "Modelling the persistence of conditional variances: a comment", *Econometric Reviews*, núm. 5, pp. 51-56.
White, H. (1980a), "Using least squares to approximate unknown regression functions", *International Economic Review*, núm. 21, pp. 149-170.
——— (1980b), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity", *Económetrica*, núm. 48, pp. 817-838.
——— (1982), "Maximum likelihood estimation of misspecified models", *Econometrica*, núm. 50, pp. 1-25.
Wu, D. (1973), "Alternative tests of independence between stochastic regressors and disturbances", *Econometrica*, núm. 41, pp. 733-750.
Zellner, A. (1962), "An efficent method of estimating seemingly unrelated regressions and tests for aggregation bias", *Journal of the American Statistical Associaton*, núm. 57, pp. 348-368.