

METODOLOGIA DE LA EVALUACION DE MODELOS ECONOMETRICOS UNIECUACIONALES DE REGRESION LINEAL MULTIPLE

Rafael E. Solórzano¹

Código JEL: C19, C20, C39

RESUMEN

En este trabajo se presentan los principales aspectos de carácter metodológico que deben considerarse al evaluar un modelo econométrico uniecuacional, estimado mediante técnicas de regresión lineal múltiple. La construcción y estimación de estos modelos constituye un recurso de muy generalizada aplicación entre los economistas y demás estudiosos de las ciencias sociales, ya que a través de ellos se suele establecer algún tipo de relación de dependencia entre variables, a partir de la cual se suministran elementos explicativos para proposiciones de carácter teórico. La evaluación del modelo estimado tiene el propósito de validar empíricamente la relación teórica involucrada, y de analizar la capacidad explicativa y el poder predictivo de las estimaciones. Dicha evaluación debe ser hecha en tres diferentes planos: Primero, una evaluación desde el punto de vista de los principios teóricos que han sido incorporados en la especificación del modelo, lo cual se hace analizando y comparando los signos y magnitudes de los coeficientes de regresión del modelo, con los de los coeficientes de regresión estimados. Segundo, una evaluación de la significación estadística de los parámetros estimados, considerados tanto individualmente, como formando parte conjunta de la ecuación de regresión. Y, finalmente, una evaluación econométrica propiamente dicha de las estimaciones, dirigida a verificar si en el proceso de estimación, se respetaron los supuestos clásicos de la regresión lineal múltiple,

¹ Economista, Master in Economics, Doctor en Ciencias Económicas. Profesor Titular Jubilado de Economía, Universidades de los Andes y del Zulia, Mérida y Maracaibo, Venezuela, respectivamente. Miembro del Consejo Académico del Doctorado en Ciencias Económicas, Universidad del Zulia. Individuo de Número de la Academia de Mérida (Sillón 11) y Miembro Correspondiente Nacional de la Academia de Ciencias Económicas del Estado Zulia. E-mail: resolorzano@gmail.com. Teléfonos: +58-274-2631912 y +58-412-7642313.

a saber, la no existencia de colinealidad entre las variables independientes del modelo, la ausencia de autocorrelación estadísticamente significativa entre las perturbaciones aleatorias, y de correlación lineal entre estas y las variables independientes, y, la existencia de varianzas constantes de las perturbaciones aleatorias. El presente trabajo hace énfasis en este tercer tipo de evaluación y en las más importantes pruebas de significación que han sido diseñadas para detectar la presencia significativa de alguno de esos problemas, los cuales afectan negativamente tanto la capacidad explicativa como el poder predictivo de las estimaciones econométricas.

Palabras clave: Autocorrelación, Multicolinealidad, Heteroscedasticidad.

ABSTRACT

This paper presents the main methodological aspects to consider when evaluating an econometric model estimated by the use of multiple linear regression techniques for single-equation econometric model. The construction and estimation of these models is a resource of widespread application among economists and other scholars in the social sciences, because through them is usually set some kind of dependency relationship between variables, from which are supplied explanatory elements for theoretical propositions. The evaluation of the estimated model aims to empirically validate the theoretical relationship involved, and to analyze the explanatory and predictive power of the estimates. This evaluation must be made in three different levels: First, an evaluation from the standpoint of theoretical principles have been incorporated into the model specification, which is done by analyzing and comparing the signs and magnitudes of the regression coefficients of the model, with the estimated regression coefficients. Second, an evaluation of the statistical significance of the estimated parameters, considered individually, and as part of the joint regression equation. And finally, a full-fledged econometric assessment of the estimates, addressed to verify whether in the estimation process, the classical assumptions of multiple linear regression were respected, namely, the absence of collinearity between the independent variables of the model, no statistically significant autocorrelation between random disturbances, and linear correlation between these and the independent variables; and the existence of

constant variances of the random perturbations. This paper emphasizes this third type of assessment and the major significance tests that have been designed to detect the significant presence of any of these problems, which negatively affect both the explanatory capability and the predictive power of the econometric estimates .

Keywords: Autocorrelation, multicollinearity, heteroskedasticity.

La Econometría como disciplina científica es una rama relativamente reciente de la Economía. La más reciente, quizás. Y aunque, en su intento de cuantificar los eventos y procesos económicos se le puedan asociar antecedentes tan remotos como la publicación de la *Political Arithmetic*, de Sir William Petty en 1690, no fue sino hasta 1930 cuando verdaderamente se produjo su nacimiento *formal* a raíz de la fundación de la Sociedad Econométrica por parte de Ragnar Frisch, Irving Fisher y Charles Roos, y de la paralela creación y puesta en circulación de la revista científica *Econometrica* en 1933 como órgano divulgativo oficial de la Sociedad, iniciativas ambas adoptadas bajo el mecenazgo del millonario financista norteamericano Alfred Cowles III y de la comisión científica especializada que para tales efectos él creó en 1932, la *Cowles Commission for Research in Economics*, institución sin fines de lucro fundada y emplazada originalmente en Colorado Springs (Colorado) y posteriormente mudada en 1939 a la Universidad de Chicago (Illinois), y a la Universidad de Yale (New Haven, Connecticut) en 1955, donde funciona actualmente y en donde cambió su nombre por el de Fundación Cowles. Como ninguna otra rama de la Economía, durante los casi noventa años de su existencia oficial, la Econometría ha experimentado un progreso impresionante básicamente por los beneficios que brinda para el desarrollo de la teoría económica, de los cuales, dos son dignos de destacar. Por un lado, la posibilidad de medir y estimar los parámetros económicos, y por otra parte, la de evaluar si dichas estimaciones son aceptables o no como estadísticamente significativas, todo lo cual puede ayudar de forma importante a clarificar las relaciones de dependencia entre variables económicas y los procesos decisorios en la economía, en la medida en que permite una mejor comprensión del futuro, ya que la mayoría de dichos procesos dependen en cierto grado de las percepciones y expectativas que los agentes económicos responsables de esas decisiones mantienen con respecto al

curso futuro de los acontecimientos. Por esta razón algunos afirman que las posibilidades de predicción han sido la principal contribución de la Econometría al progreso de la Teoría Económica (Dutta, 1982). Sin embargo, se debe ser cuidadoso de no sobrevalorar las potencialidades que para el control del mundo real puede ofrecer la posibilidad de predecir que brinda la econometría, ya que hay muchos elementos de subjetividad involucrados en la construcción de modelos econométricos que hacen que los resultados de estimar y trabajar estos modelos y su posible uso como instrumentos de predicción dependa también en buena medida de la manera cómo el formulador del modelo decidió construirlo en el momento de su especificación. En general, se suele aceptar que la economía teórica y la aplicada pueden utilizar los aportes de la Econometría en los siguientes usos (Gujarati, 2003; Loría, 2007):

1. *En el análisis estructural*, para explicar y describir los eventos económicos bajo estudio, a través del análisis de los parámetros estimados del modelo;
2. *En la evaluación de las políticas*, a través del análisis de los efectos que sobre los valores de equilibrio de las variables endógenas pueden producir los cambios en las variables políticas del modelo (es decir, los parámetros y las variables exógenas);
3. *En la predicción*: Cuando ya el modelo estimado ha confirmado la hipótesis o el teorema sometido a estudio, sus resultados estimados pueden ser usados para predecir valores futuros de la variable dependiente, correspondientes a valores futuros (extramuestrales) de las variables independientes o explicativas;
4. *Para fines de control y de política*: Esto requiere considerar a las variables independientes como variables de control y a la dependiente como variable objetivo y proceder a determinar los valores que habría que asignar a las variables de control para obtener unos valores deseados de la variable objetivo;
5. *Para elegir entre diversas especificaciones alternativas de modelos*: Cuando con arreglo a los postulados de la teoría económica se especifican diversas formas funcionales para representar la ecuación de regresión a utilizar en el modelo, la comparación entre sus resultados estimados permite establecer cuál de esas especificaciones satisface en mejor grado los diversos criterios de evaluación tanto económica como estadística y econométrica de las estimaciones hechas, y por tanto, cuál de ellas ofrece una mejor calidad explicativa y predictiva como ecuación representativa del fenómeno bajo estudio.

La metodología de la regresión lineal múltiple

La manera cómo la Econometría suele proceder para aplicar el análisis de regresión múltiple en el cumplimiento de sus propósitos de explicar y predecir hechos,

fenómenos y eventos económicos, en el caso de modelos uniecuacionales, podría ilustrarse a través del siguiente ejemplo hipotético.

Imagínese un país para el cual nos interese realizar un análisis empírico de las relaciones existentes entre el crecimiento de la economía nacional y el desarrollo tecnológico, científico y cultural alcanzado por dicho país. Se pretende alcanzar dicho objetivo a través de la formulación, estimación y evaluación de un modelo econométrico uniecuacional, es decir un modelo de regresión lineal múltiple, en el cual una variable dependiente, en este caso, *el crecimiento de la economía del país*, representado por la tasa interanual de crecimiento de la Producción Interna Bruta real, se asociará linealmente con un conjunto de variables independientes a través de una relación de linealidad tanto en las variables del modelo como también en los parámetros de las regresiones (Gujarati, 2003; Studenmund and Cassidy, 1987). Ejemplos de variables independientes que pueden ser utilizadas como representativas del desarrollo científico y tecnológico para intentar explicar la conducta de la variable dependiente y también para efectos de la posterior realización de predicciones, podrían ser las siguientes:

- *La producción científica*, representada por el número de publicaciones científicas, que durante un cierto período tienen las universidades, los centros públicos de investigación y las fundaciones y empresas privadas, etc.
- *El desarrollo tecnológico*, representado por los gastos públicos y privados en Investigación y Desarrollo (I+D), por las innovaciones tecnológicas patentadas, y por los patrones de comercio exterior de los sectores económicos líderes en materia de I+D.
- *El capital cultural y artístico* del país, representado por las variables anteriormente señaladas relacionándolas con la proporción que estas actividades representan en el PIB.

Estas variables se suponen suficientemente representativas del estadio de desarrollo tecnológico, científico, cultural del país bajo estudio y además, se espera que ejerzan un impacto directo sobre las potencialidades de crecimiento de la economía nacional a través de su influjo positivo sobre la eficiencia de los procesos productivos y sobre las productividades de los factores de producción.

La forma específica de la relación funcional entre las variables, a ser utilizada para efectos de la estimación, puede ser una función agregada de producción de las más comúnmente utilizadas, por ejemplo, una función del tipo Cobb-Douglas ², la cual

² C.E.Ferguson (1985). El uso de esta función de producción se ha hecho muy popular porque posee un conjunto de ventajas muy especiales para su tratamiento empírico, tales como el hecho de ser una función matemáticamente homogénea, lo cual a su vez permite estudiar sus rendimientos de escala; y

plantea una relación potencial de la forma:

$$y_t = A_t X_{1t}^{\alpha_1} X_{2t}^{\alpha_2} \dots X_{nt}^{\alpha_n} \quad (1)$$

expresión en la cual y_t (para $t=1,2,3,\dots,m$) es la tasa porcentual de crecimiento del PIB real (nacional o regional, según sea el caso); las X_{it} (para $i=1,2,3,\dots,n$; $t=1,2,3,\dots,m$), son las diversas variables independientes de las cuales se supone que depende y_t (es decir, la producción científica-humanística de las universidades, centros públicos de investigación, y fundaciones y empresas privadas, etc); las α_i son parámetros, de valores desconocidos, los cuales definen el tipo de relación de dependencia que existe entre y_t y las X_{it} y que deben ser estimados a través del uso del modelo econométrico; y A_t es un factor exógeno de desplazamiento temporal de la función (1) por progreso tecnológico, cuyo valor, al igual que los de las α_i también debe ser estimado con el uso del modelo.

Puestas así las cosas, el problema econométrico que se plantea es el siguiente: Dados los valores observados de la variable dependiente y_t y de las variables independientes X_{it} , los cuales pueden provenir de series cronológicas (datos temporales), o de información de corte transversal (datos atemporales); se requiere *estimar* los valores de los parámetros α_i y del factor de cambio tecnológico A_t , de forma tal que dichos estimadores sean los *mejores estimadores lineales e insesgados* (MELI) de dichos parámetros. Se puede demostrar que cuando en el proceso de estimación se respetan los supuestos clásicos de la regresión lineal múltiple, el método de estimación basado en los mínimos cuadrados ordinarios (MCO), suministra estimadores MELI de los parámetros desconocidos de la ecuación (1).

Para poder ser estimada, la expresión (1) debe ser linealizada, ya que los métodos econométricos de regresión lineal múltiple que se suele utilizar para estimar los valores de los parámetros α_i , requieren de funciones lineales en dichos parámetros. Para hacer posible la linealización de la ecuación potencial (1), tomamos su logaritmo natural:

$$(2) \quad \ln y_t = \ln A_t + \alpha_1 \ln X_{1t} + \alpha_2 \ln X_{2t} + \dots + \alpha_n \ln X_{nt} + \varepsilon_i$$

además el hecho de que los valores estimados de los exponentes de los factores de producción, son directamente las respectivas elasticidades de la función de producción con respecto a cada uno de los factores.

y a esta ecuación le incorporamos una variable o perturbación aleatoria ε_i , para convertirla de una expresión matemática exacta en una ecuación estocástica o probabilística a la que podamos aplicar métodos estadísticos de estimación.

Si la ecuación (2) no incluyese una variable aleatoria o probabilística tal como la variable de perturbación ε_i , entonces dicha ecuación sería simplemente la expresión de un modelo matemático. Lo que hace a la ecuación (2) un modelo econométrico es precisamente el hecho de que al incluir la variable aleatoria ε_i como *una más* de las variables explicativas del comportamiento de la variable dependiente y_t , el modelo económico original (1) ahora puede contener las especificaciones requeridas para su tratamiento empírico, es decir, para ser estimado y evaluado desde el punto de vista de hipótesis de significación estadística. De acuerdo con la teoría de la econometría, esta perturbación aleatoria ε_i es una variable no observable y por tanto una para la cual no existe una serie de valores observados que puedan ser usados en el proceso de estimación, y se supone en capacidad de representar y tomar en consideración la influencia de todos aquellos factores que pudiesen afectar a y_t , pero que no han sido considerados en forma explícita en el modelo por diversidad de causas, entre ellas, por razones de simplicidad frente a la multiplicidad de factores de diversa naturaleza (incluso, no económicas) que suelen actuar como determinantes de la casi totalidad de los eventos económicos y cuya consideración plena haría operacionalmente inmanejable las ecuaciones y los modelos resultantes.

Esta variable de perturbación aleatoria se supone distribuida desde un punto de vista probabilístico, como una distribución normal con media cero y varianza constante y debe cumplir satisfactoriamente los supuestos clásicos de la regresión lineal múltiple³, siendo los mismos, precisamente, las especificaciones empíricas requeridas para poder estimar y evaluar el modelo, es decir, para convertirlo en un modelo econométrico. Los supuestos en cuestión son los siguientes:

1. $E(\varepsilon_i) = 0$, para cualquier valor de i .
2. $\text{var}(\varepsilon_i) = \sigma^2$, para cualquier valor de i .
3. $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, para cualquier $i \neq j$.
4. $\text{cov}(\varepsilon_i, X_{it}) = 0$, para cualquier valor de i .

A estos supuestos sobre el comportamiento de las perturbaciones aleatorias se le debe añadir un supuesto referido a las variables explicativas o independientes,

³ Gujarati (2003), Studenmund & Cassidy (1987), Koutsoyiannis (1973).

X_{it} , el cual establece que éstas no deben estar linealmente correlacionadas entre sí. Este es el supuesto de la no existencia de *multicolinealidad* entre las variables independientes.

Además de estos supuestos explícitamente establecidos, hay un grupo de supuestos adicionales que también requieren ser formulados para poder aplicar apropiadamente el método de los Mínimos Cuadrados Ordinarios. Gujarati (2003; 321-322) ha señalado que aún cuando los mismos no suelen ser enumerados en forma explícita, ni en la mayoría de los libros de texto, ni en la mayoría de los trabajos con modelos econométricos, es importante no perder de vista que ellos también deben cumplirse, si se quiere una apropiada aplicación de los MCO. Esos supuestos adicionales son, en primer lugar, que los valores de las variables independientes son fijos en las muestras utilizadas; segundo, que el número de observaciones de las variables debe ser mayor que el número de parámetros a ser estimados, y, tercero, que el modelo de regresión a ser estimado, haya sido correctamente especificado.

Volviendo a los supuestos clásicos de la regresión lineal, el *Supuesto 1.* establece que las perturbaciones aleatorias deben tener una esperanza matemática, o media aritmética, igual a 0. *El Supuesto 2.*, que la varianza de las perturbaciones debe ser constante. A este supuesto se le llama supuesto de la *homoscedasticidad* de las perturbaciones aleatorias. *El Supuesto 3.*, que las perturbaciones aleatorias no deben estar correlacionadas entre sí, y por tanto su covarianza debe también ser igual a 0. Este es el supuesto de la *no autocorrelación* de las perturbaciones aleatorias. Y el *supuesto 4.*, sencillamente establece que las perturbaciones aleatorias tampoco deben estar linealmente correlacionadas con las variables explicativas.

Suponiendo normalidad de las perturbaciones aleatorias, el cumplimiento del primer supuesto garantiza que los estimadores de los parámetros desconocidos, obtenidos por aplicación del método de estimación de uso más común en econometría (Mínimos Cuadrados Ordinarios) serán estimadores linealmente insesgados, es decir, estimadores cuya media aritmética debe igualar al parámetro desconocido que pretenden estimar y, por tanto, capaces de representar fielmente las características de aquellos parámetros poblacionales. Si, además, se cumplen los restantes supuestos sobre las perturbaciones aleatorias, y también el supuesto de no multicolinealidad de las variables independientes, ello garantizará que además de linealmente insesgados, los estimadores mínimo cuadráticos obtenidos sean también estimadores de varianza mínima, es decir, estimadores eficientes. Y además de ello, serán los *mejores* estimadores lineales, insesgados y eficientes, es decir, que de todo el conjunto de estimadores que podamos diseñar para esos parámetros desconocidos, los

estimadores mínimo-cuadráticos que satisfacen los supuestos clásicos de la regresión lineal múltiple serán los que exhiban mejor desempeño desde el punto de vista de su capacidad de describir fielmente las características de los parámetros desconocidos que definen la relación entre las variables del modelo.

La primera tarea a cumplir es especificar el modelo, lo cual se ha hecho con auxilio de la teoría económica, en el caso de nuestro ejemplo, a través de la formulación de la ecuación (2). De inmediato se debe proceder a agotar las restantes fases metodológicas que comprende la Econometría: Estimar los parámetros del modelo, evaluar su significación económica, estadística y econométrica, evaluar su poder predictivo para pronosticar valores extramuestrales de las variables, y aplicar sus resultados estimados y evaluados, para fines de predicción y control de la política económica.

En cuanto a la estimación, este es un proceso meramente estadístico, una vez realizado el cual, podremos obtener los valores estimados de los parámetros desconocidos del modelo, para lo cual es necesario aplicar alguno de los métodos estadísticos disponibles de estimación. Aunque existe una diversidad de métodos de estimación, aplicable cada uno de ellos a diversas situaciones de especificación del modelo y de naturaleza del problema a ser estudiado, el método de los Mínimos Cuadrados Ordinarios (MCO) es el más comúnmente utilizado en Econometría debido a que bajo ciertas condiciones, el mismo garantiza la obtención de los mejores estimadores lineales e insesgados de los parámetros de un modelo.

Podemos designar a los estimadores o coeficientes de regresión estimados en esta fase por $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_n, \hat{A}_t$.⁴ Para poder ser aceptados como los estimadores representativos de los parámetros desconocidos del modelo, dichos estimadores deberán cumplir las propiedades de ser lineales en las observaciones dependientes, ser insesgados (es decir, con un valor medio igual al parámetro que pretenden estimar) y ser óptimos (es decir, con varianza mínima), en cuyo caso generarán un tipo

⁴ Para realizar esa estimación se suele utilizar algún software de computación de los que para tales fines hoy están disponibles, como por ejemplo, el *Econometric Views (Eviews)*, el cual es un programa de computación integrado por un conjunto de herramientas para trabajar datos de series temporales, que está basado en el *software* del TSP (Time Series Processor) de uso muy habitual en computadores de gran tamaño, pero con la ventaja de poder ser utilizado también en microcomputadores bajo ambiente Windows y con todas las características operacionales que ofrecen los *software* basados en Windows. Estos programas permiten la obtención de valores estimados de un conjunto bastante completo de parámetros y de estadísticos. De todos ellos, y para efectos de evaluar los parámetros de la ecuación estimada, interesan particularmente los coeficientes de regresión (es decir, el coeficiente de la constante y los de cada variable independiente); los errores estándar de tales coeficientes; los valores calculados de las *t* para cada coeficiente (para la prueba de significación de cada coeficiente de regresión estimado); el coeficiente de determinación R^2 , sin ajustar y ajustado por grados de libertad (para la prueba de la bondad del ajuste); la *F* calculada (para la prueba de la significación global de la regresión); y el estadístico Durbin-Watson (para la prueba de autocorrelación).

de información sobre los parámetros desconocidos del modelo de regresión, de una muy alta calidad explicativa y predictiva. Algunos autores interpretan que estos estimadores definen la *estructura* del modelo, siendo esta simplemente, el conjunto de las relaciones funcionales del modelo y sus correspondientes parámetros (Barbancho, 1962).

Luego de obtenidos los valores estimados de los parámetros y de la constante A_1 , se procederá a agotar la siguiente fase consistente en la evaluación de aquellas estimaciones. Para ello es necesario realizar un conjunto de pruebas, individuales y globales, dirigidas a verificar la significación estadística y econométrica de las mismas. Lo primero que hay que hacer al respecto es evaluar los signos y las magnitudes de los coeficientes de regresión estimados, lo cual se hace en base a los supuestos *a priori* suministrados por los principios de la teoría económica. Por ejemplo, si la ecuación estimada fuese una función de demanda de un bien, se sabe que la teoría microeconómica de la demanda establece normalmente una relación negativa entre la cantidad demandada y el precio del bien. Por tanto, el signo del coeficiente de regresión estimado correspondiente a la variable precio del bien, en el lado derecho de la ecuación de regresión, debería ser negativo para que la estimación realizada se conforme con el supuesto *a priori* de la teoría. Esta primera evaluación de las estimaciones es exclusivamente de carácter económico y generalmente involucra la necesidad de realizar análisis de elasticidades, para determinar los efectos proporcionales que sobre el comportamiento de la variable dependiente producen las variaciones proporcionales de las variables explicativas.

Luego de evaluar los signos y magnitudes de los coeficientes de regresión estimados, debe realizarse el análisis de la *significación estadística* de las estimaciones, o pruebas de primer orden, lo cual se hace en dos niveles: Por un lado, la evaluación de la significación estadística individual de cada uno de los parámetros estimados, para lo cual se utiliza la conocida prueba estadística de la *t* de Student. Por otro lado, el análisis de la significación estadística conjunta de toda la ecuación de regresión múltiple, a través de la prueba estadística de la *F* de Snedecor y la prueba de la bondad del ajuste a través del coeficiente de determinación múltiple R^2 . Estas pruebas de significación están dirigidas a establecer si los valores estimados de los parámetros y la propia ecuación de regresión estimada, considerada globalmente, son aceptables o no como buenas estimaciones suministradas por el modelo, dentro de unos niveles de confianza estadística determinada. Estas pruebas, aún no tienen carácter econométrico, sólo estadístico. Para tener ese carácter de pruebas econométricas, las mismas deben contrastar estadísticamente si los supuestos

clásicos de la regresión lineal formulados en el momento de la especificación del modelo, fueron o no respetados durante el proceso de estimación, ya que el irrespeto o violación de dichos supuestos implica la pérdida de una parte, algunas veces muy importante, de la capacidad explicativa y del poder predictivo de dichas estimaciones.

La evaluación econométrica del modelo

Como ya se dijo, el cumplimiento durante el proceso de estimación, de los supuestos clásicos de la regresión lineal, garantiza que los estimadores obtenidos por MCO para los parámetros desconocidos de la ecuación de regresión, sean estimadores MELI. Ello significa que no basta con evaluar el modelo estimado desde el punto de vista de la significación estadística de sus estimaciones, para poder determinar si el mismo posee suficientes capacidad explicativa y poder predictivo. Es necesario además realizar una evaluación del modelo, dirigida a determinar si aquellos supuestos fueron respetados o por el contrario violados. Este es el análisis de la *significación econométrica* del modelo estimado, o pruebas de segundo orden, el cual se realiza a través de un conjunto de pruebas estadísticas para determinar la presencia significativa en el modelo estimado de los principales problemas econométricos que surgen como consecuencia de la violación de dichos supuestos: La multicolinealidad, la autocorrelación y la heteroscedasticidad (Johnson, Johnson and Buse, 1987; Gujarati, 2003; Studenmund and Cassidy, 1987; Koutsoyiannis, 1973).

La Multicolinealidad

La presencia de multicolinealidad en el modelo econométrico se produce cuando se detecta la existencia de una relación lineal, exacta o inexacta, entre las variables explicativas del modelo. La multicolinealidad no es una patología de las perturbaciones aleatorias, sino de las variables explicativas, y por tanto, es un problema más atribuible al comportamiento de la muestra utilizada de valores de las variables, que al de la población estudiada. No es un problema del método de estimación utilizado; es meramente un problema de los datos muestrales. La multicolinealidad *perfecta* se produce cuando la variación en una variable explicativa es completamente determinada por variaciones en la otra variable explicativa. Mientras que la multicolinealidad *imperfecta* puede definirse como una relación funcional lineal entre dos ó más variables independientes, tan fuerte que puede afectar la estimación de los coeficientes de las variables. La multicolinealidad perfecta es muy rara en la práctica. Lo frecuente es la existencia de multicolinealidad imperfecta, la cual será

mucho más severa mientras más fuerte sea la relación lineal que vincule a las variables independientes involucradas. Salvo indicación en contrario, en este trabajo nos estaremos refiriendo a multicolinealidad imperfecta.

Las causas más frecuentes de este problema tienen que ver con una mala especificación del modelo a ser estimado, con una sobredeterminación del modelo (es decir, situaciones en las cuales las variables explicativas del modelo exceden en número a las observaciones muestrales que serán utilizadas) y, con el hecho de que las variables explicativas se muevan en el mismo sentido a lo largo del tiempo, es decir, compartan una tendencia común, lo cual suele ser frecuente en Economía cuando se trabaja con datos procedentes de series temporales. La mayoría de los fenómenos económicos que involucran uso de datos cronológicos, implica la existencia de una cierta dosis de multicolinealidad entre las variables independientes, lo cual no sería problema para la Econometría mientras dicha presencia no sea severa o estadísticamente significativa. De allí la necesidad de realización de pruebas estadísticas de significación para detectar si hay o no presencia en el modelo estimado, de alguna dosis significativa de multicolinealidad.

Las consecuencias más conocidas de la multicorrelación lineal se refieren al hecho de que en caso de presencia significativa de este problema, aunque se pueda realizar la estimación de los parámetros del modelo y estos continúen siendo estimadores MELI, las varianzas, los errores estándar y las covarianzas de los estimadores de dichos parámetros tenderán a aumentar de tamaño, con lo cual perdería precisión la prueba *t* de Student para evaluar la significación estadística de dichos estimadores considerados individualmente.

Para detectar la presencia significativa de multicolinealidad en el modelo estimado no se aplican pruebas estadísticas de significación, como si ocurre en el caso de los restantes problemas econométricos, debido a que el problema no es generado por causas relacionadas con las perturbaciones aleatorias sino por problemas en los valores muestrales. En su defecto lo que se suele aplicar son algunas reglas prácticas, de las cuales las de uso más frecuente son las siguientes: 1) Examinar si el valor estimado del coeficiente de determinación R^2 es alto (mayor a 0.80) mientras que los coeficientes de regresión estimados son poco significativos. En tal caso existirá una fuerte posibilidad de existencia de multicolinealidad significativa. 2) Revisar los coeficientes de correlación parcial estimados de las variables independientes ya que valores altos de estos coeficientes suelen ser indicativos de presencia significativa de multicolinealidad en el modelo estimado; y en último caso, 3) Estimar la regresión de cada variable independiente contra las restantes para obtener los valores de los coeficientes de determinación, R^2 , de cada una de dichas variables,

que en caso de ser altos, indicarían la existencia de una fuerte correlación entre esa y las restantes variables independientes del modelo.

Por las mismas razones por las cuales para la detección del problema de la multicolinealidad no suele recomendarse la aplicación de pruebas estadísticas de significación sino la aplicación de reglas prácticas, en el caso de su corrección también se sugiere el uso de ciertas reglas prácticas. Ellas son: 1) Utilizar información adicional, ya que este es el mejor remedio para corregir los valores indeseablemente altos de las varianzas de los estimadores. Dado que la multicolinealidad es esencialmente un problema de los datos, la incorporación de datos adicionales, deseablemente no colineales, siempre ayudará a corregirlo, incluso hasta en los casos en que los nuevos datos tengan problemas de multicolinealidad. Ello debido a que datos adicionales en la regresión, implican varianzas menores. 2) Combinar datos cronológicos con datos transversales, es decir, realizar una mezcla de datos y ajustar nuevamente las regresiones con los nuevos datos. 3) Eliminación de una o más variables colineales. Esto, sin embargo, suele generar sesgos o errores de especificación en el modelo, que pudieran llegar a ser más inconvenientes desde el punto de vista de sus efectos sobre la calidad de los estimadores que la propia multicolinealidad. 4) Incorporar estimaciones de otros estudios empíricos previos disponibles. 5) Utilizar un índice compuesto de las variables independientes que son colineales, que sea capaz de representar a ese grupo de variables. Ese índice solo debería ser creado y utilizado en las estimaciones si la interpretación económica combinada de dichas variables tiene algún significado útil para los fines perseguidos por el modelo.

Corregido el modelo mediante el uso de alguna de las reglas mencionadas, la realización de nuevas estimaciones de las ecuaciones de regresión corregidas debería conducir a la obtención de estimadores no afectados por una presencia significativa de multicolinealidad entre las variables independientes, pudiéndose así proceder a continuar realizando las restantes pruebas econométricas.

La heteroscedasticidad

La heteroscedasticidad es un problema econométrico que surge cuando se viola el supuesto de la regresión lineal clásica según el cual las perturbaciones aleatorias del modelo econométrico, deben poseer varianzas constantes, es decir, deben ser perturbaciones homoscedásticas. Recuérdese que el Supuesto 2 exigía que $\text{var}(\varepsilon_i) = \sigma^2$, para cualquier valor de $i = 1, 2, 3, \dots, n$. Pues bien, en el caso de

existencia de heteroscedasticidad se cumpliría que $\text{var}(\varepsilon_i) = \sigma_i^2$, para valores de $i = 1, 2, 3, \dots, n$; es decir, las varianzas de las perturbaciones habrían dejado de ser constantes e iguales entre sí. El significado de este supuesto es que la variación de las perturbaciones ε_i alrededor de su media (igual a cero), no depende de los valores de las variables independientes X_i , es decir, la varianza de ε_i no es una función de X_i . Cuando este supuesto no se cumple, entonces, ocurre una dispersión creciente, o decreciente, de las observaciones muestrales alrededor de la línea de regresión.

La heteroscedasticidad puede ser *pura*, cuando es causada por la perturbación aleatoria en una ecuación de regresión que ha sido correctamente especificada, o *impura* cuando es causada por un error de especificación, de los cuales el más frecuente es la omisión de alguna (s) variable(s) independiente(s). Cuando esto último ocurre, la fuente de la heteroscedasticidad viene dada por el hecho de que la porción del efecto que la variable omitida debió producir sobre la variable dependiente, debe ser absorbida por la perturbación aleatoria. Salvo indicación en contrario, en este trabajo nos referimos a la heteroscedasticidad pura.

Las consecuencias de la existencia de una presencia significativa de heteroscedasticidad en el modelo estimado, son variadas: 1) Si las ε_i son heteroscedásticas, entonces no es posible utilizar las varianzas de los coeficientes de regresión estimados para construir las t con las cuales establecer los intervalos de confianza para las pruebas de significación individual. 2) Cuando las ε_i son heteroscedásticas, los estimadores MCO de los parámetros de la ecuación de regresión, aún cuando continúan siendo insesgados, pierden la propiedad de varianza mínima, es decir, dejan de ser estimadores eficientes. 3) La predicción de y_i para un valor dado de X_i , tendrá una elevada varianza, es decir, será ineficiente.

Debido a que la heteroscedasticidad, a diferencia de la multicolinealidad, es un problema de las perturbaciones aleatorias, entonces, su detección requiere de la realización de pruebas de significación estadística. Las tres pruebas de heteroscedasticidad más conocidas y utilizadas son las siguientes (Gujarati, 2003; Koutsoyiannis, 1973):

1. Prueba del coeficiente de correlación por rangos de Spearman

Esta es la más sencilla de las pruebas estadísticas de heteroscedasticidad. Consiste en ajustar la regresión de y_i contra las X_i , obteniéndose los residuos o

estimaciones de las ε_i . Se ordenan los valores estimados de las ε_i ignorando sus signos y los valores de las X_i en orden ascendente (o descendente); y se procede a calcular el coeficiente de correlación por rangos:

$$r_{\varepsilon, X}^2 = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$$

expresión en la cual, D_i es la diferencia entre los rangos de pares correspondientes de X , ε_i ; n es el número de observaciones de la muestra. Un valor alto de este coeficiente, es decir, un valor próximo a la unidad, sugeriría la presencia de heteroscedasticidad. Si la ecuación de regresión es una ecuación de más de una variable independiente, entonces, se debe calcular el coeficiente de correlación por rangos entre ε_i y cada una de las variables explicativas por separado.

2. La prueba de Goldfeld y Quandt

Esta prueba es aplicable a muestras grandes, para las cuales, el número de observaciones es al menos el doble de los parámetros a ser estimados. La prueba supone normalidad e independencia (es decir, no autocorrelación) en las ε_i . Para practicar la prueba se comienza ordenando las observaciones según la magnitud de la variable explicativa X y se procede a seleccionar arbitrariamente un cierto número (c) de observaciones centrales, las cuales excluimos del análisis. Según los experimentos hechos por Goldfeld y Quandt se logró determinar que para muestras mayores a $n = 30$, el número óptimo de observaciones centrales a ser omitidas es una cuarta parte de todas las observaciones. El resto, $(n-c)$, se divide en dos sub-muestras de igual tamaño, $(n-c)/2$, una para los valores pequeños de X y otra para los valores grandes. Se ajusta, por separado, una regresión a cada sub-muestra y se obtiene para cada una de ellas la suma de los residuos al cuadrado. Los residuos son las estimaciones de las ε_i y surgen de la regresión como las desviaciones entre los valores de las Y_i y la línea de regresión estimada. Designando tales residuos por e_i se tiene que: $\sum e_1^2$ y $\sum e_2^2$ son respectivamente la suma de los residuos al cuadrado para ambas submuestras, las cuales se distribuyen como una χ^2 con $[(n-c)/2] - k$ grados de libertad, en donde k es el número total de parámetros en el modelo. Al dividir cada una de estas sumas por sus grados de libertad, se obtienen las estimaciones de las

varianzas de las ε_i para ambas sub-muestras; y el cociente de ambas varianzas, por ser el cociente de dos χ^2 , será una distribución F de Snedecor:

$$F^* = \frac{\sum e_2^2 / [\{(n-c)/2\} - k]}{\sum e_1^2 / [\{(n-c)/2\} - k]} = \frac{\sum e_2^2}{\sum e_1^2}$$

con $v_1 = v_2 = [\{(n-c)/2\} - k] = (n - c - 2k) / 2$ grados de libertad, y en donde n es el número total de observaciones, c es el número de observaciones centrales omitidas y k es el número de parámetros estimados de cada regresión.

Si ambas varianzas fuesen iguales, es decir, si las ε_i fuesen homoscedásticas, el valor de F^* deberá ser igual a 1. Si las varianzas difieren, entonces F^* será mayor que 1 ya que por el propio diseño de la prueba, se tiene que $\sum e_2^2 > \sum e_1^2$. A partir de acá, se compara el valor observado de F^* con el valor teórico tabulado de F para la región crítica y el nivel de significación previamente definidos de la prueba, a los grados de libertad ya establecidos. Si $F^* > F$ rechazaríamos la hipótesis nula de no existencia de diferencia significativa entre las varianzas de las ε_i en las dos sub-muestras y, por tanto, estaríamos aceptando que existe heteroscedasticidad al nivel de significación que hayamos fijado para la prueba. Mientras mayor sea el cociente observado F^* , más severa será la heteroscedasticidad del modelo.

3. La prueba de Glejser

Esta prueba de significación estadística se practica de la siguiente forma: Se realiza el ajuste de la regresión de y_i contra todas las variables explicativas y se calculan los residuos, e_i . A continuación se ajustan los valores absolutos de las e_i , es decir $|e_i|$, contra la X_i con la cual se piensa sobre una base *a priori* que $\sigma_{\varepsilon_i}^2$ está asociada. Dado que la forma de la relación entre $|e_i|$ y X_i es desconocida, se deben ensayar varias formas específicas de formas funcionales, las cuales se ajustan por MCO y se escoge la que arroje un mejor ajuste desde el punto de vista del coeficiente de correlación y de los errores estándar de los coeficientes de regresión estimados. La presencia de heteroscedasticidad se juzga en función de la significación estadística de los parámetros estimados de esta última regresión, evaluada la misma bien sea a través de pruebas t o de pruebas F. Si los parámetros resultan estadísticamente

significativos, entonces concluimos que en el modelo estimado las ε_i son heteroscedásticas.

La prueba de Glejser tiene la ventaja de que no solo detecta la existencia de heteroscedasticidad, sino que también suministra información sobre la forma que la misma adopta, es decir, sobre la manera como las $\sigma_{\varepsilon_i}^2$ se asocian con las X_i , lo cual es importante para los efectos de la corrección del problema.

Una vez detectada la presencia de heteroscedasticidad a través de cualquiera de las pruebas señaladas, es necesario corregir el problema, lo cual se hace simplemente transformando el modelo original de tal forma que en la nueva formulación del mismo, las perturbaciones aleatorias transformadas tengan varianza constante y se pueda aplicar el método de los MCO al modelo transformado.

En general, las tres formas más frecuentes de tratar de remediar el problema de la heteroscedasticidad son las siguientes (Studenmund and Cassidy, 1987; Koutsoyiannis, 1973): 1) Incluir en el modelo una variable previamente omitida. Si la heteroscedasticidad es impura, hay que determinar e incluir la variable que quedó fuera del modelo en el momento de su especificación y que está causando la heteroscedasticidad, una vez hecho lo cual, se procede a ajustar nuevamente la regresión por MCO. 2) Utilizar como método de estimación el método de los Mínimos Cuadrados Ponderados (MCP), el cual es un caso especial del método de los Mínimos Cuadrados Generalizados (MCG) (Aitken, 1935). 3) Redefinir las variables. Esta es una forma directa de aliviar el problema de heteroscedasticidad. La redefinición de las variables del modelo debe basarse en una revisión muy cuidadosa de los fundamentos teóricos que inspiraron la especificación del modelo, los cuales deben ser respetados en la medida de lo posible. Una vez redefinidas las variables, el ajuste del nuevo modelo que surja, debería aliviar sustancialmente la presencia severa de heteroscedasticidad.

La autocorrelación

La autocorrelación, o correlación serial, de las perturbaciones aleatorias es un problema econométrico que se produce cuando en la estimación del modelo econométrico se irrespeta el supuesto según el cual los valores sucesivos de la variable aleatoria deben ser independientes entre sí. El Supuesto 3 de la regresión lineal múltiple exigía que $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, para cualquier valor de $i \neq j$. La autocorrelación es un caso especial de la teoría de la correlación, ya que se refiere, no a la relación existente entre dos o más variables diferentes, sino entre valores

sucesivos de una misma variable. La autocorrelación es un problema exclusivo de las series temporales. Puede decirse que ella no existe como problema en los datos transversales o atemporales (a menos que la muestra sea no aleatoria), ya que los datos transversales al estar referidos a “un punto en el tiempo”, excluyen la posibilidad de interdependencia temporal entre ellos (Koutsoyiannis, 1973).

La autocorrelación puede existir, y de hecho es frecuente, en la mayoría de las variables económicas; sin embargo, nos interesa solo el caso en el que este problema se refiere a las perturbaciones aleatorias del modelo. Por consiguiente, podemos tratar la autocorrelación de las ε_i en la misma forma en que trataríamos la correlación en general. Generalmente estamos interesados en trabajar con el caso simple de una relación lineal de primer grado entre cualquier par de valores sucesivos de ε_i , de la forma:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

la cual se conoce como una *relación o esquema autorregresivo de primer orden de Markov AR(1)*. En esta expresión, v_t es una nueva variable de perturbación aleatoria, no observable, utilizada para recoger en la relación entre ε_t y ε_{t-1} todos los factores que pudiendo influir en dicha relación no están explícitamente señalados en la misma y para la cual suponemos la vigencia de los mismos supuestos clásicos de la regresión lineal múltiple; ρ es interpretado como un coeficiente de autocorrelación simple entre los valores ε_t y ε_{t-1} de la perturbación aleatoria, lo cual es simplemente un caso especial del coeficiente de correlación simple entre cualquier par de variables, estando sujeto por tanto a las mismas críticas que suelen hacerse al coeficiente de correlación simple. Por ejemplo, ese coeficiente no es apropiado para analizar relaciones no lineales entre ε_t y ε_{t-1} , ni mucho menos para esquemas autorregresivos de orden superior. Aunque la autocorrelación puede ser de cualquier signo, consideraremos solo los casos de existencia de autocorrelación *positiva* entre ε_t y ε_{t-1} , ya que la mayoría de las variables económicas tienden a aumentar en períodos de crecimiento, o a exhibir comportamientos cíclicos.

El hecho de que las perturbaciones aleatorias de un modelo econométrico estén autocorrelacionadas, puede deberse a múltiples causas, siendo las más frecuentes las siguientes: 1) La omisión de variables explicativas en la especificación del modelo, ya que la influencia de cualquier variable explicativa autocorrelacionada que haya sido excluida del modelo, se reflejará en el modelo en la variable aleatoria, la cual por tal razón estará autocorrelacionada. 2) Una mala especificación de la forma

matemática funcional del modelo. 3) La manipulación de los datos que empíricamente alimentan a los modelos econométricos. El uso de promedios para obtener valores de series cronológicas, o la interpolación o la extrapolación de datos para obtener valores faltantes en muestras temporales, son claros ejemplos de manipulación de la información estadística de uso frecuente en economía que inciden directamente en generar autocorrelación en las perturbaciones. 4) El uso de series temporales no estacionarias, es decir, cuyas características estadísticas principales (media, varianza, covarianza, etc) varían a medida que transcurre el tiempo, lo cual hace que las perturbaciones aleatorias y los residuos también sean no estacionarios.

Las principales consecuencias de la presencia de autocorrelación de las perturbaciones en los modelos econométricos son las siguientes: 1) En presencia de autocorrelación severa, los estimadores de los parámetros obtenidos por MCO siguen siendo lineales e insesgados, pero dejan de tener varianza mínima, es decir, dejan de ser eficientes. Por tanto, las pruebas t y F para evaluar la significación individual de los parámetros estimados del modelo ya no podrían continuar practicándose. 2) La varianza de la perturbación aleatoria ε_i puede quedar sustancialmente subestimada cuando las ε_i están autocorrelacionadas en grado significativo. 3) Cuando las perturbaciones aleatorias están autocorrelacionadas, entonces las predicciones de Y para valores dados de las X, realizadas a partir de los resultados suministrados por la aplicación de MCO, serán ineficientes, es decir, tendrán una varianza mayor que la que tendrían las predicciones basadas en estimaciones obtenidas mediante aplicación de otros métodos (por ejemplo, el de Mínimos Cuadrados Generalizados).

La detección de la presencia de autocorrelación de las perturbaciones aleatorias en el modelo estimado, se puede hacer de diversas maneras. Una muy sencilla, consiste en representar gráficamente los residuos de la regresión en un diagrama bidimensional, bien sea contra sus propios valores retrasados o contra el tiempo y observar si el diagrama de puntos resultante exhibe un patrón regular de comportamiento gráfico. Esta técnica ofrecería una primera percepción de si el problema está presente, pero carece de la precisión de una prueba estadística de significación, razón por la cual para detectar la autocorrelación se recomienda realizar los correspondientes tests que han sido diseñados para tales efectos.

La prueba de Durbin-Watson

De todas las pruebas que actualmente están disponibles, la más tradicional y utilizada es la prueba de Durbin-Watson. La prueba de significación estadística de Durbin-

Watson es un test, aplicable a muestras pequeñas y a esquemas autorregresivos de primer orden del tipo $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, diseñada en 1951 por los estadísticos J. Durbin y G. S. Watson y la cual se ha convertido en el test más popular y el más utilizado para detectar la presencia severa de autocorrelación en los modelos econométricos. La prueba se basa en el uso del estadístico d de Durbin-Watson para determinar si existe correlación serial de primer orden en las perturbaciones aleatorias de una ecuación de regresión a través del examen de los residuos de la estimación por MCO de esa ecuación. Se define el estadístico d de Durbin-Watson a través de la siguiente expresión:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

que no es más que la relación por cociente entre la suma de la diferencia de los residuos sucesivos al cuadrado y la suma de los cuadrados de los residuos. Si definimos el coeficiente de autocorrelación por ρ , se puede demostrar que entre el estadístico d y el coeficiente ρ existe la siguiente relación:

$$d \cong 2(1 - \hat{\rho})$$

expresión en la cual, $\hat{\rho}$ es el valor estimado del coeficiente de autocorrelación, obtenido por aplicación de los MCO según la expresión:

$$\hat{\rho} = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2}$$

. La prueba de Durbin-Watson se puede describir simplemente como un test dirigido a contrastar la hipótesis nula de que $\rho = 0$, la cual implica que las ε_i no están autocorrelacionadas con un esquema de primer orden; contra la hipótesis alternativa de que las ε_i sí están autocorrelacionadas. Se puede demostrar que los valores del estadístico d de Durbin-Watson están comprendidos entre 0 y 4, y que cuando $d = 2$, entonces $\rho = 0$. Por tanto, verificar la hipótesis nula de que $\rho = 0$, equivale a contrastar la hipótesis de que $d = 2$. La prueba se realiza, comparando el valor estimado del estadístico d contra los valores teóricos de dicho estadístico, contenidos en la tabla de valores de d elaborada por Durbin y Watson, a los niveles de significación estadística previamente establecidos y con grados de libertad

determinados por el número de observaciones muestrales n y el número de parámetros k contenidos en el modelo.

La prueba presenta el problema de que no se conoce la distribución probabilística exacta del estadístico d . Sin embargo, Durbin y Watson, al elaborar la tabla de valores teóricos de d lograron establecer unos límites superior (d_u) e inferior (d_L), para los niveles de significación del estadístico d que son apropiados para contrastar la hipótesis nula ya establecida. Los criterios de decisión de la prueba son los siguientes:

1. Si $d < d_L$, entonces rechazamos la hipótesis nula de no autocorrelación y aceptamos que existe autocorrelación positiva de primer orden.
2. Si $d > (4 - d_L)$, entonces rechazamos la hipótesis nula de no autocorrelación y aceptamos que existe autocorrelación negativa de primer orden.
3. Si $d_u < d < (4 - d_u)$, entonces aceptamos la hipótesis nula de no autocorrelación.
4. Si $d_L < d < d_u$, o si $(4 - d_u) < d < (4 - d_L)$, afirmamos que la prueba no es concluyente.

Para corregir el problema de autocorrelación es necesario tener en cuenta que la solución que se adopte en cada caso particular dependerá de la fuente de la autocorrelación, es decir, del tipo de interdependencia existente entre las perturbaciones aleatorias. Pero, como las ε_i son variables *no observables*, entonces, durante el proceso de estimación es necesario suponer la vigencia de algún mecanismo de relación entre las ε_i . El mecanismo generalmente adoptado es el ya señalado esquema autorregresivo de primer orden AR(1).

Si el origen de la autocorrelación está en la omisión de variables, la corrección apropiada es incluir esas variables en el conjunto de variables explicativas del modelo. La manera más simple de comprobar si esta inclusión funciona es ajustando los residuos e_i de la estimación original, contra las variables que, sobre bases teóricas *a priori*, se presume que sean variables explicativas relevantes para el fenómeno bajo estudio y contrastando el tipo de relación que resulte entre ellas. Del mismo modo, si la fuente de la autocorrelación fuese una mala especificación de la forma matemática del modelo, lo recomendable sería cambiar la forma inicial de la ecuación de

regresión. De las múltiples posibilidades que al respecto se podrían intentar, suelen ser frecuentes las linealizaciones en el logaritmo natural, cuando la forma inicial es una función potencial, o el uso de potencias superiores de las variables independientes. En cualquier caso, sería necesario ajustar los residuos originales contra las nuevas formas de las variables explicativas y reexaminar los nuevos residuos resultantes.

Ya detectada la autocorrelación mediante la aplicación de cualquiera de las técnicas señaladas, el procedimiento apropiado de corrección es obtener una estimación de los coeficientes de correlación ρ y aplicar MCO al conjunto de datos transformados. La transformación de los datos dependerá del patrón de estructura autorregresiva adoptada.

Un comentario final

Una vez corregida la presencia de los problemas econométricos que pueda exhibir el modelo y por tanto evaluados el poder explicativo y la capacidad predictiva de las estimaciones realizadas en el modelo, ya las mismas pueden ser utilizadas para los propósitos indicados con anterioridad: Bien sea para realizar análisis estructural, o para predecir valores futuros de la variable dependiente para valores dados de las independientes, o para fines de control y de política, o para seleccionar entre especificaciones alternativas de un modelo econométrico. Sin embargo, antes de ello, será conveniente hacer una revisión final del modelo en función de lo que algunos denominan *“la econometría de las series de tiempo”*, lo cual es un desarrollo reciente de la econometría que ha ganado creciente espacio y popularidad dentro de la profesión. Si en la investigación se usaron series temporales, debe procederse a realizar un análisis de *estacionariedad* y de *cointegración*, basado en *raíces unitarias*, el cual persigue el propósito de determinar si las estimaciones realizadas son o no espúrias. Este será un procedimiento necesario ya que todo el análisis de la metodología de la regresión lineal múltiple realizado hasta ahora, supone implícitamente que las series temporales utilizadas son series *estacionarias*, es decir, series de datos estadísticos cronológicos para las cuales su media y su varianza no varían sistemáticamente a lo largo del tiempo. Sin embargo, por la naturaleza misma de las variables económicas utilizadas, la mayoría de las series cronológicas en economía, son *no estacionarias*. Ello significa que si se hacen regresiones de variables que son series temporales no estacionarias, las pruebas convencionales basadas en los estadísticos t, F, chi cuadrada, etc, pierden efectividad ya que dichas pruebas están basadas en el supuesto de la estacionariedad de las series

cronológicas. Por otro lado, si se trabaja con series de tiempo no estacionarias, es muy probable que aún no existiendo una relación significativa entre las variables dependiente e independientes del modelo, se obtengan elevados valores de los coeficientes de determinación múltiple R^2 , lo cual sugeriría la falsa impresión de existencia de una fuerte asociación entre dichas variables. Cuando esto ocurre, se dice que lo que realmente existe es una *relación espúria*, la cual obedece más a un problema de tendencias comunes o compartidas de las series estadísticas utilizadas, que a una verdadera relación significativa entre dichas variables. Es importante entonces, para dar el toque final a la aplicación de la metodología de la regresión lineal múltiple, realizar las pruebas de significación estadísticas dirigidas a establecer si las series empleadas son o no estacionarias, lo cual se hace contrastando hipótesis relacionadas con la existencia de raíces unitarias en dichas series. Los tests de Dickey-Fuller (DF) y de Dickey-Fuller Aumentada son las pruebas estadísticas de significación más usadas a este respecto, cuyos detalles se obvian por trascender el propósito del presente trabajo (Gujarati, 2003).

Paralelamente a la realización de pruebas para contrastar la estacionariedad de las series cronológicas empleadas, también es necesario en este particular, verificar la existencia de *cointegración* entre las diversas series cronológicas utilizadas. La cointegración es una manera de enfrentar el problema de las regresiones espurias, ya que a través de ella, se puede lograr que una combinación lineal de dos o más series temporales no estacionarias, puedan servir como si fuese una serie estacionaria. Las pruebas estadísticas más habituales para verificar la cointegración de series temporales son las Engle-Granger y la de Engle-Granger Aumentada, cuyos detalles tampoco se explican en el presente ensayo (Gujarati, 2003), ya que serán objeto de ulteriores trabajos.

Bibliografía

- Aitken, A. C. (1935). "On Least Squares and Linear Combinations of Observations", *Proceedings of the Royal Statistical Society of Edimburg*, vol. 55. pp. 42-48.
- Barbancho, Alfonso G. (1962). *Fundamentos y Posibilidades de la Econometría*. Barcelona: Ediciones Ariel.
- Beach, E. F. (1965). *Economic Models: An Exposition*. New York: John Wiley & Sons, Inc.
- Bishop, Robert V. (1981). "The Use and Misuse of Summary Statistics in Regression Analysis", *Agricultural Economics Research*, vol. 33, Nº 1, January, pp. 13-30.
- Cramer, Elliot M. (1972). "Significance Tests of Models in Multiple Regression", *The American Statistician*, vol. 26, Nº 4, October. pp. 26-30.
- Dagum, Camilo y Estela de Dagum (1971). *Introducción a la Econometría*. México: Siglo Veintiuno Editores, S.A.
- Durbin, J and G. S. Watson (1951). "Testing for Serial Correlation in Least-Squares Regression", *Biometrika*, vol. 38. pp. 159-171.
- Dutta, M (1982). *Métodos Económicos*. Cincinnati: South-Western Publishing Company.
- Ferguson, C. E. (1985). *Teoría Neoclásica de la Producción y la Distribución*. México: Editorial Trillas, S. A. de C. V.
- Frisch, Ragnar (1933), "Note on term 'Econometrics' ". *Econometrita*, vol 1, issue 1, pp. 1-4. January.
- Geary, R. C. and C. E. V. Leser (1968). "Significance Tests in Multiple Regression", *The American Statistician*, vol. 22, February. pp. 20-21.
- Glejser, H. (1969). "A New Test for Heteroscedasticity", *Journal of the American Statistics Association*, vol. 64, pp. 316-363.
- Goldfeld, S. M. and R. E. Quandt (1965). "Some Tests for Homoscedasticity", *Journal of the American Statistics Association*, vol. 60, pp. 539-547.
- Gujarati, Damodar N. (2003) *Econometría*. Cuarta Edición. México: McGraw-Hill.
- Haavelmo, Trygve. (1944). "The Probability Approach in Econometrics", *Econometrita*, Supplement, vol. 12, iii.
- Intriligator, Michael D. (1978). *Econometric Models, Techniques, and Applications*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
- Johnson, Aaron C., M. V. Johnson and R. C. Buse (1987). *Econometrics: Basic and Applied*. New York: Macmillan Publishing Company.

Kennedy, Peter (1993). *A Guide to Econometrics*. Third Edition. Cambridge, Mass.:The MIT Press.

Koutsoyiannis, Anna (1973). *Theory of Econometrics*. New York: Harper & Row.

Loría, Eduardo (2007). *Econometría con Aplicaciones*. México: Pearson Educación.

Marshall, Jacob. (1948). *Introduction to Econometrics*, Buffalo: University of Buffalo Press. p.1.

Pindyck, Robert and D. Rubinfeld (1981). *Econometric Models and Economic Forecasts*. New York: McGraw-Hill Book Company.

Studemund, A. H. and Henry J. Cassidy (1987). *Using Econometrics: A Practical Guide*. Boston: Little, Brown and Company.

Valavanis, Stefan (1959). *Econometrics*. New York: McGraw-Hill.