SÁNCHEZ GONZÁLEZ, CARLOS
HERRERÍAS VELASCO, JOSÉ MANUEL
csanchez@ugr.es
jmherrer@ugr.es

Departamento de Economía Aplicada
Universidad de Granada
Campus de Cartuja
18071 GRANADA

ESTIMADORES DE TAMAÑO MUESTRAL VARIABLE (ETMV)

Palabras Clave: tamaño muestral, submuestra, estimadores jackknife,

1. Resumen.-

En la presente comunicación se analiza el comportamiento predictivo de un procedimiento de estimación que considera el conjunto de posibles estimaciones mínimo cuadráticas que pueden obtenerse para distintas submuestras de una muestra de tamaño dado. El procedimiento de estimación propuesto se articula en dos etapas, una primera en la cual para una muestra dada se determina un conjunto de estimaciones de los parámetros , en la segunda se selecciona de entre las anteriores la que resultan más adecuada para cada observación. El comportamiento del estimador propuesto se compara con los correspondientes a los estimadores de MCO y Jackknife obteniendo errores de predicción cuadrátricos medios menores para todos los horizontes de predicción analizados.

2. Introducción.-

El modelo de regresión convencional es la técnica econométrica más ampliamente utilizada para el estudio de la posible existencia de relaciones de carácter causal entre un conjunto de variables susceptibles de ser cuantificadas. La estimación de una determinada especificación nos permite obtener los valores correspondientes a los parámetros del modelo con el fin último, entre otros de poder contar con una aproximación acerca de cual será el valor que adoptará la variable endógena, dados unos valores predeterminados de las variables explicativas. El procedimiento común de estimación de los valores de estos parámetros es el de mínimos cuadrados ordinarios, que pondera de igual modo todas las observaciones muestrales, siendo única la relación causal estimada entre las variables con independencia de cuáles sean los valores de las variables explicativas.

La metodología que se presenta en esta comunicación permite ajustar o predecir el valor de la variable dependiente teniendo en cuenta para ello las observaciones cuyas características son más próximas a las de la observación que se pretende explicar.

El procedimiento de estimación de la variable dependiente se estructura en dos fases, una primera en la cual par una muestra dada se determina un conjunto de estimaciones de los parámetros, y una segunda en la que se selecciona de entre las anteriores la que resulta más adecuada para cada observación.

3. Estimadores de Tamaño Muestral Variable (ETMV).-

Supongamos expecificado un modelo de regresión convencional:

$$y_t = \vec{x}_t{}' \vec{\boldsymbol{b}} + \boldsymbol{e}_t$$

siendo el vector de parámetros de dimensiones kx1. En estas condiciones es posible obtener en principio T-k-1 estimaciones distintas $(\vec{b}^{\,j}, s)$, correspondientes al vector de parámetros $(\vec{\boldsymbol{b}}, \boldsymbol{s})$, dependiendo del número de observaciones incluidas en la respectiva estimación (desde la k+1 ésima hasta la j-ésima). Al mismo tiempo podemos obtener un vector de valores ajustados corrrespondientes a la variable dependiente, de la forma: $\vec{y}^{\,j} = X\vec{b}^{\,j}$.

El grado de ajuste alcanzado con cada uno de los vectores de simulación $\vec{y}^{\,j}$ para cada una de las componentes de los verdaderos valores de la muestra recogidos en el vector $\vec{y}$ no es el mismo para toda la muestra, de hecho existirán observaciones muestrales y_t para las cuales las simulaciones correspondientes a una determinada estimación $\vec{x}_t{}' \vec{\boldsymbol{b}}^{\,j}$ resulten menos acuradas que las correspondientes a la simulación $\vec{x}_t{}' \vec{\boldsymbol{b}}^{\,i}$, por poner un ejemplo.

Tomando el criterio de mínimos cuadrados o cualquier otra distancia como por ejemplo la diferencia en valor absoluto, podemos considerar como mejor estimador de $\vec{\boldsymbol{b}}_t$ para la observación y_t, al vector:

$$\vec{b}_t^{\,*} = \arg \min_{\vec{b}^{j} \in \left\{\vec{b}^{k+1}, \vec{b}^{k+2}, \ldots, \vec{b}^{T}\right\}} \left(y_t - \vec{x}_t^{\,'} \vec{\boldsymbol{b}}^{j} \right)^2$$

Este estimador puede considerarse un estimador en dos etapas, en el sentido de que primero se obtienen los MCO usuales para cada uno de los tamaños muestrales {k+1,k+2,...,T},y en una segunda etapa se minimiza la desviación que se obtiene en cada observación para todos y cada uno de los estimadores obtenidos en la primera fase, determinándose el estimador óptimo correspondiente a esa observación. De este modo tenemos una estimación distinta del vector de parámetros para cada uno de los valores de la muestra, de manera que podemos escribir la secuencia de estimadores (ETMV):

$$\left\{\vec{b}_t^{\,*}\right\}, t = 1,2,\ldots,T$$

y la correspondiente senda óptima estimada (SOE) para los valores de la variable dependiente:

$$\left\{\vec{x}_t^{\,'} \vec{\boldsymbol{b}}_t^{\,*}\right\}, t = 1,2,\ldots,T$$

4. Propiedades estadísticas de los ETMV.-

Con este procedimiento se obtiene una adherencia a la muestra mayor que la obtenida con el procedimiento de estimación de MCO.

En efecto, sea el estimador de MCO correspondiente a la totalidad de la muestra $\vec{b}^{T}$, en definitiva, aquel que se obtiene cuando consideramos la totalidad de las observaciones muestrales T. El vector de residuos MCO será:

$$\vec{e} = \vec{y} - \hat{\vec{y}} = \vec{y} - X\vec{b}^{T}$$

El vector de residuos que proporciona la senda óptima estimada es:

$$\vec{e}^{\,*} = \vec{y} - \vec{y}^{\,*}$$

nótese que cualquiera que sea $t \in \{1,2,\ldots,T\}$ y cualquiera que sea $\vec{b}^{j} \in \left\{\vec{b}^{k+1}, \vec{b}^{k+2}, \ldots, \vec{b}^{T}\right\}$, se verifica que: $\left|e_t^*\right| = \left|y_t - y_t^*\right| = \left|y_t - \vec{x}'_t b_t^*\right| \leq \left|y_t - \vec{x}'_t b^j\right|$, y en particular: $\left|e_t^*\right| \le \left|y_t - \vec{x}'_t b^T\right|$, y por tanto para la suma de cuadrados de los residuos se verifica: $e^{*\prime} e^* \le e' e$

Por otro lado se verifica:

$$E\left[e_t^*\right] = \left[y_t - \vec{x}_t' E\left[\vec{b}_t^*\right]\right] = y_t - \vec{x}_t' \boldsymbol{b}$$

Puesto que para cada uno de los correspondientes tamaños muestrales los distintos $\vec{b}^{\,j}$ eran ELIO al cumplir el teorema de Gauss Markov.
Respecto de la varianza tenemos:

$$\begin{aligned}
&\operatorname{var}\left[e_t^*\right] = E\left[y_t - \vec{x}_t' \vec{b}_t^*\right]^2 \\
&= E\left[\vec{x}_t' \boldsymbol{b} + \boldsymbol{e}_t - \vec{x}_t' \vec{b}_t^*\right]^2 \\
&= \vec{x}_t' E\left(\vec{b}_t^* - \boldsymbol{b}\right)\left(\vec{b}_t^* - \boldsymbol{b}\right)' \vec{x}_t + E\left(\boldsymbol{e}_t\right)^2 - 2\vec{x}_t' E\left(\vec{b}_t^* - \boldsymbol{b}\right)\boldsymbol{e}_t \\
&= \vec{x}_t' \operatorname{var}\left(\vec{b}_t^*\right)\vec{x}_t + \boldsymbol{s}^2 - 2\vec{x}_t' \operatorname{cov}\left[\left(\vec{b}_t^* - \boldsymbol{b}\right)\boldsymbol{e}_t\right]
\end{aligned}$$

Cada una de las estimaciones es insesgada puesto que $E\left(b_t^*\right) = E\left(\vec{\boldsymbol{b}}^{\,j}\right) = \vec{\boldsymbol{b}}$ para algún j; sin embargo, cuando nos referimos a la senda óptima de estimaciones $\left\{\vec{b}_t^*\right\}$ el valor correspondiente a cada observación t no tiene por qué coincidir con el correspondiente $\left\{\vec{b}_t\right\}$ por lo que no sería óptimo.

5. Predicción con ETMV.-

En la estimación de MCO obtenemos unas estimaciones únicas de los parámetros del modelo a partir de todas las observaciones T de la muestra. Sin embargo cuando utilizamos los ETMV, el resultado de los mismos es una secuencia de vectores de estimaciones $\left\{\vec{b}_t^*\right\}$, t=1,2,..., T para cada uno de los valores de la muestra. El problema de la predicción estriba en que, para una determinada observación muestral $\vec{x}_0^{\,'}$, tenemos que determinar un valor predicho de la variable endógena, y nos planteamos cual de entre los valores de la secuencia $\left\{\vec{b}_t^*\right\}$ escogemos.

Tenemos que tomar un criterio de selección del valor de $\vec{b}_0^*$ utilizado para llevar a cabo la predicción. Una alternativa puede consistir en tomar una media ponderada de los valores correspondientes a la secuencia de estimaciones óptimas, de manera que los vectores que cometen un menor error de estimación $e_t^* = y_t - \vec{x}_t' \vec{b}_t^*$ en valor absoluto tengan una ponderación mayor. Una propuesta pudiera ser la siguiente:

$$\vec{b}_0^* = \sum_{t=1}^{T} \left(\frac{\frac{1}{e_t^*}}{\sum_{t=1}^{T} \frac{1}{e_t^*}} \right) \vec{b}_t^*$$

Una segunda alternativa que también vamos a analizar consiste en lo siguiente: dado un vector de variables independientes $\vec{x}_0'$ seleccionaremos aquel vector $\vec{b}_o$ de la secuencia de estimadores óptimos, $\left\{\vec{b}_t^*\right\}$, t=1,2,..., T , que se corresponda con la posición para la cual la distancia de $\vec{x}_0'$ respecto de los vectores $\left\{\vec{x}_t'\right\}$, t=1,2,..., T en la muestra sea mínima, es decir: $\vec{b}_o = \vec{b}_t$, siendo $\boldsymbol{t}$ tal que:

$$\min_{t \in \{1,2,\ldots,t,\ldots,T\}} d(\vec{x}_t, \vec{x}_0)$$

siendo d la distancia euclidea convencional.

Intuitivamente, el fundamento de esta elección es claro. Para predecir de la mejor manera posible el valor de y_0 correspondiente a una determinada observación $\vec{x}_0{}'$ lo mejor es seleccionar, de entre el conjunto de estimaciones disponibles para los parámetros, a aquel vector $\vec{b}_o$ que corresponda con el seleccionado como el más adecuado para una observación que está muy próxima a $\vec{x}_0{}'$. En el siguiente apartado se muestran los resultados comparativos correspondientes a este método.

Una tercera alternativa que hemos analizado consiste en obtener la senda de estimadores óptimos pero omitiendo ahora en cada instante la observación correspondiente a ese periodo.

Finalmente a efectos de poder juzgar la capacidad predictiva de los estimadores propuestos se ha calculado también el estimador de jackknife definido como:

$$\vec{b}_{JK} = \frac{1}{T}\sum_{i=1}^{T}\tilde{b}_i$$

$$\tilde{b}_i = \left(\tilde{X}_i{}'\tilde{X}_i\right)^{-1}\tilde{X}_i{}'\tilde{y}_i$$

donde $\left(\tilde{X}_i{}', \tilde{y}_i\right)$ representan la matriz de observaciones de los regresores y de la variable dependiente respectivamente, una vez eliminada la observación i-ésima.

6. Análisis Empírico.-

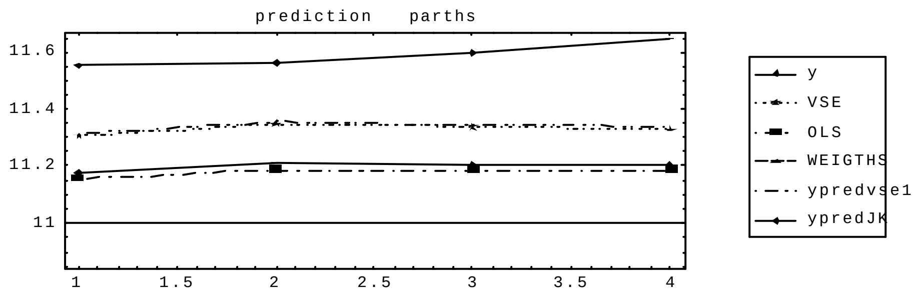
A efectos de analizar la capacidad predictiva de los ETMV se ha procedido a estimar la función de demanda de dinero $m^d = b_0 + b_1 p + b_2 y + b_3 R$ propuesta por Harris (1995) , donde las variables están tomadas en logaritmos, estando positivamente relacionada con el nivel de los precios y la renta y de manera negativa con el tipo de interés. Se utilizaron los datos de Ericsson et als (1992) para estimar esta función para el caso del reino unido, dejando como información extramuestral las últimas L observaciones, pudiendo de esta manera analizar de forma comparativa la capacidad predictiva de los ETMV frente al resto de estimadores propuestos en el epígrafe anterior. A continuación se muestran los resultados obtenidos por los ETMV para distintos horizontes de predicción.

Suma de Cuadrados de los residuos para el horizonte de predicción L
para distintos estimadores

L	1	2	3	4	5
VSE	0.071685	0.257016	0.531308	1.072560	1.787780

VSE1	0.068558	0.245165	0.502068	1.014190	1.687060
OLS	0.211182	0.758900	1.550680	2.692040	4.005270
JK	0.192063	0.684515	1.386920	2.404910	3.565300

A modo de ejemplo se compara en el gráfico siguiente las predicciones obtenidas por cada uno de los procedimientos comparados.



7. Conclusiones.-

Los estimadores ETMV se fundamentan en el concepto de proximidad, asignando para una determinada observación de las variables dependientes aquellos valores de los parámetros que se corresponden con el menor error cuadrático de entre el conjunto de todos los EMCO para todos los posibles subtamaños muestrales de uno dado. Este criterio nos permite obtener una senda óptima de valores ajustados de la variable dependiente que nos sirve para encontrar los ETMV. El error de predicción obtenido por este procedimiento es menor que el obtenido para estimadores alternativos como los EMCO o los estimadores de Jackknife. El error cuadrático de predicción mejora aún mas cuando se prescinde en el cálculo de la estimación correspondiente a esa observación de la información muestral contenida en la misma.

Bibliografía.-

Ericsson, N.R., Hendry, D.F. and H-A Tran (1992) . Cointegration, Seasonality, Ecompassing , and the Demand for Money in the United Kingdom, Discussion Paper, Board of Governors of the Federal Reserve System, Washington D.C.

Harris, Richard I.D. (1995) Using Cointegration Analysis in Econometric Modelling. Prentice Hall. Harvester Wheatsheaf.

Hendry, D.F, and Ericsson, N.R.,. (1991) Modelling the demand for narrow money in the United Kingdom and the United States, European Economic Review, 35, 833-81.

Poirier, D.J. (1995) Intermediate Statistics and Econometrics. A Comparative Approach. The MIT Press, Cambridge, Massachusetts.