Revista Colombiana de Estadística
Nº 17-18, 1988

MODELOS DE REGRESION CON PARAMETROS ESTOCASTICOS

Diego Escobar Uribe y Luz Myriam Echeverri
Departamento de Matemáticas
Universidad de los Andes

1. Introducción. Supóngase que tenemos un modelo de regresión que llamaremos aquí tradicional:

$$y_t = \sum_{i=1}^{k} \beta_i x_{it} + e_t \tag{1}$$

o en su forma matricial:

$$Y = X\beta + e \tag{2}$$

donde

$$Y = (y_1,\dots,y_T)^T,\quad X = \begin{bmatrix} x_{11}\cdots x_{k1} \\ \cdots\ \cdots\ \cdots \\ x_{1T}\cdots x_{kT} \end{bmatrix},$$

$$\beta = (\beta_1,\dots,\beta_k)^T \text{ y } e = (e_1,\dots,e_T)^T.$$

Normalmente este modelo de regresión múltiple se analiza bajo un conjunto de suposiciones estandar. A saber:

(i) Las variables independientes x_{ij}, o son valores fijos, o son variables aleatorias independientes de los terminos de error e_j.

(ii) $E(e_j) = 0 \quad \forall j$

(iii) $V(ej) = 0 \quad \forall j$ y en lo general es desconocida.

(iv) $\mathrm{COV}(e_i, e_j) = 0 \quad \forall i,j \quad i \neq j$.

(v) Para un tiempo t, las variables $x_{1t}, \ldots, x_{kt}$ son linealmente independientes.

Ahora bien, bajo estos supuestos se conoce por el teorema de Gauss-Markov que los estimadores de Mínimo Cuadrado (MC) de β son

$$b = (X^T X)^{-1} X^T Y. \tag{3}$$

Estos son BLU(Best Linear Unbiased Estimates). A su vez σ^2 se estima por medio de

$$s^2 = (y-Xb)^T (y-Xb)/T-k \tag{4}$$

Usualmente se comprueban las suposiciones descritas, sobre todo con respecto a la homoscedasticidad y a la no correlación (Prueba de Durbin-Watson). Si falla alguna de ellas, los estimadores de MC pueden llevar a conclusiones graves y dejan de tener propiedades deseables. Las predicciones pueden ser bastante imprecisas y por todo ello se recurre a otros métodos.

El modelo que queremos estudiar se basa fundamentalmente en la monografía de Newbold y Bos (1985) y difiere adicionalmente del tradicional en un aspecto más: en (1) los coeficientes β_i son fijos en el tiempo. Vamos a suponer sin embargo que estos parámetros son una Serie de Tiempo y por lo tanto nuestro modelo tiene ahora la forma

$$y_t = \sum_{i=1}^{K} \beta_{it} X_{it} + e_t. \tag{5}$$

Como una posible justificación de lo anterior podríamos, por ejemplo, tomar la relación

$$C_t = \alpha + \beta Y_t + e_t \tag{6}$$

que denota la dependencia del consumo del ingreso. Al asumir en (6) β fijo, se está diciendo que $dC_t/dY_t = \beta$, es decir que la propensión a consumir no varía con el tiempo. Es sin embargo plausible que dicha propensión cambie a través de los años por la composición del consumo, por un cambio en los gustos etc. y aún más, que esté relacionado el de hoy con el de ayer y así con los anteriores.

Antes de proceder a desarrollar el modelo (5) propuesto, necesitamos recapitular algunos conceptos básicos sobre Series de Tiempo.

2. Series de Tiempo.

Definición 1.

a) Una Serie de Tiempo β_t se llama *estacionaria* si cumple
   i) $E(\beta_t) = \mu \quad \forall t$
   ii) $V(\beta_t) = \sigma^2 = \gamma(0) \quad \forall t$
   iii) $COV(\beta_t, \beta_{t+\tau}) = \gamma(\tau)$, es decir depende de la distancia en el tiempo τ.

b) La *función de autocorrelación* se define como
   $\rho(\tau) = \gamma(\tau)/\gamma(0)$.
   La *autorrelación muestral* a su vez se define como
   $r(\tau) = c(\tau)/c(0)$

donde los c's son la covarianza y varianza muestral respectivamente.

La gráfica de $r(\tau)$ con respecto a τ se llama un *correlograma*.

Definición 2.

Una Serie de Tiempo e_t se llama de *ruido blanco* sí, tiene media 0, varianza σ^2 y no esta correlacionada, es decir $\gamma(p)=0$ $\forall p \neq 0$.

Utilizando las definiciones anteriores y pensando en los modelos que deaseamos estudiar, queremos que los parámetros β_{it} de (5) sean autocorrelacionados. Un modelo que nos refleja esta situación es el llamado Autorregresivo.

Definición 3.

La Serie y_t se llama un modelo *autorregresivo* de orden p, $AR(p)$, si es de la forma

$$y_t - \mu = \phi_1(y_{t-1}-\mu) + \phi_2(y_{t-2}-\mu)+\ldots+\mu_p(y_{t-p}-\mu) + a_t \qquad (7)$$

donde a_t es de ruido blanco y los σ's son parámetros fijos.

Es importante anotar que imponiendo ciertas condiciones, estas series son estacionarias. No viene al caso deducir aquí estas propiedades que se pueden encontrar en cualquier buen libro de Series de Tiempo (Fuller [1976]). Sin embargo anotamos que dado que la complejidad va en aumento con el orden p, muchos ejemplos y aplicaciones se trabajan con series $AR(1)$ o $AR(2)$. En especial para la predicción se tiene que si estamos situados en el tiempo T, futuros valores y_{T+1} dependen sólo de y_T si estamos en presencia de un modelo $AR(1)$, y, sólo de y_T y y_{T-1} si es un modelo $AR(2)$.

Hasta este punto hemos introducido modelos que muestran la evolución a través del tiempo de una sola serie.

Supongamos ahora que tenemos observaciones en el tiempo sobre un conjunto de K series. Para ello definimos $y_t^T = (y_{1t},\ldots, y_{Kt})$ como el vector observado y asumimos que se trata de un caso estacionario. Esto significa que tenemos un vector fijo de medias $\mu^T = (\mu_1,\ldots,\mu_K)$, una matriz fija de covarianza $\sum$ y que para todo i,j la covarianza entre y_{it} y $y_{j,t-1}$ depende sólo de 1. Una generalización natural de (7) en el caso de una serie $AR(1)$ puede entonces escribirse en la forma matricial

$$y_t - \mu = \Phi(y_{t-1}-\mu) + a_t \tag{9}$$

donde Φ es una matriz $K\times K$ de coeficientes fijos y a_t es un vector K dimensional de ruido blanco que cumple

$$E(a_t) = 0, E(a_t a_t^T) = \Omega_a \quad \text{y} \quad E(a_t a_{t-j}^T) = 0 \quad \forall j \neq 0. \tag{10}$$

Dos conclusiones son de importancia para lo que sigue: La forma (9) y las propiedades de a_t nos llevan facilmente a concluir que

1) $$\sum = \Phi\sum\Phi^T + \Omega_a \tag{11}$$

2) Si Φ es diagonal, (9) nos dice que cada serie individual es del tipo $AR(1)$. Por lo tanto para la predicción sólo se depende de su propio pasado. Si Φ no lo es entonces, la predicción $y_{i,T+1}$ puede depender de todas los $y_{i,T}$.

Con las herramientas introducidas hasta este punto sobre Series de Tiempo vamos ahora a pasar a formular el modelo sobre el cual queremos trabajar.

3. Estimación y Predicción para un Modelo Estocástico

Suponemos un proceso como en (5), que se puede escribir en forma matricial como

$$y_t = x_t^T\beta_t + e_t \tag{12}$$

Se asume adicionalmente que el vector β_t es un proceso del tipo autoregresivo de primer orden, es decir tiene la forma

$$\beta_t - \mu = \Phi(\beta_{t-1}-\mu) + a_t \tag{13}$$

Φ es una matriz $K\times K$, a_t y μ vectores $K\times 1$, $\beta_t^T = (\beta_{1t},\ldots,\beta_{Kt})$, $x_t^T = (x_{1t},\ldots,x_{Kt})$. Se dispone de T observaciones sobre las variables y y x y se supone que tanto e_t como a_t son de ruido blanco (e_t como en la Def 2 y a_t como en (10)). Además no son correlacionadas entre ellas. Finalmente se tiene para $\beta_t, E(\beta_t) = \mu$ y matriz de covariante $\sum$.

La ecuación (12) puede escribirse en la forma

$$y_t = x_t^T\beta_t + e_t = x_t^T\mu + x_t^T(\beta_t-\mu) + e_t = x_t^T\mu + u_t \tag{14}$$

donde $u_t = x_t^T(\beta_t-\mu) + e_t$.

La última ecuación muestra que nuestro modelo puede interpretarse como un caso de parámetro fijo β, pero con un término de error u_t que no posee las condiciones clásicas puesto que exhibe tanto hetereocedasticidad como autocorrelación, es decir no cumple ni (iii) ni (iv) del modelo tradicional. En efecto, dado que las β_t son autocorrelacionadas, también lo serán las u_t. Para la varianza de u_t se tiene

$$V(u_t) = x_t^T \Sigma x_t + \sigma^2 .$$

Dado las propiedades que exhibe u_t, el método de estimación que usualmente se aplica es el de Mínimos Cuadrados Generalizados. Sin embargo aquí se presenta el problema que para aplicarlo, se necesita Σ, que no conocemos.

La solución que se propone es utilizar lo que se conoce en la literatura como modelos de estado-espacio y aplicar la técnica de Filtros de Kalman. Esta permite establecer bajo el supuesto adicional de normalidad de los terminos de error, la función de verosimilitud a partir de la cual se estiman los parámetros fijos μ, σ^2, Φ y Ω_a.

3.1. El Filtro de Kalman.

Suponemos un modelo dado por la especificación (12)-(13). El propósito de la técnica que vamos a describir es establecer un sistema computacional que permite generar la media μ y matriz de covarianza Σ de la serie β_t, dada información en el tiempo t. Introducimos la siguiente notación: sean $\beta(t|n) := E(\beta_t|y_1,\ldots,y_n)$ y $P(t|n) := V(\beta_t|y_1,\ldots,y_n)$ media y varianza de β_t condicionadas a la información en el tiempo n.

En forma similar definimos, $y(t|t-1) = E(y_t|y_1,\ldots,y_{t-1})$ y $h_t := V(y_t|y_1,\ldots,y_{t-1})$.

Ahora bien, dado la suposición de normalidad y tomando expectativas condicionadas en nuestro modelo (12) y (13) obtenemos las ecuaciones que caracterizan el Filtro de Kalman: para $t = 1,\dots,T$

$$\beta(t|t-1) = \mu + \Phi|\beta(t-1|t-1)-\mu| = \Phi\beta(t-1|t-1) + [I-\Phi]\mu \quad (14.1)$$

$$P(t|t-1) = \Phi P(t-1|t-1)\Phi^T + \Omega_a \quad (14.2)$$

$$y(t|t-1) = x_t^T\beta(t|t-1) \quad (14.3)$$

$$h_t = x_t^T P(t|t-1)x_t + \sigma^2 \quad (14.4)$$

Igualmente e utilizando propiedades de la distribución normal multivariada, (Wilks [1962]) completamos el conjunto de ecuaciones anteriores con

$$\beta(t|t) = \beta(t|t-1) + P(t|t-1)x_t h_t^{-1}\left[y_t - x_t^T\beta(t|t-1)\right] \quad (14.5)$$

$$P(t|t) = P(t|t-1) - P(t|t-1)x_t h_t^{-1}x_t^T P(t|t-1) \quad (14.6)$$

El proceso o filtro se inicia dándole valores a $\beta(0|0)$ y $P(0|0)$ y calculando los restantes valores esperados y varianzas a partir de las ecuaciones (14) en forma recursiva. Como los β_t son por la especificación (13) un proceso $AR(1)$,

$$\beta(0|0) = \mu$$

y $P(0|0)$ se obtiene por (11) como solución del sistema lineal

$$P(0|0) = \Phi P(0|0)\Phi^T + \Omega_a. \quad (15)$$

El paso siguiente es establecer la función de verosimilitud basándose en la información obtenida recursivamente por el Filtro de Kalman. Para ello utilizamos la siguiente propiedad:

$$f(y_1,\dots,y_T) = f(y_1)f(y_2|y_1)f(y_3|y_2,y_1)\dots f(y_T|y_{T-1},\dots,y_1)$$

Por (14.3) y (14.4) se conoce además que $y_t | y_{t-1},\dots,y_1$ es $N(x_t^T\beta(t|t-1),h_t)$ y por lo tanto

$$\log L = \log f(y_1,\dots,y_t)$$
$$= -T/2 \log 2\pi - \tfrac{1}{2}\sum_{t=1}^{T} \log h_t - \tfrac{1}{2}\sum_{t=1}^{T}\left[y_t - x_t^T\beta(t|t-1)\right]^2 h_t^{-1}. \tag{16}$$

Maximizar log L o lo que es lo mismo minimizar

$$G = \tfrac{1}{2}\sum_{t=1}^{T} \log h_t + \tfrac{1}{2}\sum_{t=1}^{T}\left[y_t - x_t^T\beta(t|t-1)\right]^2 h_t^{-1} \tag{17}$$

con respecto a μ, σ^2, Φ y Ω nos arroja estimadores para estos parámetros.

Como se observa G involucra para $t = 1,\dots,T$ tanto $\beta(t|t-1)$ como también $P(t|t-1)$ (por medio de h_t) y estos a su vez son funciones de los parámetros a estimar. Resolver analíticamente este problema de minimización es prácticamente imposible y como veremos en el ejemplo posterior necesitamos recurrir a métodos numéricos. Sin embargo antes de exponer éstos queremos exponer brevemente como se presenta el problema de la predicción y la estimación de β_t.

3.2. Predicción y Estimación de β_t.

El propósito de la predicción es dar o indicar futuros valores, basados en la información conocida en un momento determinado del tiempo. En nuestro caso queremos predecir tanto los futuros valores de los parámetros estocásticos β_t así como también de la variable dependiente y_t.

Si se conocen los parámetros fijos que intervienen en el modelo (12)-(13), entonces lo que deseamos conocer es

$$\beta(T+1|T) = \Phi\beta(T+1-1|T)+|I-\Phi|\mu; \quad P(T+1|T) = \Phi P(T+1-1|T)\Phi^T+\Omega \tag{18}$$

$$y(T+1|T) = x_{T+1}^T\beta(T+1|T); \quad h_{T+1} = x_{T+1}^T P(T+1|T)x_{T+1}+\sigma^2 \tag{19}$$

que se obtienen directamente de las ecuaciones (14). Naturalmente no se conocen los valores de los parámetros y debemos utilizar los estimadores de máxima verosimilitud anteriormente calculados.

Finalmente queremos estimar los parámetros β_t sobre el período $t = 1,\ldots,T$. El estimador debe basarse en toda la información que se tiene y por lo tanto el estimador óptimo de β_t resulta ser $\beta(t|T)$. De igual manera para conocer los errores estándar se requiere de $P(t|T)$.

El algoritmo empleado anteriormente no da estos valores puesto que solo arroja media y varianza de β_t dada información hasta el tiempo t. Solo conocemos por lo tanto $\beta(T|T)$ y $P(T|T)$ por las ecuaciones (14). La solución nos la ofrece Ansley y Kohn [1982], donde de nuevo utilizando propiedades de la distribución multivariada normal se obtiene

$$\beta(t|T) = \beta(t|t)+P(t|t)\Phi^T P(t+1|t)^{-1}[\beta(t+1|T)-\beta(t+1|t)] \tag{20}$$

$$\begin{aligned} P(t|T) = {} & P(t|t)-P(t|t)\Phi^T P(t+1|t)^{-1}\cdot \\ & \cdot|P(t+1|t)-P(t+1|T)]P(t+1|t)^{-1}\Phi P(t|t) \end{aligned} \tag{21}$$

El cálculo se inicia poniendo $t = T-1$ y utilizando las ecuaciones (14), se obtienen recursivamente con $t = T-2, T-3,\ldots$ estimadores para $\beta(t|T)$ y $P(t|T)$. Nótese que de nuevo es necesario emplear para los valores de los parámetros fijos los estimadores de máxima verosimilitud.

4. Ejemplo de un Modelo de Inversión.

El modelo que queremos tratar tiene la siguiente especificación

$$I_t = b_t + c_t Y_t + d_t R_{t-1} + e_t \tag{22}$$

donde e_t es $N(0,\sigma^2)$, I_t denota la inversión (o Formación Bruta de Capital), Y_t el ingreso (o PIB) y R_t una tasa de interés. Definimos $\beta_t := (b_t, c_t, d_t)^T x_t := (1, Y_t, R_{t-1})^T$. Suponemos adicionalmente que β_t sigue un proceso $AR(1)$, es decir que

$$\beta_t - \mu = \Phi(\beta_{t-1} - \mu) + a_t \tag{23}$$

donde a_t es $N(0,\Omega_a)$, no hay autocorrelación y tampoco está correlacionada con el término de error de (22)

La especificación dada por (22) y (23) sigue los lineamien mientos trazados por el sistema (12)-(13) y desde el punto de vista de su interpretación económica podría arguirse que por ejemplo por cambios tecnológicos los parámetros no son fijos.

Sin embargo y antes de embarcarse en un modelo y especificarlo con parámetros estocásticos es deseable correr algunas pruebas de hipótesis que nos verifique el supuesto de estar efectivamente en presencia de un modelo no tradicional. Varias pruebas han sido propuestas para este efecto, cuya descripción trasciende el alcance de este trabajo. Dejamos como referencia en especial, la sinópsis en Chow [1983] donde se describen siete posibles alternativas con sus respectivas notas bibliográficas y los tests descritos en el artículo base de Newbold y Bos [1985]. De igual manera este trabajo sólo trató el caso en el cual los parámetros estocásticos siguen un proceso autorregresivo. Es posible extender la metodología empleada a procesos más complejos como los llamados modelos ARMA (autoregresivos y de promedio movil).

Ahora bien para simplificar un tanto las cosas vamos a suponer que tanto Φ como Ω_a son inicialmente ambas matrices diagonales con $|\phi_{ii}| < 1$. Adoptando la notación de (14) tenemos entonces que la matriz de covarianza de los u_t viene dada por

$$E(u_t u_s) = \sum_{i=1}^{3} x_{it} x_{is} \omega_{ii} \phi_{ii}^{\tau} / (1-\phi_{ii}^2) \tag{24}$$

donde $\tau = t-s$ es el rezago.

Queremos obtener estimadores para los valores de Φ y Ω_a. Harvey [1985] sugiere proceder de la siguiente forma: reemplazar $E(u_t u_s)$ por $\hat{u}_t \hat{u}_s$ y estimar por MC las siguientes dos regresiones:

$$\hat{u}_t^2 = c_1 + c_2 x_{2t}^2 + c_3 x_{3t}^2 + \varepsilon_t \tag{25}$$

$$\hat{u}_t \hat{u}_{t-1} = d_1 + d_2 x_{2t} x_{2,t-1} + d_3 x_{3t} x_{3,t-1} + \varepsilon_t \tag{26}$$

Resolviendo por (24) las ecuaciones $c_i = \omega_{ii}/(1-\phi_{ii}^2)$ y $d_i = \omega_{ii}\phi_{ii}/(1-\phi_{ii}^2)$ para ω_{ii} y ϕ_{ii} se obtinen estimadores para estos parámetros. Tenemos por lo tanto gasta este punto valores iniciales para los elementos de la diagonal de las matrices ϕ y Ω_a.

Nuestro objetivo es minimizar la función G de (17). Para obtener valores iniciales de $E(\beta_t) = \mu$ y de σ^2 es natural estimar la ecuación (22) omitiendo el carácter estocástico de esta especificación por MC. Es decir hacemos de cuenta que tenemos un modelo tradicional. El resultado específico de este ejercicio arrojó los siguientes resultados:

$$\hat{I}_t = -4.97 + \underset{(0.019)}{0.19 Y_t} - \underset{(1.72)}{4.88 R_{t-1}}$$

$$R^2_{ajus} = 0.93 \qquad \hat{\sigma}^2 = 20.31$$

$$\phi_{11} = 0.3357 \quad \phi_{22} = -0.79 \quad \phi_{33} = 0.6376 \quad \omega_{11} = 8.855$$

$$\omega_{22} = 0.0000085 \quad \omega_{33} = 0.1243.$$

La parte final de este trabajo es utilizar los resultados obtenidos como valores iniciales en la minimización de la función G de (17) y analizar los resultados. Como ya se dijo esto es esencialmente un problema numérico, por cierto bastante complejo, en el cual aún estamos trabajando.

*

BIBLIOGRAFIA

Ansley, C.F. y Kohn, R., (1982). A geometrical derivation of the fixed interval smoothing algorithm. Biometrika 69.

Chow, G., (1983). Econometrios Mac-Graw Hill, New York.

Fuller, W.A.(1976). Introduction to Statistical Time Series. J. Wiley, New York.

Harvey, A.C.(1981). Time Series Models, Phillip Allan, London.

Newbold, P. y Bos, Th., (1985). Stochastic Parameter Regression Models. Sage University Papers Nº 51.

Wilks, S., (1962). Mathematical Statistics. J. Wiley New York.

*