

CAPÍTULO 2
REGRESORES ESTOCÁSTICOS.
MODELOS PARCIALMENTE NO ESTACIONARIOS.
MÉTODOS DE MONTE-CARLO.

2.1 INTRODUCCIÓN.

Buena parte de los desarrollos presentados en el capítulo anterior se han obtenido asumiendo el cumplimiento de las Hipótesis 2 y 4.

La Hipótesis 2 establece que los regresores no son estocásticos. Este supuesto facilita la derivación de las propiedades de los estadísticos definidos para estimar y evaluar los diferentes parámetros, tanto en muestras pequeñas como en el marco asintótico.

Su incumplimiento dificulta tales derivaciones siendo necesario añadir determinados supuestos sobre la dependencia estocástica entre el vector de perturbaciones y los regresores. Según sea esta dependencia, las propiedades de los estadísticos difieren. La Sección 2 de este capítulo está dedicada a analizar las cuestiones relacionadas con estos temas.

La Hipótesis 4 se refiere a que los promedios de los cuadrados de las observaciones de las variables así como los de sus productos cruzados tienden a constantes finitas. Cuando esta hipótesis se incumple, entonces esas sumas hay que dividirlos por potencias de T superiores a 1, en muchos casos no coincidentes. Cuando esto es así sólo cabe un razonamiento analítico en el marco asintótico, por lo que es necesario hacer uso de los Métodos de Monte-Carlo para derivar resultados válidos para muestras finitas. Todos estos puntos serán tratados en la Sección 3.

Por último, la Sección 4 está dedicada a los Métodos de Monte-Carlo. La utilidad de los mismos se pondrá de manifiesto en las dos secciones anteriores en las que la mayor parte de los resultados derivados lo son asumiendo una muestra de tamaño infinito.

2.2 REGRESORES ESTOCÁSTICOS.

En esta Sección se van a mantener las hipótesis comentadas en la Sección 1.2, excepto la Hipótesis 2 que ahora cambiaremos por la siguiente hipótesis:

Hipótesis 2': Los valores de X son la observación-realización de un proceso estocástico.

El cambio a esta nueva hipótesis tiene implicaciones importantes a la hora de derivar las propiedades de los estimadores MCO.

Por ejemplo, cuando se estudiaba la insesgadez del estimador MCO en el Resultado 1.4 se aplicaban dos reglas que se justificaban por el carácter no estocástico de los regresores. Estas dos reglas permitían escribir:

$$E(X'X)^{-1}X'u = E(X'X)^{-1}E(X'u)$$

y
$$E(X'u) = 0$$

Cuando los regresores son estocásticos el uso de estas reglas sólo es posible bajo ciertos supuestos acerca de la dependencia del vector de residuos y los regresores. Cuando estos supuestos no se cumplen sólo cabe pensar en otras reglas especialmente pensadas para un marco asintótico ya que cuando la muestra es grande, se pueden sustituir las dos reglas anteriores por las siguientes:

$$\text{plim}(X'X)^{-1}X'u = \text{plim}(X'X)^{-1}\text{plim}(X'u)$$

y
$$\text{plim}X'u = 0$$

En todo caso, todo depende de los supuestos que se hagan acerca de la dependencia entre el vector de perturbaciones y regresores.

Por esta razón, vamos a distinguir tres situaciones diferentes según sea la dependencia:

- (a) Independencia entre regresores y vector de perturbaciones.
- (b) Independencia contemporánea pero dependencia temporal.
- (c) Dependencia contemporánea.

(a) Independencia

La independencia se concreta en:

Hipótesis 2'(i): El vector de perturbaciones y los regresores son independientes pudiéndose escribir:

$$E(u/X) = 0 \quad (2.1)$$

Veremos ahora algunas de las propiedades de los estimadores MCO.

Resultado 2.1: El vector de estimadores MCO de β , $\hat{\beta}$ es insesgado con una matriz de varianzas y covarianzas igual a:

$$\text{Var}(\hat{\beta}) = \sigma^2 E_x (X'X)^{-1} \quad (2.2)$$

En donde E_x denota la esperanza respecto a la distribución marginal de X .

Prueba: Comenzaremos haciendo referencia a un resultado estándar en estadística que vamos a utilizar en esta sección y las que siguen. El resultado dice así: Sean u y v dos variables aleatorias distribuidas conjuntamente. Entonces se cumple que:

$$E[h(u, v)] = E_u \left[E_{v/u} h(u, v) \right] \quad (2.3)$$

En donde $E_{v/u}$ indica la esperanza de la distribución condicional de v dada u y E_u representa la esperanza con respecto a la distribución marginal de u .

Utilizando (2.3) se tiene que:

$$\begin{aligned} E\hat{\beta} &= E_x \left\{ E \left(\beta + (X'X)^{-1} X'u/X \right) \right\} = \\ &= E(\beta) = \beta \end{aligned}$$

demostrándose la insesgadez.

En cuando a la matriz de varianzas y covarianzas hay que tener en cuenta que:

$$\begin{aligned} \text{Var}(\beta) &= E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \left\{ E_x \left[E(X'X)^{-1} X'uu'X(X'X)^{-1} / X \right] \right\} = \\ &= E_x \left\{ (X'X)^{-1} X' E(uu'/X) X(X'X)^{-1} \right\} = \\ &= E_x \left\{ (X'X)^{-1} X' (\sigma^2 I_T) X(X'X)^{-1} \right\} = \\ &= \sigma^2 E_x (X'X)^{-1} \end{aligned} \quad \square$$

Resultado 2.2 : El estimador MCO de σ^2 es insesgado.

Prueba: $E\hat{\sigma}^2 = E_x [E(\hat{\sigma}^2/X)] = E_x(\sigma^2) = \sigma^2$



Si se sigue manteniendo la Hipótesis 8 que establece la normalidad del vector u , entonces condicionados sobre X , los estimadores $\hat{\beta}$ y $\hat{\sigma}^2$ tienen las mismas distribuciones de probabilidad comentadas en el capítulo anterior. Por ser estimadores máximo-verosímiles son consistentes y asintóticamente eficientes.

Si la hipótesis de normalidad no se cumple entonces para poder demostrar que los estimadores son consistentes y asintóticamente eficientes es necesario asumir nuevas hipótesis acerca del proceso estocástico multivariante que genera los valores de X . Ver Fomby, Hill y Johnson (1984, Capítulo 5) y Hamilton (1994, Capítulo 8) para más detalles.

(b) Independencia Contemporánea pero Dependencia Temporal.

Dentro de este apartado vamos a estudiar el modelo autoregresivo que escribimos como:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + u_t \quad (2.4)$$

o, con notación matricial

$$y = X \beta + u$$

en donde ahora:

$$X = \begin{bmatrix} y_0 & y_{-1} & \dots & y_{-p+1} \\ y_1 & y_0 & \dots & y_{-p} \\ \dots & \dots & \dots & \dots \\ y_{T-1} & y_{T-2} & \dots & y_{T-p} \end{bmatrix}$$

Para la perturbación u_t mantendremos las mismas hipótesis comentadas en la Sección 1.2, añadiendo que el cuarto momento de u_t es finito.

Respecto a la parte sistemática, añadiremos la hipótesis que garantiza la estacionariedad:

Hipótesis 2'(ii): Todas las raíces del polinomio

$$1 - \phi_1 \lambda - \phi_2 \lambda^2 - \dots - \phi_p \lambda^p = 0 \quad (2.5)$$

están fuera del círculo unitario.

Cumplida ésta hipótesis, el proceso puede escribirse como:

$$y_t = \psi(L)u_t \quad (2.6)$$

en donde $\psi(L)$ es un polinomio de retardos de orden infinito que es convergente en error cuadrático medio.

A partir de (2.6) y_{t-j} depende de $u_{t-j}, u_{t-j-1}, \dots$ pero es independiente de u_t . Esto justifica el título del apartado dependencia temporal pero de independencia contemporánea.

Resultado 2.3: La matriz

$$Q = \text{plim} \left(\frac{X'X}{T} \right) \quad (2.7)$$

tiene elementos finitos y es no singular.

Prueba:

$$T^{-1}(X'X) = \begin{bmatrix} 1 & T^{-1} \sum y_{t-1} & \dots & T^{-1} \sum y_{t-p} \\ T^{-1} \sum y_{t-1} & T^{-1} \sum y_{t-1}^2 & \dots & T^{-1} \sum y_{t-1} y_{t-p} \\ \dots & \dots & \dots & \dots \\ T^{-1} \sum y_{t-p} & T^{-1} \sum y_{t-p} y_{t-1} & \dots & T^{-1} \sum y_{t-p}^2 \end{bmatrix} \quad (2.8)$$

Los elementos de la primera fila o columna tienden a la media del proceso que, en este caso, es cero. Por otro lado, tal como puede verse en el Capítulo 7 de Hamilton (1994), también se tiene que:

$$\frac{1}{T} \sum y_{t-i} y_{t-j} \xrightarrow{P} E y_{t-i} y_{t-j} = \gamma_{(i-j)}$$

Llegamos así al resultado pudiéndose escribir:

$$Q = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ 0 & \gamma_0 & \dots & \dots & \gamma_{p-1} \\ \vdots & \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_0 \end{bmatrix}$$

Se ve como los elementos de Q, exceptuando los de la primera fila y columna son los de la función de autocovarianza del proceso. □

Resultado 2.4 : La distribución de probabilidad asintótica de $\frac{1}{\sqrt{T}} X' u$ viene dada por:

$$\frac{1}{\sqrt{T}} X' u \xrightarrow{d} N(0, \sigma^2 Q) \quad (2.9)$$

Prueba: Tomaremos uno de los p elementos del vector que escribiremos como:

$$\frac{1}{\sqrt{T}} \sum y_t - j^u_t \quad (2.10)$$

La forma que toma (2.10) nos indica que se trata del promedio de una martingala en diferencias. Bastaría aplicar el Resultado A.18 del Anexo 1 para completar la prueba.

Hay que tener en cuenta que:

$$E\left(y_t - j^u_t\right)^2 = \sigma^2 E y_{t-j}^2$$

y que:
$$\frac{1}{T} \sum \left(y_t - j^u_t\right)^2 \xrightarrow{p} \sigma^2 E y_{t-j}^2$$

tal como puede verse en el Ejemplo 7.15 de Hamilton (1994). □

Utilizando todos estos resultados se llega a la distribución de $\sqrt{T} (\hat{\beta} - \beta)$

Resultado 2.5 : La distribución asintótica de probabilidad de $\sqrt{T} (\hat{\beta} - \beta)$ viene dada por:

$$\sqrt{T} (\hat{\beta} - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1}) \quad (2.11)$$

Prueba : Basta tomar en cuenta el Resultado A.12 del Anexo 1.

La consecuencia inmediata de este resultado es que el estimador $\hat{\beta}$ es consistente. □

$$\begin{aligned}\text{plim}\hat{\beta} &= \text{plim}\left(\beta + (X'X)^{-1}X'u\right) = \\ &= \beta + \left(\text{plim}\frac{X'X}{T}\right)^{-1} \text{plim}\frac{X'u}{T} = \\ &= \beta\end{aligned}$$

ya que por (2.7)

$$\left(\text{plim}\frac{X'X}{T}\right)^{-1} = Q^{-1}$$

Por otra parte, tomando el elemento genérico de $\frac{X'u}{T}$ se tiene que su esperanza es:

$$E\left(\frac{1}{T} \sum_{t=1}^T y_{t-i} u_t\right) = \frac{1}{T} \sum E y_{t-i} E u_t = 0$$

y su varianza puede escribirse:

$$\begin{aligned}E\left(\frac{1}{T} \sum_{t=1}^T y_{t-k} u_t\right)^2 &= \frac{1}{T^2} \left[\sum E y_{t-k}^2 E u_t^2 + \right. \\ &\left. + 2 \sum_{\substack{i \\ i < j}} \sum_j E(u_{t-i}) E(y_{t-k-i} y_{t-k-j} u_{t-j}) \right]\end{aligned}$$

El segundo término de la derecha es cero. El límite del primero es:

$$\frac{1}{T} \frac{1}{T} T \gamma_0 \sigma^2 \approx \frac{\gamma_0 \sigma^2}{T} \xrightarrow{p} 0$$

La esperanza cero y la varianza en el límite cero son condiciones suficientes para la convergencia en error cuadrático medio y, como consecuencia para la convergencia en probabilidad, tal como puede verse en el Anexo 1.

Ejemplo 2.1 : Consideremos el proceso autoregresivo de primer orden con media cero:

$$y_t = \phi y_{t-1} + u_t \quad (2.12)$$

En este caso, el cumplimiento de la Hipótesis 2'(ii) implica que:

$$|\phi| < 1 \quad (2.13)$$

El proceso puede escribirse como:

$$\begin{aligned} y_t &= \phi y_{t-1} + u_t = \phi^2 y_{t-2} + u_t + \phi u_{t-1} = \\ &= \dots = \sum_{i=0}^{\infty} \phi^i u_{t-i} \end{aligned} \quad (2.14)$$

Claramente se ve que y_{t-j} depende de u_{t-j} y perturbaciones de periodos anteriores pero no de u_t u otras posteriores a $t-j$.

El estimador MCO de ϕ es:

$$\hat{\phi} = \frac{\sum y_t y_{t-1}}{\sum y_{t-1}^2} = \phi + \frac{\sum y_{t-1} u_t}{\sum y_{t-1}^2} \quad (2.15)$$

En este caso se tiene que:

$$\frac{1}{T} X'X = \frac{\sum y_{t-1}^2}{T}$$

Por lo que Q será:

$$\frac{\sum y_{t-1}^2}{T} \xrightarrow{P} Q = \gamma_0 = \frac{\sigma^2}{1-\phi^2}$$

Por lo tanto, la distribución asintótica viene dada por:

$$\sqrt{T}(\hat{\phi} - \phi) \sim N[0, (1-\phi^2)] \quad (2.16)$$

La consistencia se deriva de:

$$\hat{\phi} = \phi + \frac{\sum y_{t-1} u_t / T}{\sum y_{t-1}^2 / T} \quad (2.17)$$

Respecto al denominador del segundo término de la derecha ya hemos demostrado que tiende a γ_0 .

Para el numerador se tiene que su esperanza es igual a cero por ser independientes y_{t-1} y u_t .

$$E\left(\sum y_{t-1} u_t\right) = \sum E y_{t-1} E u_t = 0$$

y su varianza se puede escribir:

$$\begin{aligned} E\left(\frac{1}{T} \sum y_{t-1} u_t\right)^2 &= \frac{1}{T^2} \left(\sum E y_{t-1}^2 E u_t^2 \right) + \\ &+ \frac{2}{T^2} \sum_i \sum_{j \atop i < j} E(u_{t-i}) E(y_{t-i-1} y_{t-j-1} u_{t-j}) \end{aligned}$$

Claramente se ve que el segundo término de la derecha es cero. Respecto al primero:

$$\frac{1}{T^2} \sum E y_{t-1}^2 E u_t^2 = \frac{1}{T} \left[\frac{1}{T} \sum \gamma_0 \sigma^2 \right] = \frac{\gamma_0 \sigma^2}{T} \rightarrow 0$$

Por lo tanto, la esperanza es cero y la covarianza tiende a cero por lo que hay convergencia en error cuadrático medio, y, como consecuencia, hay convergencia en probabilidad.

c) Dependencia Contemporánea.

Hemos visto que, en las dos situaciones anteriores, los estimadores MCO mantenían alguna de las propiedades derivadas en el caso en que los regresores no eran estocásticos. Cuando existía independencia se mantenían todas las propiedades y cuando se tenía dependencia temporal siendo contemporáneamente independientes se mantenían las propiedades asintóticas.

Cuando existe dependencia contemporánea entre perturbación y regresores los estimadores MCO ya no mantienen ninguna de las propiedades y se hace necesario pensar en estimadores alternativos que recuperen alguna de esas propiedades: ser insesgados, consistentes, etc. La solución ha consistido en utilizar los llamados estimadores de variable instrumental. Veamos a continuación, algunos ejemplos para ilustrar esta problemática.

Ejemplo 2.2: Considerar el modelo autorregresivo del Ejemplo 2.1 pero asumiendo que la perturbación sigue, a su vez, una proceso autorregresivo de primer orden:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (2.18)$$

con $\varepsilon_t \sim \text{iid } N(0, \sigma^2)$ y $|\rho| < 1$

El estimador MCO de ϕ sigue siendo el escrito en (2.17). En este caso tenemos:

$$\begin{aligned} E y_{t-1} u_t &= E(u_{t-1} + \phi u_{t-2} + \phi^2 u_{t-3} + \dots) u_t = \\ &= \gamma_{u_1} + \phi \gamma_{u_2} + \phi^2 \gamma_{u_3} + \dots \end{aligned} \quad (2.19)$$

En donde γ_{u_j} es la función de autocovarianza de u que, a partir, de (2.18) puede escribirse:

$$\begin{aligned} \gamma_{u_0} &= \frac{\sigma^2}{1-\rho^2} \\ \gamma_{u_j} &= \rho^j \gamma_{u_0} = \frac{\rho^j \sigma^2}{1-\rho^2} \end{aligned}$$

Utilizando estos resultados, (2.19) puede escribirse:

$$\begin{aligned} E y_{t-1} u_t &= \frac{\sigma^2}{1-\rho^2} [\rho + \phi \rho^2 + \phi^2 \rho^3 + \dots] = \\ &= \frac{\sigma^2 \rho}{(1-\rho^2)(1-\phi\rho)} \end{aligned}$$

Por lo tanto:

$$\frac{1}{T} \sum y_{t-1} u_t \xrightarrow{p} \frac{\sigma^2 \rho}{(1-\rho^2)(1-\phi\rho)}$$

En lo que respecta al denominador de (2.17):

$$E y_{t-1}^2 = E(\phi y_{t-2} + u_{t-1})^2 = \phi^2 E y_{t-2}^2 + E u_{t-1}^2 + 2\phi E y_{t-2} u_{t-1}$$

y por ser estacionario:

$$E y_{t-1}^2 = E y_{t-2}^2 = \gamma_0$$

$$\gamma_0(1 - \phi^2) = \frac{\sigma^2}{1 - \rho^2} + \frac{2\rho\sigma^2}{(1 - \rho^2)(1 - \phi\rho)}$$

de donde se obtiene:

$$E y_{t-1}^2 = \frac{\sigma^2(1 - \phi\rho) + 2\rho\sigma^2}{(1 - \rho^2)(1 - \phi^2)(1 - \phi\rho)} = \gamma_0$$

por lo que:

$$\text{plim} \frac{\sum y_{t-1}^2}{T} = \gamma_0$$

Utilizando todos estos resultados se llega a:

$$\begin{aligned} \text{plim} \hat{\phi} &= \phi + \frac{\sigma^2 \rho / (1 - \rho^2)(1 - \phi\rho)}{(\sigma^2(1 - \phi\rho) + 2\sigma^2 \rho) / (1 - \rho^2)(1 - \phi\rho)(1 - \phi^2)} = \\ &= \phi + \frac{\rho(1 - \phi^2)}{1 - \phi\rho + 2\rho} \neq \phi \end{aligned}$$

quedando demostrado que el estimador MCO es inconsistente.

Ejemplo 2.3: (Errores en variables)

Suponer que el modelo que relaciona dos variables, y_t y X_t , viene dado por

$$y_t = \beta X_t + u_t$$

En donde $u_t \sim \text{iidN}(0, \sigma^2)$. X_t no puede ser observada y, en su lugar, se

observa Z_t de forma que:

$$Z_t = X_t + v_t$$

en donde también, $v_t \sim \text{iidN}(0, \sigma_v^2)$ y es independiente de u_t y X_t .

El modelo en términos de variables observables puede escribirse como:

$$y_t = \beta(Z_t - v_t) + u_t = \beta Z_t + w_t$$

en donde $w_t = u_t - \beta v_t$. La dependencia contemporánea se refleja en:

$$\text{Cov}(Z_t, w_t) = -\beta \text{Var}(v_t) \quad \square$$

El estimador MCO es:

$$\hat{\beta} = \frac{\sum y_t Z_t}{\sum Z_t^2} = \beta + \frac{\sum Z_t w_t}{\sum Z_t^2} \quad (2.20)$$

$$\frac{1}{T} \sum Z_t w_t \xrightarrow{P} E(Z_t w_t) = -\beta \text{Var}(v_t)$$

$$\frac{1}{T} \sum Z_t^2 \xrightarrow{P} E Z_t^2 = \text{Var}(X) + \text{Var}(v_t)$$

Sustituyendo estos resultados en (2.20) se llega a:

$$\text{plim} \hat{\beta} = \beta - \frac{\beta \text{Var}(v_t)}{\text{Var}(X_t) + \text{Var}(v_t)} = \beta - \frac{\beta}{\frac{\text{Var}(X_t)}{\text{Var}(v_t)} + 1}$$

Este resultado pone de manifiesto que el estimador MCO es inconsistente y que la inconsistencia depende de la importancia de la varianza del error de observación con respecto a la varianza de la variable no observada.

2.3 MODELOS CON VARIABLES NO ESTACIONARIAS.

En la sección anterior hemos flexibilizado alguna de las hipótesis del MLG aceptando la posibilidad de que los regresores fueran estocásticos. Pero incluso en ese caso, el análisis se ha mantenido en un marco estacionario.

Hemos visto que, en este marco, bastaba dividir por T a las sumas de cuadrados o sumas cruzadas de las variables para que tendieran a límites finitos. Esto hacía que expresiones del tipo $\sqrt{T}(\hat{\beta} - \beta)$ tuvieran distribuciones asintóticas bien definidas, en general todas ellas gaussianas. Además, los estadísticos tipo Wald

propuestos para el contraste de determinadas hipótesis nulas tienden asintóticamente a distribuciones estándar tipo χ^2 sin ninguna dependencia de parámetros desconocidos.

Estos resultados pueden verse retomando el modelo autoregresivo estudiado en el Ejemplo 2.1.

El estimador MCO del parámetro lo hemos escrito en (2.15). Se ve como el denominador dividido por T tiende a un escalar finito obteniendo la distribución de $\sqrt{T}(\hat{\phi} - \phi)$ que hemos escrito en (2.16).

Cuando se relaja la hipótesis de estacionariedad, el planteamiento varía substancialmente. Las sumas de los cuadrados y de productos cruzados hay que dividirlos por T^v con $v \geq 1$, siendo necesarios varios valores de v según las expresiones. Por otra parte, en lugar de expresiones del tipo $\sqrt{T}(\hat{\beta} - \beta)$, hay que considerar expresiones del tipo $T^v(\hat{\beta} - \beta)$ en donde $v \geq 1/2$; esto nos conduce al concepto de superconsistencia frente al de consistencia que es el que, hasta este momento, hemos venido manejando. En otros casos se presentan situaciones en las que los estadísticos, sean éstos estimadores o contrastes, siguen distribuciones diferentes a las estándar, dependiendo, en ciertos casos, de parámetros no conocidos.

A continuación, vamos a presentar una serie de desarrollos que permiten entender algunos de éstos puntos.

Comenzaremos haciendo referencia al concepto de variable Browniana o proceso de Wiener. Partiremos del modelo de paseo aleatorio sin deriva que escribiremos como:

$$y_t = y_{t-1} + u_t \quad (2.21)$$

en donde suponemos que los u_t son i.i.d. con media cero y varianza 1 y, además, que $y_0 = 0$.

Sustituyendo sucesivamente se llega a:

$$y_t = \sum_{i=1}^t u_i \quad (2.22)$$

A partir de esta expresión se tiene que:

$$y_t \sim N(0, t) \quad (2.23)$$

de forma que la no estacionariedad queda reflejada en la varianza.

Para diferentes valores de t , (2.22) toma la forma siguiente:

$$y_1 = u_1; y_2 = u_1 + u_2; y_3 = u_1 + u_2 + u_3; y_4 = u_1 + u_2 + u_3 + u_4; \dots$$

Una característica importante de estas variables es que diferencias no solapadas de las mismas son independientes. Por ejemplo, $y_2 - y_1 = u_2$ e $y_4 - y_3 = u_4$, son independientes entre sí.

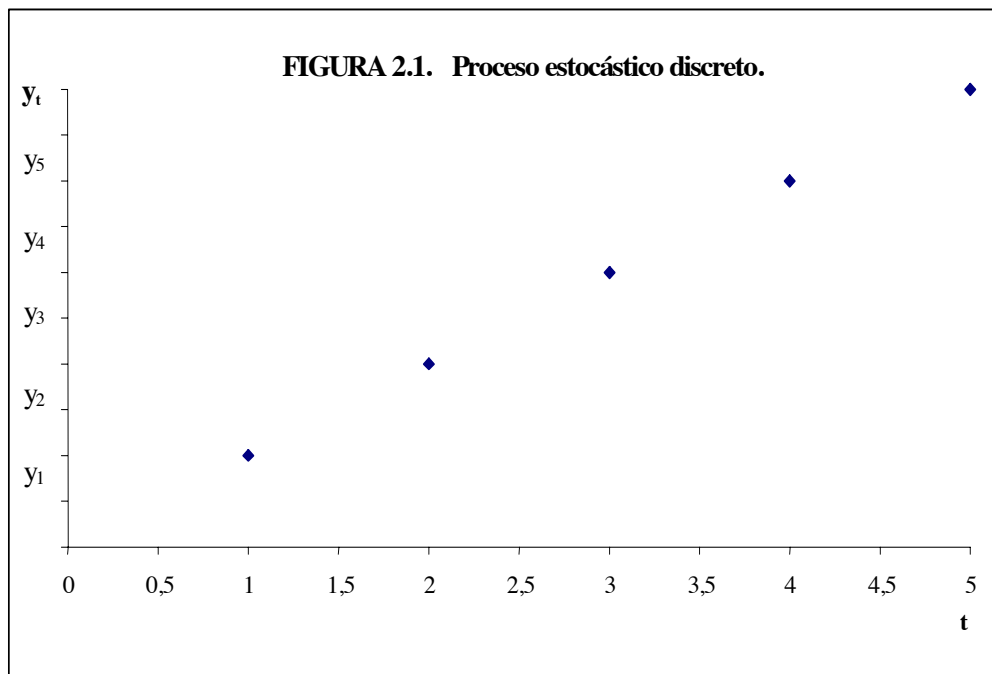
La generalización de este resultado en un marco continuo es lo que conduce al concepto de variable Browniana.

La realización de un proceso estocástico multivariante discreto experimenta una especie de metamorfosis para acomodarse a un marco en el que el argumento del proceso en lugar de pertenecer al conjunto de todos los números enteros se limita al intervalo $(0,1)$. Veamos esta metamorfosis gráficamente.

Partimos de la realización de un proceso discreto con cinco observaciones que escribimos como:

$$y_1, y_2, y_3, y_4, y_5$$

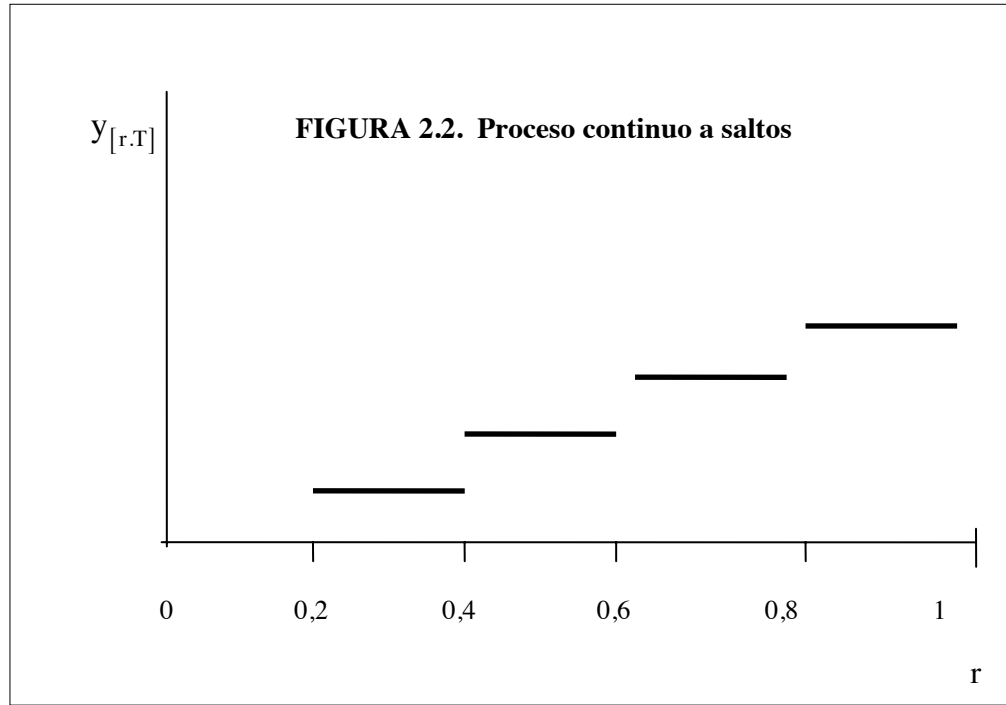
Su representación gráfica aparece en la Figura 2.1



Definamos lo que vamos a llamar un proceso estocástico continuo a saltos. Sea $[rT]$ la parte entera de rT con $0 \leq r \leq 1$. El proceso estocástico continuo a saltos se define como:

$$y_{[rT]} = y_{i-1} \text{ para } \frac{i-1}{T} \leq r \leq \frac{i}{T} \quad i = 0, 1, 2, \dots, 5$$

Su representación puede verse en la Figura 2.2.



Cuando el tamaño de la muestra observada se va haciendo grande, la amplitud de los intervalos en el intervalo $(0,1)$ se va haciendo menor hasta llegar a una línea continua que es la variable Browniana. Es decir, se tiene que:

$$\frac{y_{[rT]}}{\sqrt{T}} \xrightarrow{d} W(r) \quad (2.24)$$

en donde $W(r)$ es la variable Browniana.

Siguiendo a Hamilton (1994) adoptaremos la siguiente definición:

Definición 2.1: La variable Browniana estándar es un proceso estocástico continuo que asocia cada valor de $r \in [0,1]$ con el escalar $W(r)$ tal que:

$$(a) \quad W(0) = 0 \quad (2.25)$$

(b) Para cualesquiera valores, $0 \leq r_1 \leq r_2 \leq \dots r_k \leq 1$ los cambios $[W(r_2) - W(r_1)], \dots [W(r_k) - W(r_{k-1})]$ son independientes entre sí, cumpliéndose $[W(s) - W(t)] \sim N(0, s - t)$

(c) Para cualquier realización, $W(r)$ es continua en r con probabilidad 1.

Observar que, como consecuencia de (b) se tiene que:

$$W(r) \sim N(0, r) \quad (2.26)$$

La variable Browniana es como un proceso aleatorio continuo definido sobre $[0, 1]$.

Un resultado importante que utilizaremos insistentemente en lo que queda de sección es el llamado Mapping Theorem que dice lo siguiente: Si $f(\cdot)$ es una función continua en $[0, 1]$ entonces:

$$f\left(\frac{y_{[rT]}}{\sqrt{T}}\right) \xrightarrow{d} f(W(r)) \quad (2.27)$$

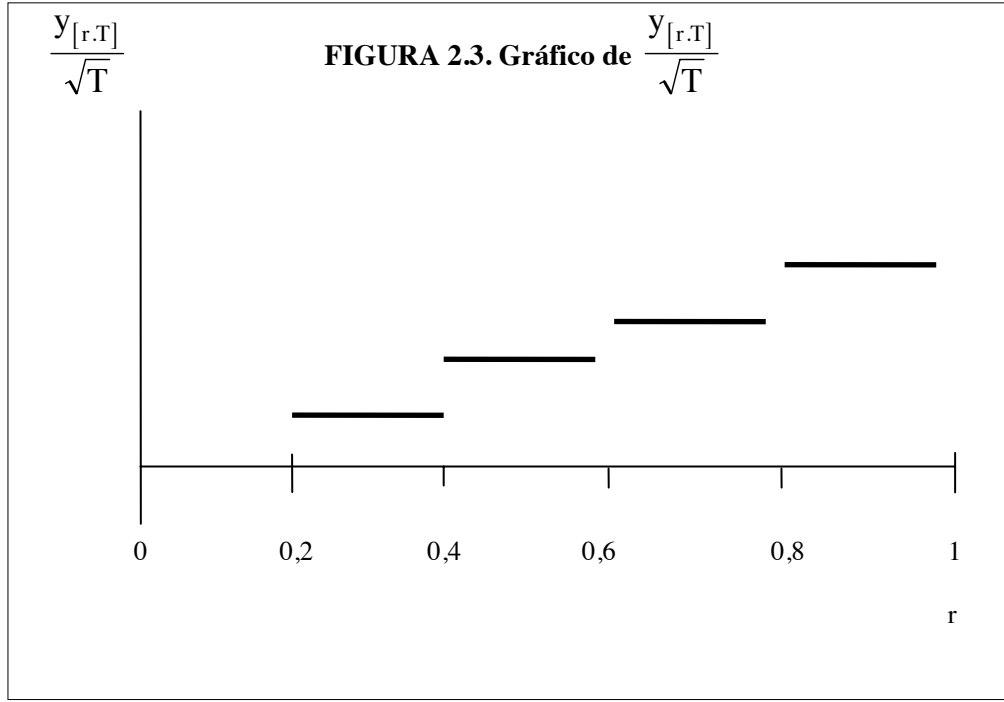
Por ejemplo, con la función integral:

$$\int_0^1 \frac{y_{[rT]}}{\sqrt{T}} dr \xrightarrow{d} \int_0^1 W(r) dr \quad (2.28)$$

Habiendo definido el concepto de variable Browniana estudiemos la convergencia de algunas expresiones habituales en la evaluación de ciertos estadísticos, prestando especial atención al factor de normalización necesario para lograr la convergencia. Comenzaremos analizando el siguiente sumatorio:

$$\sum_{t=1}^T y_{t-1}$$

Considerar lo que hemos llamado el proceso estocástico continuo a saltos dividido por \sqrt{T} . Gráficamente puede verse en la Figura 2.3



Se ve que es una línea quebrada. El área de los rectángulos determinados por el eje de abscisas y esta línea quebrada es:

$$\int_0^1 \frac{y_{[r.T]}}{\sqrt{T}} dr = \sum \frac{y_{[r.T]}}{\sqrt{T}} \times \frac{1}{T} = T^{-3/2} \sum_{t=1}^T y_{t-1}$$

ya que se trata de rectángulos con base igual a $\frac{1}{T}$ y altura igual a $\frac{y_{[r.T]}}{\sqrt{T}}$.

Por lo tanto, utilizando (2.28) se llega a:

$$T^{-3/2} \sum y_{t-1} \xrightarrow{d} \int_0^1 W(r) dr \quad (2.29)$$

Destacar que el factor de normalización en este caso es $T^{3/2}$ (es decir, el factor por el que hay que dividir) y que converge no a una constante sino a una expresión estocástica que cae fuera de lo que es habitual y estándar.

Resultado 2.6 : Considerar un proceso generado de acuerdo con (2.21). Entonces se tiene que:

a) $T^{-1/2} \sum u_t \xrightarrow{d} W(1) = N(0,1)$

b) $T^{-3/2} \sum y_{t-1} \xrightarrow{d} \int_0^1 W(r) dr$

$$\begin{aligned}
\text{c) } T^{-1/2} \bar{y} &\xrightarrow{d} \int_0^1 W(r) dr \\
\text{d) } T^{-1} \sum y_{t-1} u_t &\xrightarrow{d} \frac{1}{2} [W(1)^2 - 1] \\
\text{e) } T^{-3/2} \sum t u_t &\xrightarrow{d} W(1) - \int_0^1 W(r) dr \\
\text{f) } T^{-2} \sum y_{t-1}^2 &\xrightarrow{d} \int_0^1 W(r)^2 dr \\
\text{g) } T^{-5/2} \sum t y_{t-1} &\xrightarrow{d} \int_0^1 r W(r) dr \\
\text{h) } T^{-3} \sum t y_{t-1}^2 &\xrightarrow{d} \int_0^1 r W(r)^2 dr \\
\text{i) } T^{-(v+1)} \sum t^v &\xrightarrow{p} \frac{1}{v+1}
\end{aligned}$$

Prueba:

a) Es una consecuencia de la aplicación del Resultado A.6 del Anexo 1

(resultado de Lindeberg-Levy).

b) Ya demostrado anteriormente.

c) $T^{-1/2} \bar{y} = T^{-3/2} \sum y_{t-1}$ que es b).

d) Tener en cuenta que:

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + u_t^2 + 2y_{t-1}u_t$$

de donde:

$$y_{t-1} \cdot u_t = \frac{1}{2} (y_t^2 - y_{t-1}^2 - u_t^2)$$

y sumando:

$$\sum y_{t-1} u_t = \frac{1}{2} (y_T^2 - y_0^2) - \frac{1}{2} \sum u_t^2$$

y como: $y_T \sim N(0, T)$ y suponiendo que $y_0 = 0$:

$$\frac{1}{T} \sum y_{t-1} u_t = \frac{1}{2} \left(\frac{y_T}{\sqrt{T}} \right)^2 - \frac{1}{2} \frac{\sum u_t^2}{T} \xrightarrow{d} \frac{1}{2} \chi^2(1) - \frac{1}{2} = \frac{1}{2} [W(1)^2 - 1]$$

e) Primeramente vamos a demostrar que:

$$T^{-3/2} \sum t u_t = T^{-1/2} \sum u_t - T^{-3/2} \sum y_{t-1} \quad (2.30)$$

Escribimos:

$$\begin{aligned} T^{-3/2} \sum_{t=1}^T y_{t-1} &= T^{-3/2} [u_1 + (u_1 + u_2) + (u_1 + u_2 + u_3) + \dots \\ &+ (u_1 + u_2 + \dots + u_{T-1})] = T^{-3/2} [(T-1)u_1 + (T-2)u_2 + \dots \\ &+ (T - (T-1))u_{T-1}] = T^{-3/2} \sum_{t=1}^T (T-t)u_t \end{aligned}$$

llegándose así a (2.30).

El resultado se obtiene aplicando a) y b) en (2.30).

f) Basta aplicar (2.27).

$$g) T^{-5/2} \sum t y_{t-1} = T^{-3/2} \sum \left(\frac{t}{T} \right) y_{t-1} \xrightarrow{d} \int_0^1 r W(r) dr \text{ para } r = \frac{t}{T}$$

y aplicando (2.27).

$$h) T^{-3} \sum t y_{t-1}^2 = T^{-2} \sum \left(\frac{t}{T} \right) y_{t-1}^2 \xrightarrow{d} \int_0^1 r W(r)^2 dr .$$

i) Inductivamente:

$$\begin{aligned} \sum_{t=1}^T t &= \frac{T(T+1)}{2} \text{ que requiere una normalización por el factor } T^2 \\ \sum_{t=1}^T t^2 &= \frac{T(T+1)(2T+1)}{6} \text{ que requiere una normalización por el factor } T^3 \end{aligned}$$

y así sucesivamente. □

Se ve que todas las expresiones van precedidas por una potencia de T . Este es el factor de normalización. Es el factor por el que hay que dividir la expresión para poder decir que tiene un límite finito. Por esta razón en ocasiones a la expresión original se la denomina $O(T^v)$, queriendo indicar que hay que dividirla por T^v para conseguir la convergencia. Por ejemplo, diremos que $\sum u_t$ es $O(T^{1/2})$, que $\sum y_{t-1}$ es $O(T^{3/2})$ y así sucesivamente.

Hay que destacar que los resultados comentados se han derivado asumiendo hipótesis bastante restrictivas, en concreto que el proceso no tiene deriva, que la varianza de la perturbación es 1 y que ésta no tiene autocorrelación. Pero la extensión de los resultados a los casos en que se incumple alguno de estos supuestos es inmediata. En la Proposición 17.3 de Hamilton (1994) pueden encontrarse los resultados para cuando hay autocorrelación y la perturbación tiene una varianza diferente de la unidad.

A continuación, derivaremos algunos resultados para aquellos casos en que la deriva es diferente de cero.

Considerar el siguiente proceso:

$$y_t^* = \delta + y_{t-1}^* + u_t \quad (2.31)$$

en donde δ es una constante.

Sustituyendo sucesivamente y suponiendo $y_0^* = 0$, podemos escribir:

$$y_t^* = \delta t + \sum_{i=1}^t u_i = \delta t + y_t \quad (2.32)$$

Resultado 2.7 : Suponer que el proceso es el escrito en (2.31). Entonces se tiene que:

$$a) T^{-3/2} \sum y_{t-1}^* u_t \xrightarrow{d} \delta \left\{ W(1) - \int_0^1 W(r) dr \right\}$$

$$b) T^{-2} \sum y_{t-1}^* \xrightarrow{p} \frac{\delta}{2}$$

$$c) T^{-3} \sum y_{t-1}^{*2} \xrightarrow{p} \frac{\delta^2}{3}$$

$$d) T^{-3} \sum t y_{t-1}^* \xrightarrow{p} \frac{\delta}{3}$$

$$e) T^{-4} \sum_t y_{t-1}^{*2} \xrightarrow{P} \frac{\delta^2}{4}$$

Prueba:

a) Teniendo en cuenta (2.32):

$$\sum y_{t-1}^* u_t = \delta \sum_t u_t - \delta \sum u_t + \sum y_{t-1} u_t$$

$$\text{Por el Resultado 2.6.e) } \sum_t u_t \text{ es } O\left(T^{3/2}\right),$$

por el a) $\sum u_t$ es $O\left(T^{1/2}\right)$ y por el d) $\sum y_{t-1} u_t$ es $O(T)$.

Por lo tanto:

$$T^{-\frac{3}{2}} \sum y_{t-1}^* u_t \longrightarrow \delta T^{-3/2} \sum_t u_t = \delta \left[W(1) - \int_0^1 W(r) dr \right]$$

$$b) \sum y_{t-1}^* = \delta \sum (t-1) + \sum y_{t-1}$$

A partir del Resultado 2.6. se ve que por i), $\sum (t-1)$ es $O(T^2)$

y por b) $\sum y_{t-1}$ es $O\left(T^{3/2}\right)$.

Por lo tanto:

$$T^{-2} \sum y_{t-1}^* \longrightarrow \frac{\delta}{2} \text{ aplicando i)}$$

$$c) \sum y_{t-1}^{*2} = \delta^2 \sum (t-1)^2 + \sum y_{t-1}^2 + 2\delta \sum (t-1) y_{t-1}$$

Utilizando de nuevo el Resultado 2.6 se ve que:

por i) $\sum (t-1)^2$ es $O(T^3)$, por f) $\sum y_{t-1}^2$ es $O(T^2)$

y por g) $\sum (t-1) y_{t-1}$ es $O\left(T^{5/2}\right)$.

Por lo tanto:

$$T^{-3} \sum y_{t-1}^{*2} \xrightarrow{P} \frac{\delta^2}{3}$$

$$d) \sum t y_{t-1}^* = \delta \sum t(t-1) + \sum t y_{t-1}$$

A partir del Resultado 2.6 se ve que $\sum t(t-1)$ es $O(T^3)$ y que:

$$\sum t y_{t-1} \text{ es } O(T^{5/2}). \text{ Por lo tanto, } T^{-3} \sum t y_{t-1}^* \xrightarrow{p} \frac{\delta}{3}$$

$$e) \quad \sum t y_{t-1}^{*2} = \delta^2 \sum t(t-1)^2 + \sum t y_{t-1}^2 + 2\delta \sum t(t-1) y_{t-1}$$

Por los resultados vistos, el primer término de la derecha es $O(T^4)$,

el segundo $O(T^3)$ y el tercero puede verse fácilmente que es $O(T^{7/2})$

Por lo tanto:

$$T^{-4} \sum t y_{t-1}^{*2} \xrightarrow{p} \frac{\delta^2}{4} \quad \square$$

La generalización a un proceso multivariante no presenta dificultades especiales.

Considerar el siguiente proceso k-variante:

$$\Delta Z_t = \begin{bmatrix} \Delta y_t \\ \Delta X_{2t} \\ \vdots \\ \Delta X_{kt} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{bmatrix} \quad (2.33)$$

en donde suponemos que $u_t = (u_{1t} \cdots u_{kt})$ se distribuyen i.i. como $N(0, I_k)$.

Sustituyendo sucesivamente en (2.33) se tiene (suponiendo que los valores iniciales son cero):

$$Z_t = \begin{bmatrix} y_t \\ X_{2t} \\ \vdots \\ X_{kt} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^t u_{1i} \\ \sum_{i=1}^t u_{2i} \\ \vdots \\ \sum_{i=1}^t u_{ki} \end{bmatrix} \quad (2.34)$$

Siguiendo las pautas comentadas para el proceso univariante podemos definir el concepto de vector de k-variables Brownianas, que seguiremos llamando $W(r)$. Siguiendo a Hamilton (1994) y Johansen (1995), la definición que adoptamos es:

Definición 2.2 El vector de k-variables Brownianas $W(\circ)$ es un proceso k-variante continuo que asocia cada valor $r \in [0,1]$ al vector $W(r)$ de forma que:

a) $W(0) = 0$

b) $W(r)$ es $N_k(0, rI_k)$

c) Para cualesquiera valores $0 \leq r_1 < r_2 < \dots r_s \leq 1$, los cambios

$[W(r_2) - W(r_1)], [W(r_3) - W(r_2)], \dots, [W(r_s) - W(r_{s-1})]$ son procesos

multivariantes e independientes entre sí que siguen una distribución Normal

con $[W(s) - W(r)] \sim N(0, (s - r)I_k)$

d) Para cualquier realización, $W(r)$ es continuo en r con probabilidad 1.

La conexión entre el proceso escrito en (2.34) y el vector $W(r)$ es la misma que la comentada para el caso univariante. Podemos pensar ahora en un proceso multivariante continuo a saltos, que escribiremos como Z_{rT} y reproducir (2.24) en su versión multivariante:

$$\frac{Z_{[rT]}}{\sqrt{T}} \xrightarrow{d} W(r) \quad (2.35)$$

Resultado 2.8 Sea Z_t el proceso k-variante generado según (2.33). Sea Z la matriz $T \times k$ de las observaciones del proceso k variante y sea U la matriz $T \times k$ de las observaciones del vector de perturbaciones. Entonces se tiene que:

a) $T^{-1/2} \sum u_t \xrightarrow{d} W(1) = N(0, I_k)$

b) $T^{-1} Z'_{-1} U \xrightarrow{d} \int_0^1 W(r) dW(r)$

c) $T^{-3/2} \sum_1^T Z_{t-1} \xrightarrow{d} \int_0^1 W(r) dr$

d) $T^{-2} Z' Z \xrightarrow{d} \int_0^1 W(r) W(r)' dr$

$$e) T^{-5/2} \sum t z_{t-1} \xrightarrow{d} \int_0^1 r W(r) dr$$

$$f) T^{-3} t(Z'Z) \xrightarrow{d} \int_0^1 r W(r) W(r)' dr$$

Prueba: Ver Proposición 18.1 de Hamilton (1994). □

Hay que destacar que también en este caso los resultados se derivan asumiendo un marco muy restringido en el que el vector u_t tiene una matriz de varianzas y covarianzas igual a la unidad y no esta sujeta a autocorrelación. Pero la extensión a otros casos en que la matriz de varianzas y covarianzas es diferente de la matriz unidad y hay autocorrelación no presenta ninguna complicación. De hecho, la demostración de Hamilton corresponde a estas situaciones más complicadas.

Bastaría suponer que el vector u_t definido en (2.33) es tal que:

$$E u_t u_t' = \Sigma$$

$$y \quad E \sum_{j=1}^{\infty} u_t u_{t-j} = \Lambda$$

y se define la matriz de varianzas y covarianzas a largo plazo como:

$$\Omega = \Sigma + \Lambda + \Lambda'$$

Entonces utilizando la Definición 2.2 definiríamos el movimiento Browniano $B(r)$ con matriz de varianzas y covarianzas igual a Ω y el Resultado 2.8 se seguirá manteniendo cambiando $W(r)$ por $B(r)$.

Examinemos ahora dos conceptos que han tenido mucha relevancia en el tratamiento de los modelos con variables no estacionarias: correlación espúrea y superconsistencia.

Comencemos considerando el proceso k variante escrito en (2.33). Como se ve, todas las variables y_{it} son independientes entre sí. Por lo tanto, si especificamos un modelo del tipo:

$$y_t = \beta_2 x_{2t} + \cdots + \beta_k x_{kt} + u_t$$

o, matricialmente:

$$y = X^* \beta^* + u \tag{2.36}$$

y lo estimamos utilizando MCO, las estimaciones deberían indicar que las variables son realmente independientes. Vamos a ver que se llega a un resultado diferente.

Los estimadores MCO de β^* en (2.36) pueden escribirse:

$$\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} y$$

Particionemos el vector de variables Brownianas como:

$$W(r) = \begin{bmatrix} W_1(r) \\ W_2(r) \end{bmatrix}$$

en donde suponemos que $W_1(r)$ corresponde a y_t y $W_2(r)$ tiene los $(k-1)$ elementos correspondientes a $x_{2t} \cdots x_{kt}$.

Entonces, a partir del Apartado d) del Resultado 2.8 podemos escribir:

$$T^{-2} (X^{*'} X^*) \xrightarrow{d} \int_0^1 W_2(r) [W_2(r)]' dr$$

$$T^{-2} (X^{*'} y) \xrightarrow{d} \int_0^1 W_2(r) W_1(r) dr$$

Usando estos resultados se llega a:

$$\hat{\beta} \xrightarrow{d} \left[\int_0^1 W_2(r) [W_2(r)]' dr \right]^{-1} \left[\int_0^1 W_2(r) W_1(r) dr \right]$$

Este resultado tiene varios aspectos que merecen destacarse. En primer lugar, el vector de estimadores tiende a un vector de elementos estocásticos y no a un vector de parámetros. En segundo lugar, nada parece indicar que los valores que toma el estimador sean cero que es lo que cabría esperar por ser independiente y de las X .

Supongamos ahora que el proceso generador de los datos viene dado por:

$$y_t = \beta_2 x_{2t} + \dots + \beta_k x_{kt} + u_{1t} \quad (2.37)$$

$$\Delta x_{it} = u_{it} \quad i = 2, \dots, k \quad (2.38)$$

El vector de estimadores MCO puede escribirse como:

$$\hat{\beta} = (X^{*'} X^*)^{-1} X^{*'} y = \beta + (X^{*'} X^*)^{-1} X^{*'} u_{1t}$$

También podemos escribir:

$$T(\hat{\beta} - \beta) = \left(\frac{1}{T^2} X^{*'} X^* \right)^{-1} \frac{1}{T} (X^{*'} u_{1t})$$

Haciendo uso de los Apartados b) y d) del Resultado 2.8 se obtiene:

$$T(\hat{\beta} - \beta) \xrightarrow{d} \left[\int_0^1 W_2(r) W_2(r)' dr \right]^{-1} \int_0^1 W_2(r) dW_1(r) \quad (2.39)$$

A partir de este resultado se pueden hacer los siguientes comentarios. Primero, que para lograr la convergencia tenemos que multiplicar a $(\hat{\beta} - \beta)$ por T y no por \sqrt{T} ; eso significa que en el marco no estacionario la tasa de convergencia de $\hat{\beta}$ a β es superior a la encontrada en el marco estacionario. Esto puede valorarse como una ventaja. Pero, y esto debe valorarse como un inconveniente, la distribución de $T(\hat{\beta} - \beta)$ se aleja de una distribución estándar como la distribución Normal.

Cuando en la especificación aparece una relación como la escrita en (2.37) decimos que en el modelo existe cointegración entre las variables. Sobre este tema volveremos en el Capítulo 5.

2.4 MÉTODOS DE MONTE-CARLO

Hasta este momento, hemos definido estadísticos y hemos derivado sus propiedades utilizando reglas analíticas de la Estadística Matemática, principalmente asentadas en torno a la obtención de la esperanza de determinadas expresiones.

Pero este método tiene límites claros. En bastantes ocasiones los estadísticos adoptan una forma compleja que hace imposible derivar sus propiedades. En algunos casos, el razonamiento asintótico permite obviar las dificultades. Pero en otros, no es así y se hace necesario pensar en métodos de simulación. Esta es la razón última de los Métodos de Monte-Carlo. Pero incluso en aquellos casos en que se ha obtenido la distribución asintótica uno está interesado en calibrar si los resultados que son válidos para una muestra grande también lo son para una muestra más reducida, más próxima a lo que es habitual en el trabajo aplicado.

Por lo tanto, la metodología de Monte-Carlo persigue dos objetivos:

- Derivar las propiedades estocásticas de determinados estadísticos cuando su derivación no puede abordarse con un método analítico.
- Determinar en qué medida los resultados asintóticos son aplicables en muestras pequeñas.

A continuación, vamos a presentar algunas ilustraciones que ayudan a entender la intuición que subyace a este tipo de métodos. Para una tratamiento detallado de otros aspectos puede consultarse el trabajo de Davidson y Mackinnon (1994).

El Método de Monte-Carlo deriva las propiedades de los estadísticos muestrales especificando los valores concretos de los parámetros del Proceso Generador de Datos (PGD) y un tamaño muestral concreto. Una vez especificado el PGD se generan los valores de las variables y, a partir de ellos, los de los estadísticos muestrales. Este ejercicio se repite tantas veces como se haya especificado en el "número de simulaciones". Hechas las repeticiones se dispone de valores de los estadísticos muestrales, tantos como sea el número de simulaciones. A partir de estos valores se pueden calcular los momentos que caracterizan la distribución de probabilidad de los estadísticos y derivar sus propiedades.

Todo estudio de Monte-Carlo requiere las siguientes especificaciones:

- 1) los elementos del PGD
- 2) tamaño muestral
- 3) número de simulaciones.

Como ilustración de la aplicación de este método, tratemos de derivar el valor esperado de la media muestral. Para ello consideremos las siguientes especificaciones:

- 1) PGD: Supongamos una variable aleatoria normal con media $\mu_x = 5$ y varianza igual a $\sigma_x^2 = 1$
- 2) Tamaño Muestral: $T = 10$
- 3) Número de Simulaciones: $N = 100$.

A partir de estas especificaciones, el proceso es como sigue:

Utilizando un paquete informático cualquiera de los disponibles se generan 10 valores de una variable distribuida normalmente con media 5 y varianza la unidad. Con estos diez valores se calcula la media y este será el primer valor que toma el estadístico media muestral; lo denotaremos por \bar{X}_1 .

SIMULACIÓN N° 1: \bar{X}_1

A continuación se repite ejercicio, obteniendo un segundo valor del estadístico muestral que denotaremos por \bar{X}_2 .

SIMULACIÓN N° 2: \bar{X}_2 .

Se sigue repitiendo el proceso, obteniéndose sucesivamente, $\bar{X}_3, \bar{X}_4, \dots$, hasta llegar a \bar{X}_{100} . Al final del proceso disponemos de 100 valores del estadístico muestral que se está estudiando, en este caso, la media muestral.

A partir de estos 100 valores se puede caracterizar la distribución de probabilidad de la media muestral. En particular, se puede obtener la media de estos 100 valores mediante la expresión:

$$\bar{X}^* = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_{100}}{100} \quad (2.40)$$

También puede calcularse la varianza mediante la expresión:

$$\frac{1}{N} \sum_{i=1}^{100} (\bar{X}_i - \bar{X}^*)^2 \quad (2.41)$$

Si el estadístico media muestral es insesgado parece razonable esperar que:

$$\bar{X}^* \approx 5$$

Frente al método analítico el método de Monte-Carlo tiene la ventaja de proporcionar resultados en situaciones en las que con el primer método no puede llegarse a ninguna conclusión. Tal es el caso, por ejemplo, de modelos con regresores estocásticos con muestras finitas.

Pero frente a esa ventaja, hay que hacer notar que la validez de los resultados que se obtienen con el método de Monte-Carlo queda limitada al marco estocástico o PGD especificado en el punto de partida. Por esa razón, se recomienda que el marco estocástico supuesto sea lo suficientemente amplio como para dar cabida a las diferentes situaciones que uno puede encontrar en el trabajo aplicado. También es conveniente tener en cuenta diferentes tamaños muestrales y que el número de simulaciones sea grande.

Para ilustrar y justificar estas recomendaciones se ha ampliado el ejercicio anterior siguiendo las siguientes direcciones:

1) PGD: Se ha mantenido la media igual a 5 pero, para la varianza, se han considerado 3 valores: 1, 4 y 16.

2) Se han considerado tres tamaños muestrales: $T = 10, 100$ y 500.

3) El ejercicio se ha repetido 10, 100 y 500 veces.

Para el caso con varianza igual a 1 y 10 observaciones los resultados pueden verse en el Cuadro 2.1. En la segunda columna de este cuadro aparecen los 10 valores generados usando el paquete informático GAUSS versión 3.2.12. El valor de la media muestral calculado aparece en la parte de la derecha.

CUADRO 2.1. Valores generados de $N(5,1)$

Observación	Valores Generados	
1	4,86	
2	3,97	
3	6,16	
4	5,83	
5	4,74	$\bar{X} = 5,131$
6	6,50	
7	5,40	
8	3,39	
9	5,74	
10	4,72	

Manteniendo la misma varianza y el mismo tamaño muestral el ejercicio se repite 10 veces obteniéndose 10 valores de la media muestral. La media de estos diez valores así como su varianza aparecen en la primera fila del Cuadro 2.2. La primera columna se refiere al tamaño muestral; la segunda a la varianza supuesta; las columnas 3 y 4 recogen los resultados, en la tercera la media de las 10 simulaciones y, la cuarta, la varianza de las mismas.

CUADRO 2.2. Media Muestral. Simulación con 10 repeticiones

TM	VP	\bar{X}^*	$V\bar{X}$
10	1	4.3743	1.9819
10	4	4.4413	2.3216
10	16	5.1540	3.8031
100	1	4.5597	2.0851
100	4	4.4562	2.0011
100	16	4.6722	2.3688
500	1	4.5454	2.0675
500	4	4.5096	2.0381
500	16	4.4602	2.0409

En este cuadro puede apreciarse cómo los valores de \bar{X}^* tienden a aproximarse a 5 aunque la calidad de la aproximación es todavía discutible. Destacar también que la varianza calculada en la última columna crece conforme lo hace la varianza del PGD.

Los resultados para cuando el número de simulaciones es 100 y 500 aparecen en los Cuadros 2.3 y 2.4, respectivamente.

CUADRO 2.3. Media Muestral. Simulación con 100 Repeticiones

TM	VP	\bar{X}^*	$V\bar{X}$
10	1	4.9164	0.3318
10	4	5.0174	0.6309
10	16	5.0540	1.4548
100	1	4.9494	0.2542
100	4	4.9455	0.2707
100	16	4.9510	0.3900
500	1	4.9533	0.2470
500	4	4.9463	0.2518
500	16	4.9465	0.2852

CUADRO 2.4. Media Muestral. Simulación con 500 repeticiones

TM	VP	\bar{X}^*	$V\bar{X}$
10	1	4.9982	0.1479
10	4	4.9736	0.4327
10	16	4.9596	1.7296
100	1	4.9938	0.0584
100	4	4.9820	0.0887
100	16	4.9802	0.2117
500	1	4.9902	0.0516
500	4	4.9866	0.0582
500	16	4.9949	0.0809

Analizando la secuencia de estos cuadros se pone de manifiesto que, conforme el número de simulaciones crece, la calidad de la aproximación a la media se va haciendo mayor y, además, la dispersión se hace menor.

A partir de lo comentado cabe hacer dos observaciones. Primero, que es posible considerar marcos para el PGD muy amplios que alivian el carácter restrictivo de todo estudio de Monte-Carlo al que ya nos hemos referido. Segundo, es muy importante que el número de simulaciones sea elevado pues cuanto mayor sea mejores serán los resultados a los que se llega.

A continuación, presentaremos una segunda ilustración referida al Modelo Lineal Simple.

Este modelo, para la observación t -ésima puede escribirse como:

$$y_t = \alpha + \beta X_t + u_t \quad t = 1, 2, \dots, T$$

Los estimadores MCO de α y β son:

$$\hat{\beta} = \frac{\sum (X_t - \bar{X})(y_t - \bar{y})}{\sum (X_t - \bar{X})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{X}$$

Las propiedades de estos estimadores son bien conocidas porque, en este caso, el método analítico proporciona los resultados esperados. Bastaría particularizar los resultados obtenidos en el capítulo anterior.

Pese al conocimiento de las propiedades de estos estimadores, a efectos ilustrativos, nos ha parecido interesante derivarlas también utilizando los métodos de Monte-Carlo.

Repitiendo los pasos comentados anteriormente las etapas del proceso son las siguientes:

1. Se fija el tamaño muestral T .
2. Se concreta el Proceso Generador de Datos determinando:
 - (i) Los valores de α y β .
 - (ii) El valor de x en cada observación.
 - (iii) Se especifica una distribución concreta para u y mediante un proceso generador de números aleatorios se obtienen los diferentes valores de u .
3. Se genera el valor de y en cada observación utilizando el modelo y los valores de α , β , x y u .
4. Utilizando sólo los valores de y así generados y los datos para x se obtienen los estimadores MCO de α y β .
5. Se fija el número de veces que se repite el experimento. Este número determina los valores de $\hat{\alpha}$ y $\hat{\beta}$ disponibles.
6. A partir de los valores obtenidos para $\hat{\alpha}$ y $\hat{\beta}$ se calculan los momentos y , en su caso, se analiza la forma de la distribución de esos valores.

Veamos ahora un ejemplo con las siguientes especificaciones.

1. El tamaño muestral se fija en 10.
2. (i) Los valores de α y β son: $\alpha = 2$ y $\beta = 1$.
- (ii) Los valores de x son 10 valores generados a partir de una variable distribuida normalmente con media 5 y varianza 1. Estos 10 valores son los que aparecen en la primera columna del Cuadro 2.5.

(iii) Para u suponemos una distribución normal con media cero y varianza la unidad. Una primera orden al proceso generador con esta especificación proporciona los valores que aparecen en la segunda columna del Cuadro 2.5.

3. Utilizando los valores comentados de α , β , x y u se generan los de y utilizando el modelo lineal simple. Los valores generados aparecen en la tercera columna del Cuadro 2.5.

4. Introduciendo estos valores de y y x en las fórmulas que definen $\hat{\alpha}$ y $\hat{\beta}$ obtendríamos estimaciones concretas de α y β .

$$\hat{\alpha} = -0,53 \quad \hat{\beta} = 1,54$$

resultando el modelo estimado:

$$y_t = -0,53 + 1,54 X_t$$

Pero un valor de $\hat{\alpha}$ y $\hat{\beta}$ no es suficiente para dar cuenta de las propiedades de una variable aleatoria. Hay que repetir el experimento más veces.

5. El número de veces que se va a repetir es 10. La diferencia entre una simulación y otra radica solamente en los valores generados para la variable u que son diferentes una a otra. Todos los otros datos -valores de α , β y x - no varían.

CUADRO 2.5. Simulación para el Modelo Lineal Simple

Valores de X	Valores de u	Valores de y
4,86	0,39	7,25
3,97	-0,60	5,37
6,16	2,34	10,50
5,83	-0,57	7,26
4,74	0,25	6,99
6,50	0,51	9,01
5,40	0,67	8,07
3,39	-0,57	4,82
5,74	0,17	7,91
4,72	-0,39	6,33

Con los 10 valores obtenidos de $\hat{\alpha}$ y $\hat{\beta}$ se calculan sus medias y varianzas. Los cuatro valores correspondientes a este caso aparecen en la tercera fila del Cuadro 2.6.

Como se ha indicado antes, para que un estudio de Monte-Carlo sea útil, el PGD debe plantearse de forma amplia y llevarse a cabo para un número grande de repeticiones.

Siguiendo esta recomendación el ejercicio se ha repetido para cuatro tamaños muestrales - $T = 10, 50, 100$ y 500 y 4 valores de la varianza de la perturbación aleatoria - $VP = 0,0001, 0,01, 1$ y 100 .

Los resultados para cuando el número de simulaciones es 10 quedan recogidos en el Cuadro 2.6. Para cuando el número de simulaciones es 100 y 500 los resultados pueden verse en los Cuadro 2.7 y 2.8, respectivamente.

CUADRO 2.6. Modelo Lineal Simple. 10 Simulaciones

T	VP	$\bar{\alpha}$	$\bar{\beta}$	$\text{Var}(\hat{\alpha})$	$\text{Var}(\hat{\beta})$
10	0.0001	1.8233	0.9077	0.3328	0.0824
10	0.01	1.7603	0.9200	0.3253	0.0853
10	1	1.2764	1.0097	2.3465	0.1953
10	100	-1.2611	1.5749	75.2609	3.7613
50	0.0001	1.8164	0.9094	0.3300	0.0827
50	0.01	1.7860	0.9153	0.3206	0.0838
50	1	1.5365	0.9704	0.7092	0.1107
50	100	6.0569	0.1294	129.1226	4.4240
100	0.0001	1.8182	0.9092	0.3306	0.0827
100	0.01	1.8239	0.9077	0.3340	0.0824
100	1	1.8063	0.9041	0.4965	0.0883
100	100	1.8789	0.9259	11.1020	0.4704
500	0.0001	1.8177	0.9092	0.3304	0.0827
500	0.01	1.8080	0.9110	0.3275	0.0830
500	1	1.8562	0.9015	0.3681	0.0822
500	100	1.1778	1.0960	5.3713	0.3326

CUADRO 2.7. Modelo Lineal Simple. 100 Simulaciones

T	VP	$\bar{\hat{\alpha}}$	$\bar{\hat{\beta}}$	$\text{Var}(\hat{\alpha})$	$\text{Var}(\hat{\beta})$
10	0.0001	1.9790	0.9904	0.0393	0.0098
10	0.01	1.9961	0.9873	0.0564	0.0104
10	1	1.9170	0.9971	1.7208	0.0884
10	100	4.3417	0.5475	161.1749	7.0211
50	0.0001	1.9802	0.9901	0.0393	0.0098
50	0.01	1.9778	0.9906	0.0450	0.0101
50	1	1.9793	0.9883	0.5640	0.0309
50	100	2.4139	0.8881	61.8963	2.6021
100	0.0001	1.9802	0.9901	0.0392	0.0098
100	0.01	1.9794	0.9904	0.0414	0.0099
100	1	1.8988	1.0093	0.2107	0.0172
100	100	1.5615	1.0671	24.3169	0.9247
500	0.0001	1.9800	0.9901	0.0392	0.0098
500	0.01	1.9804	0.9900	0.0397	0.0098
500	1	2.0094	0.9849	0.0925	0.0116
500	100	1.9833	0.9932	5.0425	0.1956

CUADRO 2.8. Modelo Lineal Simple. 500 Simulaciones

T	VP	$\bar{\hat{\alpha}}$	$\bar{\hat{\beta}}$	$\text{Var}(\hat{\alpha})$	$\text{Var}(\hat{\beta})$
10	0.0001	1.9930	0.9986	0.0084	0.0020
10	0.01	2.0080	0.9956	0.0593	0.0039
10	1	2.1993	0.9601	4.6325	0.1793
10	100	2.7349	0.8806	448.8690	17.0787
50	0.0001	1.9964	0.9979	0.0080	0.0020
50	0.01	1.9964	0.9979	0.0134	0.0022
50	1	1.9962	0.9966	0.6002	0.0252
50	100	1.0610	1.1700	57.2608	2.2329
100	0.0001	1.9960	0.9980	0.0080	0.0020
100	0.01	1.9963	0.9978	0.0106	0.0021
100	1	2.0225	0.9946	0.3044	0.0130
100	100	2.3621	0.9350	29.1138	1.1062
500	0.0001	1.9960	0.9980	0.0080	0.0020
500	0.01	1.9972	0.9979	0.0085	0.0020
500	1	1.9736	1.0029	0.0670	0.0044
500	100	2.0126	0.9938	5.1917	0.2052

La principal conclusión que se deriva de estos cuadros es que cualquiera que sea el par tamaño muestral-varianza de la perturbación que se elija, siempre el valor

medio de los estadísticos muestrales estimadores MCO, $\hat{\alpha}$ y $\hat{\beta}$ tienden a estar en torno a los valores de los parámetros que se pretendían estimar. Es la evidencia clara a favor del carácter de estimadores insesgados. Como era de esperar, esto es así en mayor medida cuanto mayor es el número de simulaciones.

EJERCICIOS

2.1). Considerar el siguiente modelo:

$$y_t = \beta x_t + u_{1t}$$

$$x_t = \delta + x_{t-1} + u_{2t}$$

en donde β y δ son constantes y u_{1t} y u_{2t} son ruidos blancos independientes entre sí.

Sabiendo que el orden de probabilidad de $\sum u_{2t}$ es $O(T^{\frac{3}{2}})$:

1). Derivar los órdenes de probabilidad de :

$$\sum x_t \quad \text{y} \quad \sum x_t^2$$

2). Analizar las propiedades del estimador MCO de β . En caso de que este estimador no sea insesgado derivar el sesgo.

2.2). Considerar el siguiente modelo:

$$y_t = y_{t-1} + u_t$$

$$x_t = x_{t-1} + v_t$$

en donde u_t y v_t son ruidos blancos independientes entre sí.

Apartir del modelo:

$$y_t = \alpha + \beta x_t + e_t$$

se estiman α y β utilizando MCO.

1). Derivar las propiedades asintóticas de estos dos estimadores.

2). Derivar las propiedades asintóticas del estadístico t-ratio para la hipótesis nula: $\beta=0$.

3). Comentar las implicaciones que se derivan de los resultados obtenidos en los dos puntos anteriores.

2.3). Considerar el siguiente modelo:

$$y_t = \lambda + y_{t-1} + u_t$$

$$x_t = \delta + x_{t-1} + v_t$$

en donde u_t y v_t son ruidos blancos independientes entre sí.

Apartir del modelo:

$$y_t = \alpha + \beta x_t + e_t$$

se estiman α y β utilizando MCO.

1). Derivar las propiedades asintóticas de estos dos estimadores.

2). Derivar las propiedades asintóticas del estadístico t-ratio para la hipótesis nula: $\beta=0$.

3). Comentar las implicaciones que se derivan de los resultados obtenidos en los dos puntos anteriores.

2.4). Suponer que el proceso generador de datos es:

$$y_t = \alpha + \beta x_t + u_t \quad (1)$$

en donde u_t es un ruido blanco con varianza σ^2 .

No se dispone de observaciones sobre x y, en su lugar, utilizamos otra variable no estocástica, z . Es decir, estimamos:

$$y_t = \alpha + \beta z_t + u_t^* \quad (2)$$

Se pide:

1). Las propiedades, media y varianza, de los estimadores MCO de α y β definidos a partir de (2), indicando como dependen dichas propiedades de la pendiente de la regresión de x sobre z .

2). Demostrar que la esperanza del residuo MCO definido a partir de (2), u_t^* , es diferente de cero pero que se cumple que: $E(\frac{\sum u_t^*}{T}) = 0$.

2.5). Considerar el siguiente modelo:

$$\begin{aligned} y_t + x_t &= v_t \\ x_t - x_{t-1} &= \varepsilon_{2t} \end{aligned} \quad (1 - \rho_1 L)v_t = \varepsilon_{1t}$$

con:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim \text{i.i.d.} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right]$$

Suponiendo que $|\rho_1| < 1$ se pide:

1). La esperanza y varianza de v_t, y_t y x_t .

2). La esperanza y varianza condicionadas a la información del periodo anterior de esas tres variables.

3). Suponiendo ahora que $\rho_1 = 1$ obtener la esperanza y varianza de Δy_t y Δx_t y de la distribución condicional de Δy_t dado Δx_t .

4). Para $\rho_1 = 1$ demostrar que el estimador MCO de Δy_t sobre Δx_t es un estimador sesgado del coeficiente de x en la primera relación del modelo. Demostrar también que, asintóticamente, el sesgo coincide con el coeficiente de Δx_t en la esperanza de la distribución condicional de Δy_t dado Δx_t .

2.6). Para el vector $z_t = (y_t, x_t)'$ considerar el siguiente modelo:

$$\begin{aligned} y_t + x_t &= v_t & (1 - \rho_1 L)v_t &= \varepsilon_{1t} \\ y_t + 2x_t &= u_t & (1 - \rho_2 L)u_t &= \varepsilon_{2t} \end{aligned}$$

con:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim \text{i.i.d.} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right]$$

Se pide:

- 1). Suponiendo que $|\rho_1| < 1$ y $|\rho_2| < 1$, obtener la esperanza y matriz de varianzas y covarianzas de z_t . Derivar ambos momentos para la distribución condicional de y_t dado x_t . Hacer la regresión MCO de y_t sobre x_t y derivar las propiedades del estimador.
- 2). Suponer ahora que $|\rho_1| = 1$ y $|\rho_2| < 1$. Obtener la esperanza y matriz de varianzas y covarianzas de z_t . Derivar las propiedades del estimador MCO de y_t sobre x_t .
- 3). Suponer ahora que $|\rho_1| = 1$ y $|\rho_2| = 1$. Derivar las propiedades del estimador MCO de y_t sobre x_t . Obtener las propiedades del estimador MCO de Δy_t sobre Δx_t . Obtener la esperanza y varianza de la distribución condicional de y_t dado x_t y los valores pasados de ambas variables.

2.7). Suponer el modelo:

$$\begin{aligned} y_t &= \phi y_{t-1} + \beta_1 x_{t-1} + \beta_2 x_{t-2} + u_{1t} \\ x_t &= \delta + \rho x_{t-1} + u_{2t} \end{aligned}$$

u_{1t} y u_{2t} son ruidos blancos independientes entre sí. Se pide:

- 1). Las condiciones que deben cumplir los parámetros del modelo para que las dos variables sean estacionarias en sentido débil. Suponiendo que se cumplen estas dos condiciones derivar la media y la varianza de las dos variables. Derivar, también, la función de autocorrelación de x_t y dibujar, aproximadamente, el correlograma.
- 2). Suponiendo que los parámetros cumplen las condiciones que garantizan la estacionariedad se estiman los parámetros de la primera relación utilizando los estimadores MCO ¿Son esos estimadores insesgados? ¿Son eficientes para todo tamaño muestral? Derivar las propiedades asintóticas de estos estimadores.
- 3). Suponiendo que los parámetros del modelo son conocidos definir el predictor óptimo de y_{T+2} situados en T . Derivar el error de predicción correspondiente su media y su varianza.

2.8). Suponer el modelo:

$$c_t = \beta y_t + u_t \quad (1)$$

$$y_t = c_t + a_t \quad (2)$$

en donde c_t , y_t y a_t son, respectivamente, consumo, renta y gasto autónomo; u_t es un ruido blanco. Se supone que a_t no es aleatorio.

1). Obtener la forma reducida las medias, varianzas y covarianza de c_t e y_t y la media y varianza de c_t dada y_t .

2). Estimar β aplicando MCO en (1). Derivar las propiedades de este estimador tanto en muestras pequeñas como asintóticamente y determinar el signo del sesgo asintótico en caso de que exista.

3). Estimar β utilizando el estimador de los mínimos cuadrados en dos etapas y derivar las propiedades de este estimador.
