PESQUIMAT 20(2): 65–83 (2017) ISSN:1560-912X/ ISSN-E:1609-8439
http://dx.doi.org/10.15381/pes.v20i2.14516 Facultad de Ciencias Matemáticas – UNMSM

Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia

Juvert Huaranga Narvajo[1]

Resumen: El presente trabajo tiene el propósito de obtener la distribución asintótica de los estimadores de mínimos cuadrados ordinarios en el modelo de regresión lineal donde los regresores siguen procesos estocásticos strong-mixing y tendencia, el uso de procesos strong-mixing permiten obtener resultados más generales dado que se incluye variables tanto estacionarias como no estacionarias. En la derivación de la distribución asintótica se utilizan resultados de la teoría de la probabilidad y análisis real para procesos dependientes. La distribución obtenida difiere de las distribución normal standard y depende de los parámetros de las variables.
Palabras clave: proceso strong-mixing; regresión lineal, distribución asintótica.

Aymptotic distribution of OLS estimators in linear regressions with strong-mixing explanatory variable and trend

Abstract: The aim of the paper is to obtain MCO estimators in a linear regression with strong-mixing explanatory variables and trend. Strong-mixing process offers a greater degree of generalization given that stationary and non-stationary regressors can be included. To obtain the asymptotic distribution results from real analysis and probability theory for dependent processes are used. Unlike classical results for estimators in linear models the derived asymptotic distribution is not normally standard and depends on the parameters of the variables.
Keywords: strong mixing process; linear regression; asymptotic distribution.

Recibido: 01/10/2017. *Aceptado*: 20/12/2017. *Publicado online*: 31/12/2017

[1]UNMSM, Facultad de Ciencias Matemáticas, e-mail: juvert.huaranga@unmsm.edu.pe

1. Introducción

La existencia de fenómenos que siguen procesos estocásticos no estacionarios ocurre en diversas disciplinas como: astronomía, química, economía, finanzas, biología y entre otros. El modelamiento estadístico de tales fenomenos se ha convertido en un área de constante crecimiento dentro de cada disciplina, sin embargo tal modelamiento requiere métodos inferenciales diferentes a los tradicionales basados en la independencia de las observaciones, en consecuencia la teoría probabilidad desarrolló nuevas herramientas analíticas que permiten realizar inferencias en tales procesos. En el contexto de las regresiones de series temporales bajo el método de los minimos cuadrados ordinarios (MCO), estos desarrollos toman la forma de distribuciones de Wiener, teoremas del límite central funcional (TLCF) y varios teoremas del análisis real y funcional.
No obstante los desarrollos en teoría de la probabilidad dados en décadas recientes, no es posible hallar de manera general la distribución de los estimadores MCO en regresiones con variables no estacionarias. Esto se debe a que el comportamiento de los procesos estocásticos no estacionarios varía según el valor de los parámetros y el tipo de proceso no estacionario, en cambio, se han hallado distribuciones asintóticas en regresiones con variables que siguen ciertos tipos de procesos tales como: procesos autoregresivos de orden 1 AR(1), procesos autoregresivos de orden 2 AR(2), procesos integrados de orden (1) y de orden (2), procesos estacionarios en diferencia de orden(1), entre otros. De modo que existe espacio para desarrollos posteriores en la obtención de distribuciones asintóticas usando diferentes tipos de procesos no estacionarios, en consecuencia el presente trabajo tiene por objetivo obtener la distribución asintótica de los estimadores de mínimos cuadrados ordinarios (MCO) en regresiones de series temporales con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia.
La investigación se justifica porque el conocer la verdadera distribución de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia permite implementar procedimientos inferenciales más generales en el marco del modelo de regresión ya que se pueden al incluir variables dependientes con propiedades estadísticas más débiles que las requeridas en otros modelos de regresión y al mismo tiempo se puede incluir una tendencia en el modelo, y en consecuencia evitar las relaciones causales espurias en regresiones de series temporales no estacionarias, esto es, sostener que existe una relación causal entre las variables cuando en realidad no la hay. Tal procedimiento es de gran importancia en trabajos aplicados.
Finalmente, el presente trabajo se divide en 5 secciones. La sección 1 es la introducción. En la sección 2 se revisa la literatura en el tema. Mientras que en la sección 3 se realiza una exposición de definiciones y resultados en teoría de la probabilidad, análisis real y funcional necesarios para entender el presente trabajo así como del modelo de regresión lineal múltiple bajo el método de los mínimos cuadrados ordinarios. En la sección 4 se deriva de manera rigurosa los resultados de la presente investigación, y finalmente en la sección 5 se presentan las conclusiones.

2. Revisión de la literatura

El estudio de las propiedades asintóticas de los estimadores MCO en regresiones con series temporales es de larga data. En el plano univariado, se tienen los trabajos de Park y Phillips [15] que desarrollan la teoría asintótica en regresiones multivariantes con procesos integrados de diferente orden, tales regresiones pueden tener interceptos, derivas y tendencias temporales. En Durlauf y Phillips [6] se analizan las propiedades asintóticas de los estimadores MCO cuando se realiza la extracción errónea de la tendencia determinística en procesos que son integrados. Mientras que en Haldrup [11] se deriva las propiedades de los estimadores MCO en regresiones que cointegran con variables que siguen procesos integrados de orden uno y dos. Además en

Hasseler [12] son derivadas las propiedades asintóticas de los estimadores MCO en regresiones con procesos que siguen tendencias lineales. Y en Phillips [24] se presenta las distribuciones tanto de los estimadores MCO como de los test estadísticos en regresiones cuyas variables siguen un proceso auto-regresivo. Por otro lado Kramer y Marmol [20] se derivan las propiedades asintóticas de los estimadores MCO y mínimos cuadrados generalizados (GLS, en adelante) en regresiones lineales donde los errores siguen un proceso autoregresivo (AR) de orden p. En Phillips y Durlauf [25] se desarrolla la teoría asintótica para variables integradas de orden uno en regresiones multivariadas, vectores auto-regresivos y modelos de corrección del error; y bajo diferentes tipos de errores, entre ellos: errores con dependencia débil y heterogeneidad, en el proceso generador de datos (PGD, en adelante). Mientras que en Stock, Sims y Watson [27] que desarrollan las propiedades asintóticas y de las pruebas de hipótesis de los estimadores MCO cuando algunas de las variables siguen procesos integrados de diferente orden. Por otro lado en Park y Phillips [16] se analiza las propiedades asintóticas de variables que siguen procesos integrados de orden uno en regresiones multivariadas así como las distribuciones de los estadísticos en diversas pruebas de hipótesis. Adicionalmente en Mynbaev [23] se deriva la distribución de los estimadores MCO en regresiones con regresores espaciales mixtos y en Zhang y Zhang [30] se deriva la distrubición asintótica en procesos estocásticos dependientes cuya estructura de dependencia esta relacionada con las sumas parciales del mismo proceso.De modo que la literatura es extensa y variada en el área en cuestión, dejando mucho espacio para desarrollos posteriores.

3. Definiciones y Preliminares

3.1. Procesos estocásticos y tipos de procesos estocásticos

Definición 1 Sea $\big(\Omega, \mathcal{F}, P\big)$ un espacio de probabilidad, $\big(X, \mathfrak{B}\big)$ un espacio medible y T un conjunto de índices. Un proceso estocástico[1] $\{X_t\}_{t \in T}$ es una familia o conjunto de variables aleatorias en un espacio de probabilidad común, tal que: $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ para $t \in T$, y se denota como $\{X_t\}$.

Por definición un proceso estocástico, se compone de un espacio muestral, funciones indexadas temporalmente y una medida de probabilidad, de modo que un proceso estocástico es una función que asigna a cada elemento del espacio muestral una trayectoria temporal asociada a una distribución de probabilidad; y en general un proceso estocástico $\{X_t\}$ tiene momentos probabilísticos (por ejemplo: media y varianza) que son funciones del tiempo.
Existen diversas formas de clasificar los procesos estocásticos según el tipo de propiedades estadísticas y probabilísticas que posean. En el presente trabajo los diversos desarrollos analíticos y teoremas requieren un uso intensivo de los siguientes tipos de procesos estocásticos:

Procesos dependientes e independientes

Un proceso estocástico $\{X_t\}$ es independiente[2] si se cumple lo siguiente:

$$F_{X_1 X_2 \ldots X_n}(X_1 X_2 \ldots X_n) = F_{x_1}(X_1, t_1) \times F_{x_2}(X_2, t_2) \ldots \times F_{x_n}(X_n, t_n) = \prod_{i=1}^{n} F_{x_i}(X_i, t_i).$$

Lo cual quiere decir que la distribución conjunta del proceso estocástico $\{X_t\}$ para t instantes temporales $(t_1, t_2, \ldots t_n)$ se puede expresar como el producto de la función de distribución mismo proceso para los n instantes del tiempo. En caso contrario el proceso $\{X_t\}$ se dice que es dependiente.

[1] Ver Capasso [23] para mayores detalles sobre los procesos estocásticos.
[2] Véase Kobayashi [19, Pág. 318].

Procesos estacionarios y no estacionarios

Un proceso estocástico $\{X_t\}$ es estacionario en sentido estricto[3], también llamado estrictamente estacionario, si su distribución conjunta en el intervalo temporal $\{t_i : i = 1, 2, ...n\}$ es la misma que en el intervalo $\{t_i + h \,:\, i = 1, 2, ...n\}$ $\forall h$. En otras palabras:

$$F_{X_1 X_2 \cdots X_n}(X_1 X_2 \cdots X_n) = F_{X_{1+h} X_{2+h} \cdots X_{n+h}}(X_1 X_2 \cdots X_n)$$

De lo cual se desprende que la función de distribución un proceso estocástico estricamente estacionario $\{X_t\}$se mantiene invariante ante cualquier desplazamiento temporal. Mientras que un proceso estocástico $\{X_t\}$es estacionario de segundo orden, también llamado estacionario en covarianza o débilmente estacionario, si se cumple que[4]:

1. $E\left[X(t)\right] = \mu_x$para todo $-\infty < t < +\infty$
2. $E\left[X(t)^2\right] < \infty$ para todo $-\infty < t < +\infty$
3. $E\left[\left(X(s) - \mu\right)\left(X\left(t\right) - \mu\right)\right]$ depende de la diferencia temporal $|s - t|$

De modo que un proceso estacionario $\{X_t\}$ es un proceso estocástico cuyas leyes probabilísticas no cambian por desplazamientos temporales. En otras palabras un proceso débilmente estacionario $\{X_t\}$ tiene momentos de primer y segundo orden que no son funciones del tiempo.
Un proceso estocástico $\{X_t\}$ es no estacionario cuando sus leyes probabilisticas no satisfacen los supuestos necesarios para ser estacionario, en otras palabras, las leyes probabilisticas de un proceso no estacionario son funciones del tiempo. Un ejemplo conocido es la caminata aleatoria.

Procesos de Wiener

Un proceso de Wiener[5], también llamado proceso Browniano o movimiento Browniano, es un proceso estocástico de tiempo continuo que satisface las siguientes propiedades:

- Homogeneidad espacial:

$$\Pr\left[W\left(t\right) \leq w \mid W\left(0\right) = a\right] = \Pr\left[W\left(t\right) \leq w + b \mid W\left(0\right) = a + b\right].$$

- Homogeneidad temporal:

$$\Pr\left[W\left(t\right) \leq w \mid W\left(0\right) = a\right] = \Pr\left[W\left(t + s\right) \leq w \mid W\left(s\right) = a\right].$$

- Incrementos independientes:

  Para un conjunto de intervalos disjuntos $\left(s_i, t_i\right], i = 1, 2, \ldots$ se tiene que:

$$W\left(t_i\right) - W\left(t_s\right), i = 1, 2, \ldots$$

  son incrementos independientes.

- Propiedad Gausiana:

  Cualquier incremento $W\left(t_i\right) - W\left(s_i\right), t_i > s_i$ es normalmente distribuido:

$$\Pr\left[W\left(t\right) \leq x \mid W\left(t_0\right) = x_0\right] = \frac{1}{\sqrt{2\pi\alpha\left(t - t_0\right)}} \int_{-\infty}^{x - x_0} \exp\left\{-\frac{y^2}{2\alpha\left(t - t_0\right)}\right\} dy.$$

- Propiedad de Markov:

$$\Pr\left[W\left(t + s\right) \leq w \mid W\left(u\right), u \leq t\right] = \Pr\left[W\left(t + s\right) \leq w \mid W\left(t\right)\right], \forall s \geq 0.$$

[3]Mayores detalles en Doob [5, Pág. 94].
[4]Véase Karlin y Taylor [17, Pág. 445].
[5]Véase Kobayashi [19, Pág. 491]. Un extenso análisis de los procesos de Wiener y sus propiedades véase [2].

3.2. Procesos Strong-Mixing

Dos eventos A y B son independientesen un mismo espacio de probabilidad cuando:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B).$$

Y son dependientes cuando $\Pr(A \cap B) - \Pr(A) \times \Pr(B) \neq 0$. De la misma manera dos eventos generados por los sub-espeacios de probabilidad $\mathcal{A}$ y $\mathcal{B}$, son independientes si se cumple:

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B) \quad \textit{con } A \in \mathcal{A} \textit{ y } B \in \mathcal{B}$$

De modo que una medida útil para cuantificar la dependencia entre dos eventos es la siguiente:

$$\Big|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\Big|$$

Definición 2 *Sea $\mathcal{A}$ y $\mathcal{B}$, dos sub-σ álgebras de $\mathcal{F}$, entonces la medida de dependencia α-mixing se define*[6] *:*

$$\alpha\left(\mathcal{A}, \mathcal{B}\right) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} \Big|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\Big|$$

Definición 3 *Sea $\mathcal{F}_{-\infty}^{t} = \sigma\big(X_t, X_{t-1}, X_{t-2}, \ldots\big)$ y $\mathcal{F}_{t}^{\infty} = \sigma\big(X_t, X_{t+1}, X_{t+2}, \ldots\big)$ un par de σ-álgebras generados por$\{X_t\}$. Un coeficiente α-mixing*[7] *se define como:*

$$\alpha\left(t\right) = \sup_{t} \alpha\left(\mathcal{F}_{1}^{t}, \mathcal{F}_{t+n}^{\infty}\right)$$

Definición 4 *Sea $\mathcal{F}_{-\infty}^{t} = \sigma\big(X_t, X_{t-1}, X_{t-2}, \ldots\big)$ y $\mathcal{F}_{t}^{\infty} = \sigma\big(X_t, X_{t+1}, X_{t+2}, \ldots\big)$, un proceso $X(t)$ es α-mixing o strong-mixing*[8] *si:*

$$\alpha\left(t\right) = \sup_{t} \alpha\left(\mathcal{F}_{1}^{t}, \mathcal{F}_{t+n}^{\infty}\right) \longrightarrow 0 \quad \textit{cuando} \quad t \to \infty \tag{1}$$

O equivalentemente:

$$\alpha\left(t\right) = \sup_{t} \sup_{A \in \mathcal{F}_{-\infty}^{t}, B \in \mathcal{F}_{t}^{\infty}} \Big|\Pr(A \cap B) - \Pr(A) \times \Pr(B)\Big| \longrightarrow 0 \quad \textit{cuando} \quad t \to \infty$$

De la definición (4) se tiene que un proceso estocástico $\{X_t\}$ es strong mixing si la sucesión de coeficientes α-mixing tiende a 0, y se desprende que a medida que el coeficiente α-mixing tiende a 0, la dependencia dentro del proceso va disminuyendo y en el límite el proceso $\{X_t\}$ es asintóticamente independiente.
Una cualidad resaltante de los procesos strong-mixing es que pueden ser tanto procesos estacionarios como no estacionarios, mientras que otros procesos asintóticamente independientes como los procesos ergódicos solo pueden ser estacionarios, de modo que los procesos strong mixing son más generales que los procesos ergódicos.

Definición 5 *Sea $\alpha\left(t\right)$ una sucesión tal que $\alpha\left(t\right) = O\left(m^{-a-\varepsilon}\right)$ para algún $\varepsilon > 0$, entonces $\alpha\left(t\right)$ es de tamaño*[9] *$-a$.*

Dado que el coeficiente α-mixing tiende a 0 se puede interpretar como una sucesión de coeficientes α-mixing que tiene una tasa de convergencia. Si la sucesión de coefientes α-mixing es a lo mucho de orden $t^{-a-\epsilon}$ $\epsilon > 0$, denotado como $\alpha(t) = O\big(t^{-a-\epsilon}\big)$ $\epsilon > 0$, entonces se dice que α es de tamaño $-a$.

[6] Véase Bradley [3].
[7] Ver Gut [9, Pág. 450].
[8] Mayores detalles en Zhengyan y Chuanrong [30, Pág. 4].
[9] Véase White [28, Pág. 49].

Proposición 1 *Sea $X_t = g\big(Z_t, Z_{t-1}, \ldots, Z_{t-\tau}\big)$ donde $g\,(.)$ una función medible y τ es un entero positivo. Si Z_t es $\alpha-$mixing con coeficiente mixing $-a$, entonces $\big\{X_t\big\}$ sigue un proceso $\alpha-$mixing con coeficiente mixing $-a$.*[10]

Y generalización para vectores de variables aleatorias es la siguiente:

Proposición 2 *Sea el vector de procesos estocásticos $\{\mathbf{X_t}, \mathbf{y_t}\}$, donde $\mathbf{X_t}$ es un vector de procesos estocásticos y $\mathbf{y_t}$ es un proceso estocástico, α-mixing con tamaño $-a$, entonces $\Big\{\mathbf{X_t}\mathbf{X_t}'\Big\}$, $\{\mathbf{X_t}\mathbf{y_t}\}$ y $\Big\{\mathbf{y_t}'\mathbf{y_t}\Big\}$ es un vector de procesos estocásticos α-mixing con tamaño $-a$*[11].

Un resultado importante para desarrollos posteriores es una SLLN que corresponde a procesos strong mixing y que se presenta a continuación:

Proposición 3 *SLLN para sucesiones $\alpha-$mixing.*[12] *Sea $\big\{X_t\big\}$ una sucesión de variables aleatorias $\alpha-$mixing , con coeficiente α de tamaño $r/(r-1)$, $r > 1$ y $\mu_t \equiv E\big[X_t\big]$, tal que $E\big|X_t\big|^{r+\delta} \leq k < \infty$, $\delta > 0\,\forall t$. Y sea $S_n = X_1 + X_2 + \cdots + X_n$, $n \geq 1$. Entonces:*

$$\frac{S_n}{n} - \bar{\mu}_n \xrightarrow{a.s} 0 \tag{2}$$

donde $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} E\big[X_i\big]$.

Proposición 4 *FCLT para sucesiones de variables aleatorias α-mixing. Sea $\{X_n\}$ una sucesión de variables aleatorias en el espacio de probabilidad $\big(\Omega, X, P\big)$, y sea $S_n = \sum_{i=1}^{n} X_i$ con $n \in \mathbb{N}$, que satisface:*

$$\begin{array}{ll}
(i)\, E\big[X_n\big] = 0 & ,\, n \in \mathbb{N} \\
(ii)\, E\big[X_n^2\big] < \infty & ,\, n \in \mathbb{N} \\
(iii)\, \lim\limits_{n \longrightarrow \infty} \dfrac{E\big[S_n^2\big]}{n} = \sigma^2 & ,\, \sigma > 0 \\
(iv)\, \limsup E^{1/\beta}\big|X_n\big|^{\beta} < \infty & ,\, \beta \in \big(2, \infty\big] \\
(v)\, \sum\limits_{i \in \mathbb{N}} \alpha(i)^{1-\gamma} < \infty & ,\, \gamma = 2/\beta
\end{array}$$

Y se define la función aleatoria $W_n : \mathit{\Omega} \longrightarrow \mathcal{D}$ como: $W_n(t) = S_{[nt]}/\sigma\sqrt{n}\,, t \in [0,1]$. Entonces:

$$W_n(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}} \xrightarrow{d} W\,(t) \tag{3}$$

siendo $W\,(t)$ un proceso de Wiener standar.[13]

Corolario 1 *Si $t = 1$ entonces $W_n(1) = S_n/\sigma\sqrt{n} \xrightarrow{d} W\,(1) = N\,(0,1)$*[14].

Proposición 5 *Se cumple que $\displaystyle\int_0^1 s dW(s) \sim N\,(0, 1/3)$*[15]

[10] Ver lema 2.1 de White and Domowitz [10].
[11] Véase White [28, Pág. 50].
[12] El teorema se presenta como corolario 3.48 en White [28] y esta basado en el teorema 2.10 de McLeish [22].
[13] La demostración del teorema se encuentra colorario 1 en Herrndorf [14, Pág. 142].
[14] Véase Hassler [13, Pág. 307].
[15] Véase Ibid [13, Pág. 203] para una demostración de este resultado.

3.3. Método de los MCO en regresiones lineales.

El método de los Mínimos Cuadrados Ordinarios (MCO) es un método de optimización matemático, basado en el principio de minimos cuadrados, que se usa para hallar la mejor aproximación a la solución de un sistema de ecuaciones. En estadistica el método MCO es uno de varios métodos utilizados para hallar los estimadores en un modelo de regresión.[16] En el caso de una regresion lineal el método MCO minimiza la suma de los cuadrados de los residuos, que se definen como:

$$e_t = y_t - \hat{y}_t = y_t - \mathbf{x_t'}\widetilde{\beta}$$

con $\widetilde{\beta}$ un valor hipotético de β en $\boldsymbol{Y_t = X_t}\beta\boldsymbol{+\varepsilon_t}$.
El término e_t es el residuo de la observación t, que se expresa como la diferencia entre el valor actual de y_t y el valor de Y_t predicho que se obtiene al reemplazar $\tilde{\beta}$ por β , en $\mathbf{x_t'}\beta$. De modo que la expresión:

$$\sum_{i=1}^{T} e_i^2 = \sum_{i=1}^{T} \left(y_t - \mathbf{x_t'}\widetilde{\beta}\right)^2 = \mathbf{e'e} = \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)'\left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right).$$

Es la suma de los cuadrados de los residuos (SCR), de la expresión anterior se desprende que la SCR es una función de $\tilde{\beta}$ porque los residuos de cada observación dependen de $\tilde{\beta}$. Entonces el estimador MCO $\mathbf{b}$ de β, es el que minimiza la siguiente función:

$$\mathbf{b} \equiv \arg\min_{\widetilde{\beta}} SSR(\tilde{\beta}) = \left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right)'\left(\mathbf{y} - \mathbf{X'}\widetilde{\beta}\right).$$

Y mediante manipulaciones algebraicas se obtiene que[17] :

$$\mathbf{b} = \left[\mathbf{X'X}\right]^{-1}\mathbf{X'y}. \tag{4}$$

A modo de ejemplo sea $y_t = \alpha + \beta_1 x_t + \varepsilon_t$ el modelo de regresión con intercepto y una variable independiente x_t , entonces el vector de estimadores $\mathbf{b}$ de la ecuación (4) toma la siguiente forma:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \left[\begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_T \end{pmatrix} \times \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_T \end{pmatrix}\right]^{-1} \left[\begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_T \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}\right]$$

y queda como:

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} y_t \\ \sum_{t=1}^{T} x_t y_t \end{bmatrix}. \tag{5}$$

[16]Otros métodos son: Método de Máxima Verosimilitud, Método General de los Momentos, Método Bayesiano, etc. Véase Golberg y Cho [7] para la aplicación de éstos métodos en el modelo de regresión lineal.

[17]Una completa explicación de la derivación de los resultados presentados en esta sección se puede obtener en Greene [8, Pág. 26].

De manera alternativa se puede hallar el vector de estimadores $(\mathbf{b}-\beta)$ para el modelo de regresión. El punto de partida es la de la ecuación (4), tal que:

$$\begin{aligned}
\mathbf{b} &= \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{y}\\
\mathbf{b} &= \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\left[\mathbf{X}'\beta+\varepsilon\right], \qquad \textit{dado que } \mathbf{y}=\mathbf{X}'\beta+\varepsilon.\\
\mathbf{b} &= \beta+\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\varepsilon\\
\mathbf{b}-\beta &= \left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\varepsilon
\end{aligned}$$

De modo que:

$$(\mathbf{b}-\beta)=\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}'\varepsilon. \tag{6}$$

Y alternativamente:

$$(\mathbf{b}-\beta)=\left[\sum_{i=1}^{T}\mathbf{x}_i\mathbf{x}_i'\right]^{-1}\left[\sum_{i=1}^{T}\mathbf{x}_i\varepsilon_i\right]. \tag{7}$$

Cuando $y_t=\alpha+\beta_1x_t+\varepsilon_t$, la ecuación (6) se expresa como:

$$\begin{bmatrix}\hat{\alpha}-\alpha\\ \hat{\beta}_1-\beta_1\end{bmatrix}=\begin{bmatrix}T & \sum_{t=1}^{T}x_t\\ \sum_{t=1}^{T}x_t & \sum_{t=1}^{T}x_t^2\end{bmatrix}^{-1}\times\begin{bmatrix}\sum_{t=1}^{T}\varepsilon_t\\ \sum_{t=1}^{T}x_t\varepsilon_t\end{bmatrix}. \tag{8}$$

4. Distribución asintótica de los estimadores MCO en una regresión lineal con variables explicativas que siguen procesos estocásticos strong-mixing y tendencia

4.1. El modelo

La variable dependiente y_t sigue el siguiente proceso generador de datos:

$$y_t=\alpha+\delta t+\beta x_t+\varepsilon_t \qquad para\, t=1,2\ldots,T \tag{9}$$

donde α es un término constante, t es un componente tendencial determinístico, x_t es una variable que sigue un proceso α-mixing y ε_t son las perturbaciones que también siguen un proceso α-mixing. Esta especificacion indica que la variable y_t tiene un componente tendencial autónomo y a la vez depende otra variable x_t. El modelo que se presenta en la ecuación (9) es bastante general, ya que al seguir x_t un proceso α-mixing permite incorporar variables estacionarias como no estacionarias y así como permite heterogeneidad en las distribuciones de las observaciones. Otras especificaciones de x_t como la de un un proceso estacionario o deterministico que no permite incorporar variables no estacionarias y heterogeneas. Al mismo tiempo, al especificar que las perturbaciones siguen procesos α-mixing se permite también que las perturbaciones esten correlacionadas y sean heterogeneas.

4.2. Derivación de la distribución asintótica

Para hallar la distribución asintótica del estimador $\mathbf{b}$ en el modelo (9) se aplica el método M.C.O . Dado que el modelo (9) se puede expresar como $\mathbf{y}=\mathbf{X}'\beta+\varepsilon$ con $\beta'=\left[\alpha\ \delta\ \beta\right]$, el estimador $\mathbf{b}$ de β toma la forma de la ecuación (4):

$$\mathbf{b}=\left[\mathbf{X}'\mathbf{X}\right]^{-1}\mathbf{X}`\mathbf{y}.$$

De manera análoga la ecuación(8) para el vector $(\mathbf{b} - \beta)$ cuando $y_t = \alpha + \delta t + \beta x_t + \varepsilon_t$ es:

$$\begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\delta} - \delta \\ \hat{\beta} - \beta \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix}. \tag{10}$$

Ahora se pre-multiplica ambos lados de la ecuación matricial (10) por la siguiente matriz de escala[18]:

$$\mathit{\Gamma}_T = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}.$$

De modo que la ecuación (10) queda como:

$$\begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \times \begin{bmatrix} \hat{\alpha} - \alpha \\ \hat{\delta} - \delta \\ \hat{\beta} - \beta \end{bmatrix} = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \times \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix}.$$

Y dado que $\mathit{\Gamma}_T \mathit{\Gamma}_T^{-1} = \mathbf{I}$

$$\begin{bmatrix} \sqrt{T}(\hat{\alpha} - \alpha) \\ T^{3/2}(\hat{\delta} - \delta) \\ \sqrt{T}(\hat{\beta} - \beta) \end{bmatrix} = \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix} \begin{bmatrix} T & \sum_{t=1}^{T} t & \sum_{t=1}^{T} x_t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 & \sum_{t=1}^{T} t x_t \\ \sum_{t=1}^{T} x_t & \sum_{t=1}^{T} t x_t & \sum_{t=1}^{T} x_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}$$

$$\times \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \varepsilon_t \\ \sum_{t=1}^{T} t\varepsilon_t \\ \sum_{t=1}^{T} x_t\varepsilon_t \end{bmatrix}$$

[18]Se hace uso de matrices de escala porque las variables independientes tienen diferentes tasas de convergencia que implica que si se usa una escala única algunas variables divergen o colapsan.

$$\begin{bmatrix} \sqrt{T}\big(\hat{\alpha}-\alpha\big) \\ T^{3/2}\big(\hat{\delta}-\delta\big) \\ \sqrt{T}\big(\hat{\beta}-\beta\big) \end{bmatrix} = \left\{ \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} T & \sum\limits_{t=1}^{T} t & \sum\limits_{t=1}^{T} x_t \\ \sum\limits_{t=1}^{T} t & \sum\limits_{t=1}^{T} t^2 & \sum\limits_{t=1}^{T} tx_t \\ \sum\limits_{t=1}^{T} x_t & \sum\limits_{t=1}^{T} tx_t & \sum\limits_{t=1}^{T} x_t^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \right\}^{-1}$$

$$\times \left\{ \begin{bmatrix} \sqrt{T} & 0 & 0 \\ 0 & T^{3/2} & 0 \\ 0 & 0 & \sqrt{T} \end{bmatrix}^{-1} \begin{bmatrix} \sum\limits_{t=1}^{T} \varepsilon_t \\ \sum\limits_{t=1}^{T} t\varepsilon_t \\ \sum\limits_{t=1}^{T} x_t\varepsilon_t \end{bmatrix} \right\}$$

De modo que

$$\begin{bmatrix} \sqrt{T}\big(\hat{\alpha}-\alpha\big) \\ T^{3/2}\big(\hat{\delta}-\delta\big) \\ \sqrt{T}\big(\hat{\beta}-\beta\big) \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{\sum\limits_{t=1}^{T} t}{T^2} & \dfrac{\sum\limits_{t=1}^{T} x_t}{T} \\ \dfrac{\sum\limits_{t=1}^{T} t}{T^2} & \dfrac{\sum\limits_{t=1}^{T} t^2}{T^3} & \dfrac{\sum\limits_{t=1}^{T} tx_t}{T^2} \\ \dfrac{\sum\limits_{t=1}^{T} x_t}{T} & \dfrac{\sum\limits_{t=1}^{T} tx_t}{T^2} & \dfrac{\sum\limits_{t=1}^{T} x_t^2}{T} \end{bmatrix}^{-1} \begin{bmatrix} \dfrac{\sum\limits_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \\ \dfrac{\sum\limits_{t=1}^{T} t\varepsilon_t}{T^{3/2}} \\ \dfrac{\sum\limits_{t=1}^{T} x_t\varepsilon_t}{\sqrt{T}} \end{bmatrix}. \tag{11}$$

Para hallar la distribución asintótica se aplican los siguientes supuestos:

Suposición 1 *Sea x_t un proceso strong mixing con coeficiente α de tamaño $r/(r-1)$, $r > 1$ y $\mu_t \equiv E\big[X_t\big]$, tal que $\sup X_t \in \{X_t\}$ y $\sup E\big|X_t\big|^{r+\delta} \leq k < \infty$,$\delta > 1\,\forall t$. Y que $\bar{\mu}_T = \frac{1}{T}\sum_{t=1}^{T} E\big[X_t\big]$ con $\lim\limits_{T\longrightarrow\infty} \bar{\mu}_T = \mu$.*

Suposición 2 *Sea ε_t un proceso strong mixing con coeficiente α de tamaño $r/(r-1)$, $r > 1$, tal que $E\big[\varepsilon_t\big] = 0 \quad \forall t$, $\sup \varepsilon_t \in \{\epsilon_t\}$ y $\sup E\big|\varepsilon_t\big|^{r+\delta} \leq k < \infty$,$\delta > 0\,\forall t$ y con $\sigma_\varepsilon^2 = \lim\limits_{T\longrightarrow\infty} E\big[\frac{1}{T}S_\varepsilon^2\big]$,$\sigma_\varepsilon^2 > 0$[19]. Y que para x_t y ε_t se cumple: $E\big[x_t\varepsilon_t\big] = 0 \quad \forall t$ y $\sigma_{x\varepsilon}^2 = \lim\limits_{T\longrightarrow\infty} E\big[\frac{1}{T}S_{x\varepsilon}^2\big]$, $\sigma_{x\varepsilon}^2 > 0$[20].*

[19] Se define $S_\varepsilon^2 = \sum_{t=1}^{T} var\varepsilon_t$ y $S_{x\varepsilon}^2 = \sum_{t=1}^{T} var x_t\varepsilon_t$.
[20] Los términos σ_ε^2 y $\sigma_{x\varepsilon}^2$ reciben el nombre de varianza a largo plazo.

Dado los supuestos 1 y 2 se procede a determinar las propiedades asintóticas de cada elemento de la ecuación matricial (11).
El término de la primera fila y primera columa de la primera matriz de la ecuación (11) es 1, tomando límites se tiene:

$$\lim_{T\longrightarrow\infty} 1 = 1$$

y luego se tiene que[21]:

$$1 \xrightarrow{P} 1. \tag{12}$$

Mientras que en la primera matriz el término $\sum_{t=1}^{T} t/T^2$ de la ecuación (11) se puede expresar como:

$$\frac{\sum_{t=1}^{T} t}{T^2} = \frac{1}{T^2} \times \left[\frac{T(T+1)}{2}\right] = \frac{1}{2} + \frac{1}{(2T)}.$$

Y al igual que en (12) se tiene que:

$$\frac{\sum_{t=1}^{T} t}{T^2} \xrightarrow{P} \frac{1}{2}. \tag{13}$$

De manera análoga $\sum_{t=1}^{T} t^2/T^3$ se puede expresar como:

$$\frac{\sum_{t=1}^{T} t^2}{T^3} = \frac{1}{3} + \frac{1}{(2T)} + \frac{1}{6T^2}$$

Entonces:

$$\frac{\sum_{t=1}^{T} t^2}{T^3} \xrightarrow{P} \frac{1}{3} \tag{14}$$

Mediante la suposición 1, x_t sigue un proceso α- mixing con coeficiente mixing $r/(r-1),\ r>1$ y satisface los requisitos de la proposición 3[22]. De modo que para $\sum_{t=1}^{T} x_t/T.$ en la ecuación (11) se tiene:

$$\frac{\sum_{t=1}^{T} x_t}{T} - \bar{\mu}_T \xrightarrow{a.s} 0$$
$$\frac{\sum_{t=1}^{T} x_t}{T} \xrightarrow{a.s} \mu$$

de modo que[23]:

$$\frac{\sum_{t=1}^{T} x_t}{T} \xrightarrow{p} \mu \tag{15}$$

De la proposición 1 se tiene que si x_t es $\alpha-$mixing con coeficiente mixing $r/(r-1),\ r>1$ entonces x_t^2 es $\alpha-$mixing con coeficiente mixing $r/(r-1),\ r>1$. Adicionalmente se define

[21] Sea $\{a_n\}$ una sucesión en $\mathbb{R}$. Si $\lim_{n\longrightarrow\infty} a_n = c$, entonces $plim\, a_n = c$. Véase Polansky [26, Pág. 106] para una demostración.

[22] Notar que $E|X_t|^{r+\delta} \leq k < \infty$,$\delta > 1$ $r > 1 \forall t$ implica que$E|X_t|^{r+\delta} \leq k < \infty$,$\delta > 0$ $r > 1 \forall t$ dado que: $E\big[|X|^s\big] < \infty \Longrightarrow E\big[|X|^r\big] < \infty$ $para\, 0 < r \leq s$ Véase Loeve [21, Pág. 157].

[23] Notar que $X_n \xrightarrow{a.s} X \implies X_n \xrightarrow{p} X$. Véase Gut [9, pág. 209] para una demostración.

$\bar{\mu}_T^2 = \frac{1}{T}\sum_{t=1}^T E\left[X_t^2\right]$ y como $\sup E\left|X_t\right|^{r+\delta} \leq k < \infty$, $\delta > 1\,\forall t$ se deriva[24] que $\lim\limits_{T\longrightarrow\infty} \bar{\mu}_T^2 = \mu^2$. Entonces para $\sum_{t=1}^{T} x_t^2/T$ se tiene:

$$\frac{\sum_{t=1}^{T} x_t^2}{T} - \bar{\mu}_T^2 \xrightarrow{a.s} 0$$
$$\frac{\sum_{t=1}^{T} x_t^2}{T} \xrightarrow{a.s} \mu^2.$$

Por lo tanto:

$$\frac{\sum_{t=1}^{T} x_t}{T} \xrightarrow{p} \mu^2. \tag{16}$$

Asímismo, la expresión $\sum_{t=1}^{T} tx_t/T^2$ se puede expresar de la siguiente manera:

$$\begin{aligned}
\frac{\sum_{t=1}^{T} tx_t}{T^2} =& \frac{1}{T^2}\left[x_1 + 2x_2 + \ldots + Tx_T\right] \\
=& \frac{1}{T^2}\Big[x_1 + 2x_2 + \ldots + Tx_T + \big(x_2 + x_3 + \ldots + x_T\big) - \big(x_2 + x_3 + \ldots + x_T\big)\Big] \\
=& \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_2 + 2x_3 \ldots + (T-1)\,x_T\Big] \\
=& \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_2 + 2x_3 \ldots + (T-1)\,x_T + (x_3 + \ldots + x_T) - (x_3 + \ldots + x_T)\Big] \\
=& \frac{\sum_{t=1}^{T} x_t}{T^2} + \frac{\sum_{t=2}^{T} x_t}{T^2} + \frac{1}{T^2}\Big[x_3 + x_4 \ldots + (T-2)\,x_T\Big] \\
=& \,\vdots \\
=& \frac{1}{T^2} \times \sum_{t=1}^{T} x_t + \frac{1}{T^2} \times \sum_{t=2}^{T} x_t + \ldots + \frac{1}{T^2} \times \sum_{t=T-1}^{T} x_t + \frac{1}{T^2} \times x_t.
\end{aligned}$$

Dado que por la suposición 1 x_t cumple con $\sup E\left|X_t\right|^{r+\delta} \leq k < \infty$, $\delta > 1$ $r > 1\,\forall\, t$, lo cual implica que x_t es as acotada en L_p y se tiene que x_t es acotada estocásticamente[25] de modo que $x_t = O_p(1)\ \forall t$. Por otro lado, $\lim\limits_{T\longrightarrow\infty} \frac{1}{T^2} = 0$ implica que $\frac{1}{T^2} = o(1)$. Entonces:

$$\begin{aligned}
\frac{\sum_{t=1}^{T} tx_t}{T^2} =& \frac{1}{T^2} \times \sum_{t=1}^{T} x_t + \frac{1}{T^2} \times \sum_{t=2}^{T} x_t + \ldots + \frac{1}{T^2} \times \sum_{t=T-1}^{T} x_t + \frac{1}{T^2} \times x_t \\
=& o(1) \times O_p(1) + o(1) \times O_p(1) + \ldots + o(1) \times O_p(1) \\
=& o_p(1) \times O_p(1) + o_p(1) \times O_p(1) + \ldots + o_p(1) \times O_p(1) \\
=& o_p\,(1) + o_p\,(1) + \ldots + o_p\,(1) \\
=& o_p\,(1)\,.
\end{aligned}$$

[24] La sucesión $E\left[X_t^2\right]\ \forall\, t$ es convergente dado que es acotada y tiene una sub-sucesión convergente.

[25] Si $\sup_n E\left|X_n\right|^p < \infty$, entonces $X_n = O_p\,(1)$. Esto es, si $\{X_n\}$ es acotada en $L^p\,(P)$ entonces es acotada estocásticamente, Véase Bierens [1, Pág. 39].

Lo cual implica que:

$$\frac{\sum_{t=1}^{T} tx_t}{T^2} \xrightarrow{p} 0. \tag{17}$$

Por otro lado, aplicando la suposición 2 se tiene que la variable ε_t cumple con los supuestos de la proposición (4) dado que $\sup E\left|\varepsilon_t\right|^{r+\delta} \leq k < \infty$ $,\delta > 1,\ r > 1\ \forall\ t$ establece que la sucesión de expectativas de ε_t es acotada e integrable de orden $P = r + \delta$, lo que implica que $\limsup E\left|\varepsilon_t\right|^{r+\delta} < \infty$ $,\delta > 1,\ r > 1$. Mientras que $\sum_{i\in\mathbb{N}} \alpha(i)^{1-2/\beta} < \infty$ de la proposición (4) se cumple si $\alpha\left(i\right) = O\left(m^{\frac{-r}{(r-2)}-\varepsilon}\right)$[26], en la definición 5 para el tamaño de los coeficientes α-mixing de ε_t , $r/(r-1)$, de la suposición 2 se establece que $\alpha\left(t\right) = O\left(m^{\frac{-r}{(r-1)}-\varepsilon}\right)$ e implica que[27] $\alpha\left(t\right) = O\left(m^{\frac{-r}{(r-2)}-\varepsilon}\right)$ que cumple el supuesto de la proposición (4).

De modo que para el término $\sum_{t=1}^{T} \varepsilon_t/\sqrt{T}$ de la segunda matriz del lado derecho de la ecuación (11) se tiene que:

$$\begin{aligned}
&\frac{\sum_{t=1}^{T} \varepsilon_t}{\sigma_\varepsilon\sqrt{T}} \xrightarrow{d} W\left(t\right)\\
&\frac{\sum_{t=1}^{T} \varepsilon_t}{\sigma_\varepsilon\sqrt{T}} \xrightarrow{d} N\left(0,1\right)\\
&\frac{\sum_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_\varepsilon N(0,1).
\end{aligned} \tag{18}$$

Para el término $\sum_{t=1}^{T} t\varepsilon_t/T^{3/2}$ de la segunda matriz del lado derecho de la ecuación (11), se define la variable $Z_t = Z_{t-1} + \varepsilon_t$ con $Z_0 = 0$, realizando reemplazos sucesivos se tiene que $Z_t = \sum_{t=1}^{T} \varepsilon_t$. De lo anterior se deriva la siguiente expresión:

$$\begin{aligned}
\frac{\sum_{t=1}^{T} Z_{t-1}}{T} =& \frac{1}{T}\Big[0 + \varepsilon_1 + (\varepsilon_1 + \varepsilon_2) + \ldots + (\varepsilon_1 + \varepsilon_2 + \ldots + \varepsilon_{T-1})\Big]\\
=& \frac{1}{T}\Big[\left(T-1\right)\varepsilon_1 + \left(T-2\right)\varepsilon_2 + \ldots + \varepsilon_{T-1}\Big]\\
=& \frac{1}{T}\sum_{t=1}^{T}\left(T-t\right)\varepsilon_t\\
=& \sum_{t=1}^{T} \varepsilon_t + \frac{1}{T}\sum_{t=1}^{T} t\varepsilon_t.
\end{aligned}$$

Multiplicando ambos lados por $1/\sqrt{T}$ y desplazando términos se obtiene:

$$\frac{1}{T^{3/2}}\sum_{t=1}^{T} t\varepsilon_t = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \varepsilon_t - \frac{1}{T^{3/2}}\sum_{t=1}^{T} Z_{t-1}.$$

Previamente bajo la suposición 2 se demosotró que $\sum_{t=1}^{T} \varepsilon_t/\sqrt{T} \xrightarrow{d} \sigma_\varepsilon W(1)$, por otro lado

[26] Véase Phillips [24, Pág. 280].

[27] Si $x_n = O(n^\lambda)$ entonces para un $\delta > 0$, $x_n = O(n^{\lambda+\delta})$, véase White [28, Pág. 17].

se tiene que $\sum_{t=1}^{T} Z_{t-1}/T^{3/2} \xrightarrow{d} \sigma_\varepsilon \int_0^1 W(s)ds$ [28]. Entonces:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{d} \sigma_\varepsilon W(1) - \sigma_\varepsilon \int_0^1 W(s)ds = \int_0^1 sdW(s).$$

Y se tiene que[29]:

$$\int_0^1 sdW(s) \sim N(0, 1/3)$$

De modo que:

$$\frac{1}{T^{3/2}} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{d} N(0, 1/3) \tag{19}$$

De la suposición 1 se tiene que x_t es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1),\ r > 1$ y de la suposición 2 ε_t es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1),\ r > 1$ y mediante proposición (2) se tiene que $x_t\varepsilon_t$ es $\alpha-$mixing con coeficientes mixing de tamaño $r/(r-1),\ r > 1$. Sea $X = \sup\left|X_t\right|^{r+\delta}$ y $Y = \sup\left|\varepsilon_t\right|^{r+\delta}$, entonces se tiene

$$\begin{aligned} E\left|XY\right| &= E\left[\sup\left|X_t\right|^{r+\delta} \sup\left|\varepsilon_t\right|^{r+\delta}\right] \\ &\geq E\left[\sup\left|X_t\varepsilon_t\right|^{r+\delta}\right] \\ &\geq \sup E\left|X_t\varepsilon_t\right|^{r+\delta}. \end{aligned}$$

De la desiguladad de Holder[30]

$$\begin{aligned} E\left|XY\right| &\leq \sqrt{E\left|X\right|^2} \times \sqrt{E\left|Y\right|^2} \\ &= \left[-1 \cdot \sqrt{E\left|X\right|^2}\right] \cdot \left[-1 \cdot \sqrt{E\left|Y\right|^2}\right]. \end{aligned} \tag{20}$$

Usando la desigualdad de Jensen[31]

$$\begin{aligned} E\left|XY\right| &\leq E\left[-\sqrt{\left|X\right|^2}\right] \times E\left[-\sqrt{\left|Y\right|^2}\right] \\ &\leq E\big[-\left|X\right|\big] \times E\big[-\left|Y\right|\big] \\ &\leq E\left[X\right] \times E\left[Y\right] \\ &= E\Big[\sup\left|X_t\right|^{r+\delta}\Big] \times E\Big[\sup\left|\varepsilon_t\right|^{r+\delta}\Big] \\ &< \infty \end{aligned} \tag{21}$$

De (20) y (21) se obtiene

$$\sup E\left|X_t\varepsilon_t\right|^{r+\delta} < \infty \tag{22}$$

[28]Véase Hassler [13, Pág. 326] para una demostración de este resultado.

[29]Ibid [13, Pág. 307] para una demostración.

[30]Sea $p^{-1} + q^{-1} = 1$. Si $E\left|X\right|^p < \infty$ y $E\left|Y\right|^p < \infty$, entonces $\left|E\left[XY\right]\right| \leq E\left[\left|XY\right|\right] \leq \left\|X\right\|_p \times \left\|Y\right\|_q$, siendo $\left\|X\right\|_p := \left[\left\{E\left|X\right|^p\right\}\right]^{1/p}$. Este resultado es conocido como la desigualdad de Holder, una demostración en Gut [9, Pág. 129].

[31]La desigualdad de Jensen establece que para una variable aleatoria X, g una función convexa y que X y $g\left(X\right)$ son integrables, entonces $g\left(EX\right) \leq Eg\left(X\right)$. Véase Gut [9, Pág. 132].

Adicionalmente de la suposición 2 se cumple que $E\left[x_t\varepsilon_t\right] = 0$, para todo t y $\sigma^2_{x\varepsilon} = \lim_{T\longrightarrow\infty} E\left[\frac{1}{T}S^2_{x\varepsilon}\right]$, $\sigma^2_{x\varepsilon} > 0$ y junto al resultado (22) satisfacen los supuestos del teorema (4) para $x_t\varepsilon_t$. De modo que para $\sum_{t=1}^{T} x_t\varepsilon_t/\sqrt{T}$, se tiene:

$$\frac{\sum_{t=1}^{T} x_t\varepsilon_t}{\sigma_{x\epsilon}\sqrt{T}} \xrightarrow{d} W(t)$$
$$\frac{\sum_{t=1}^{T} x_t\varepsilon_t}{\sigma_{x\epsilon}\sqrt{T}} \xrightarrow{d} N(0,1)$$
$$\frac{\sum_{t=1}^{T} \varepsilon_t}{\sqrt{T}} \xrightarrow{d} \sigma_{x\epsilon}N(0,1) \tag{23}$$

Aplicando los resultados de las ecuaciones (12), (13), (14), (15), (16), (17), (18), (19) y (23) en la ecuación matricial (11) se tiene que:

$$\begin{bmatrix} \sqrt{T}\left(\hat{\alpha}-\alpha\right) \\ T^{3/2}\left(\hat{\delta}-\delta\right) \\ \sqrt{T}\left(\hat{\beta}-\beta\right) \end{bmatrix} \xrightarrow{L} \begin{bmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \mu^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{\varepsilon}N(0,1) \\ N(0,1/3) \\ \sigma_{x\epsilon}N(0,1) \end{bmatrix}. \tag{24}$$

Ahora se procede a hallar la matriz inversa de la primera matriz del lado derecho de la ecuación matricial (24). Primero se halla el determinante:

$$\begin{vmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \mu^2 \end{vmatrix} = \frac{\mu^2 - 4\left[\mu\right]^2}{12}.$$

Luego se halla la matriz adjunta:

$$adj\mathbf{A} = \begin{bmatrix} \frac{\mu^2}{3} & -\frac{\mu^2}{2} & -\frac{\mu}{3} \\ -\frac{\mu^2}{2} & \mu^2 - \left[\mu\right]^2 & \frac{\mu}{2} \\ -\frac{\mu}{3} & \frac{\mu}{2} & -\frac{2}{3} \end{bmatrix}.$$

Y la inversa:

$$\begin{bmatrix} 1 & \frac{1}{2} & \mu \\ \frac{1}{2} & \frac{1}{3} & 0 \\ \mu & 0 & \bar{\mu}^2_T \end{bmatrix}^{-1} = \begin{bmatrix} \frac{4\mu^2}{\triangle} & -\frac{6\mu^2}{\triangle} & -\frac{4\mu}{\triangle} \\ -\frac{6\mu^2}{\triangle} & \frac{12\left[\mu^2-\left[\mu\right]^2\right]}{\triangle} & \frac{6\mu}{\triangle} \\ -\frac{4\mu}{\triangle} & \frac{6\mu}{\triangle} & -\frac{8}{\triangle} \end{bmatrix}, donde\ \triangle = \mu^2 - 4\left[\mu\right]^2.$$

La ecuación matricial (24) queda como:

$$\begin{bmatrix}\sqrt{T}\big(\hat{\alpha}-\alpha\big)\\ T^{3/2}\big(\hat{\delta}-\delta\big)\\ \sqrt{T}\big(\hat{\beta}-\beta\big)\end{bmatrix} \xrightarrow{L} \begin{bmatrix}\dfrac{4\mu^2}{\triangle} & -\dfrac{6\mu^2}{\triangle} & -\dfrac{4\mu}{\triangle}\\ -\dfrac{6\mu^2}{\triangle} & \dfrac{12\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu}{\triangle}\\ -\dfrac{4\mu}{\triangle} & \dfrac{6\mu}{\triangle} & -\dfrac{8}{\triangle}\end{bmatrix}\begin{bmatrix}\sigma_\varepsilon N(0,1)\\ N\left(0,1/3\right)\\ \sigma_{x\epsilon}N(0,1)\end{bmatrix}$$

$$= \begin{bmatrix}\dfrac{4\mu^2}{\triangle} & -\dfrac{6\mu^2}{\triangle} & -\dfrac{4\mu}{\triangle}\\ -\dfrac{6\mu^2}{\triangle} & \dfrac{12\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu}{\triangle}\\ -\dfrac{4\mu}{\triangle} & \dfrac{6\mu}{\triangle} & -\dfrac{8}{\triangle}\end{bmatrix}\begin{bmatrix}\sigma_\varepsilon & 0 & 0\\ 0 & \dfrac{1}{3} & 0\\ 0 & 0 & \sigma_{x\epsilon}\end{bmatrix}\begin{bmatrix}N(0,1)\\ N\left(0,1\right)\\ N(0,1)\end{bmatrix}$$

$$= \begin{bmatrix}\dfrac{4\mu^2\sigma_\varepsilon}{\triangle} & -\dfrac{2\mu^2}{\triangle} & -\dfrac{4\mu\sigma_{x\epsilon}}{\triangle}\\ -\dfrac{6\mu^2\sigma_\varepsilon}{\triangle} & \dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle} & \dfrac{6\mu\sigma_{x\epsilon}}{\triangle}\\ -\dfrac{4\mu\sigma_\varepsilon}{\triangle} & \dfrac{2\mu}{\triangle} & -\dfrac{8\sigma_{x\epsilon}}{\triangle}\end{bmatrix}\begin{bmatrix}N(0,1)\\ N\left(0,1\right)\\ N(0,1)\end{bmatrix}$$

$$= \begin{bmatrix}\dfrac{4\mu^2\sigma_\varepsilon}{\triangle}N(0,1)-\dfrac{2\mu^2}{\triangle}N\left(0,1\right)-\dfrac{4\mu\sigma_{x\epsilon}}{\triangle}N(0,1)\\ -\dfrac{6\mu^2\sigma_\varepsilon}{\triangle}N\left(0,1\right)+\dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle}N\left(0,1\right)+\dfrac{6\mu\sigma_{x\epsilon}}{\triangle}N\left(0,1\right)\\ -\dfrac{4\mu\sigma_\varepsilon}{\triangle}N(0,1)+\dfrac{2\mu}{\triangle}N\left(0,1\right)-\dfrac{8\sigma_{x\epsilon}}{\triangle}N(0,1)\end{bmatrix}.$$

Y en consecuencia[32].

$$\begin{bmatrix}\sqrt{T}\big(\hat{\alpha}-\alpha\big)\\ T^{3/2}\big(\hat{\delta}-\delta\big)\\ \sqrt{T}\big(\hat{\beta}-\beta\big)\end{bmatrix} \xrightarrow{d} \begin{bmatrix}N\left(0,\left[\dfrac{4\mu^2\sigma_\varepsilon}{\triangle}\right]^2\right)+N\left(0,\left[-\dfrac{2\mu^2}{\triangle}\right]^2\right)+N\left(0,\left[-\dfrac{4\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right)\\ N\left(0,\left[-\dfrac{6\mu^2\sigma_\varepsilon}{\triangle}\right]^2\right)+N\left(0,\left[\dfrac{4\left[\mu^2-[\mu]^2\right]}{\triangle}\right]^2\right)+N\left(0,\left[\dfrac{6\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right)\\ N\left(0,\left[-\dfrac{4\mu\sigma_\varepsilon}{\triangle}\right]^2\right)+N\left(0,\left[\dfrac{2\mu}{\triangle}\right]^2\right)+N\left(0,\left[-\dfrac{8\sigma_{x\epsilon}}{\triangle}\right]^2\right)\end{bmatrix}.$$

[32]Notar que si $X\sim N\left(\mu,\sigma^2\right)$ entonces para una constante $c\in\mathbb{R}$ se tiene que $aX\sim N\left(c\mu,c\sigma^2\right)$, véase Khuri [18, Pág. 66] para una demostración.

Finalmente

$$\begin{bmatrix}\sqrt{T}(\hat{\alpha}-\alpha)\\ T^{3/2}(\hat{\delta}-\delta)\\ \sqrt{T}(\hat{\beta}-\beta)\end{bmatrix} \xrightarrow{d} \begin{bmatrix} N\left(0,\left[\frac{4\mu^2\sigma_\varepsilon}{\triangle}\right]^2+\left[-\frac{2\mu^2}{\triangle}\right]^2+\left[-\frac{4\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0,\left[-\frac{6\mu^2\sigma_\varepsilon}{\triangle}\right]^2+\left[\frac{4\left[\mu^2-[\mu]^2\right]}{\triangle}\right]^2+\left[\frac{6\mu\sigma_{x\epsilon}}{\triangle}\right]^2\right) \\ N\left(0,\left[-\frac{4\mu\sigma_\varepsilon}{\triangle}\right]^2+\left[\frac{2\mu}{\triangle}\right]^2+\left[-\frac{8\sigma_{x\epsilon}}{\triangle}\right]^2\right)\end{bmatrix}. \qquad (25)$$

De la ecuación matricial (25) se evidencian tres hechos importantes, primero que cada uno de los elementos los estimadores de $(\mathbf{b}-\beta)$ para los estimadores $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$ del modelo de regresión (9) convergen a una distribución normal con media 0 pero con diferentes varianzas, segundo cada una de las varianzas de $(\mathbf{b}-\beta)$ en la ecuación matricial (25) depende no solo de los propios parámetros poblacionales sino de los parámetros de otras variables, y tercero cada una de las variables en $(\mathbf{b}-\beta)$ converge a la distribución normal a una tasa distinta para el caso de los estimadores de $\hat{\alpha}$ y $\hat{\beta}$ convergencen a una tasa $\sqrt{T}$ mientras que $\hat{\delta}$ converge a una tasa $T^{3/2}$.

5. Conclusiones

La determinación de la distribución de los estimadores en modelos de regresión juega un papel preponderante en los procedimientos inferenciales en los modelos de regresión como se evidenció en la sección 4 del presente trabajo.

Los tres puntos del último párrafo de la sección 4 tienen consecuencias importantes en el proceso de inferencia en los modelos de regresión lineal cuandos las variables siguen procesos strong mixing con componentes tendenciales dado que en el proceso inferencial tradicional de regresiones con variables explicativas determinísticas, el vector $(\mathbf{b}-\beta)$ sigue distribución normal $N(0,1)$ de modo que es posible usar tablas estadísticas convencionales para los diversos test de significancia individual y conjunta, mientras que los estimadores de los regresores strong mixing se distribuen de acuerdo a (25). De modo que aplicar estos procedimientos en regresiones con variables strong mixing puede llevar erroneamente a establecer relaciones estadísticas cuando puede no existir ninguna. Por otro lado, dado que las varianzas de las distribuciones para las variables $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$ en la ecuación matricial (25) dependen de parametros poblacionales, es necesario estimar estos parámetros a la hora de realizar el proceso inferencial sobre los estimadores $\hat{\alpha}$, $\hat{\delta}$, $\hat{\beta}$.

De manera general los procesos de inferencia estadística en los modelos de regresión con datos observacionales, datos provenientes de la observación pasiva en contraste con datos que provienen de muestras o experimentos, no deben de ser llevados de forma mecánica, y de manera particular con regresores que siguen procesos dependientes strong mixing y componentes tendenciales dado que se corre el riesgo de cometer errores tipo I o tipo II y asignar una relación estadística cuando en realidad no existe ninguna.

Referencias bibliográficas

[1] Bierens, H. J. (1994). *Topics in advanced econometrics: estimation, testing, and specification of cross-section and time series models.* New York: Cambridge University Press.

[2] Borodin, A. y Salminen, P. (2002). *Handbook of brownian motion: facts and formulae* (2a ed.). Boston: Birkhäuser.

[3] Bradley, R. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*, 2, 107-144.

[4] Capasso, V. y Bakstein, D. (2012). *An introduction to continuous-time stochastic processes: theory, models, and applications to finance, biology, and medicine* (2a ed.). Boston: Birkhäuser.

[5] Doob, J. (1990). *Stochastic processes.* United States of America: John Wiley & Sons.

[6] Durlauf, S. y Phillips, P. (1988). Trends versus random walks in time series analysis. *Econometrica*, 56(6), 1333-1354.

[7] Golberg, M. y Cho, H. (2003). *Introduction to regression analysis.* USA: WIT Press

[8] Greene, W. H. (2012). *Econometric analysis* (2a ed.). New York: Pearson/Prentice Hall.

[9] Gut, A. (2013). *Probability: a graduate course.* New York: Springer Science+Business Media.

[10] Halbert White, I. D. (1984). Nonlinear regression with dependent observations. *Econometrica*, 52(1), 143-161.

[11] Haldrup, N. (1994). The asymptotics of single-equation cointegration regressions with i(1) and i(2) variables. *Journal of Econometrics*, 63(1), 153-181.

[12] Hasseler, U. (2000). Simple regressions with linear time trends. Journal of Time Series Analysis, 21(1), 27-32.

[13] Hassler, U. (2016). *Stochastic Processes and Calculus. An Elementary Introduction with Applications.* New York: Springer.

[14] Herrndorf, N. (1984, 02). A functional central limit theorem for weakly dependent sequences of random variables. *Annals of probability*, 12(1), 141-153.

[15] Joon Y. Park, P. C. B. P. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3), 468-497.

[16] Joon Y. Park, P. C. B. P. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5(1), 95-131.

[17] Karlin, S., y Taylor, H. M. (1975). *A first course in stochastic processes.* (2.a ed.) New York: Academic Press.

[18] Khuri, A. (2010). *Linear Model Methodology.* London: CRC Press.

[19] Kobayashi, H., Mark, B. L., y Turin, W. (2012). *Probability, random processes, and statistical analysis.*Cambridge; New York: Cambridge University Press.

[20] Krämer, W., y Marmol, F. (2002). Ols-based asymptotic inference in linear regression models with trending regressors and ar(p)-disturbances. *Communications in Statistics - Theory and Methods*, 31(2), 261-270.

[21] Loeve, M. (1978). *Probability theory II* (4a ed.). New York: Springer-Verlag .

[22] McLeish, D. L. (1975, 10). A maximal inequality and dependent strong laws. *The Annals of Probability*, 3(5), 829-839.

[23] Mynbaev, K. T. (2010). Asymptotic distribution of the ols estimator for a mixed spatial model. *Journal of Multivariate Analysis*, 101(3), 733 - 748.

[24] Phillips, P. C. B. (1987). Time series regression with a unit root. *Econometrica*, 55(2), 277-301.

[25] Phillips, P. C. B., y Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, 53(4), 473-495.

[26] Polansky, A. M. (2011). *Introduction to statistical limit theory*. London: CRC Press.

[27] Stock, J., Sims, C., y Watson, M. (1990). Inference in linear time series models with some unit roots. *Econometrica*, 58(1), 113-144.

[28] White, H. (2000). *Asymptotic Theory for Econometricians*. New York: Academic press.

[29] Zhang, L.-X., y Zhang, Y. (2015). Asymptotics for a class of dependent random variables. *Statistics & Probability Letters*, 105(Supplement C), 47 - 56.

[30] Zhengyan, L., y Chuanrong, L. (1996). *Limit theory for mixing dependent random variables*. Netherlands: Springer.