Econometría 2: Notas de clase

Andrés Vargas

Contents

List of Figures

List of Tables

1 Estimador MCO en datos de corte transversal

1.1 Mecánica del estimador, valor esperado y varianza

Considere el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_k x_k + u \tag{1}$$

Tenemos $k+1$ parámetros poblacionales desconocidos. A partir de una muestra de datos debemos estimar los parámetros. Si $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k$ es el valor estimado de y, entonces el residual del modelo es $y - \hat{y}$. Dado un conjunto dado de observaciones $\{(y_i, x_{i1}, x_{i2}, ..., x_{ik}) : i = 1, ..., n\}$. La idea es encontrar los $\hat{\beta}_0, ..., \hat{\beta}_k$ que minimicen

$$\sum_{i=1}^{n} (y_i - \hat{\beta}0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik})^2 \tag{2}$$

Como los parámetros poblacionales los estimamos a partir de una muestra, entonces debemos tener en cuenta que si hubiésemos tenido una muestra diferente entonces habríamos obtenidos valores estimados diferentes. En otras palabras, el conjunto de datos que utilizamos es uno de tantos posibles. Esto quiere decir que debemos tener en cuenta la variabilidad muestral. En este sentido, los parámetros podrían verse como variables aleatorias y por lo tanto debemos preguntarnos por su valor esperado, varianza, covarianza, y distribución de probabilidad. Para ello, hacemos los siguientes supuestos

- S1: El modelo poblacional puede ser escrito como en 1
- S2: Tenemos una muestra aleatoria de n observaciones. Las variables son aleatorias porque no conocemos su valor a priori. Cada individuo es seleccionado aleatoriamente de la población y las observaciones de las variables son estadísticamente independientes
- S3: No hay colinealidad perfecta
- S4: Exogeneidad estricta. $E(u|x_1, x_2, ..., x_k) = 0$. Los valores de x_j no pueden ser utilizados para predecir u

Bajo los supuestos S1 a S4 podemos mostrar que

$$E(\hat{\beta}_j) = \beta_j, \quad j = 0, ..., k \tag{3}$$

Es decir, el estimador MCO es un estimador insesgado de los parámetros poblacionales. Esta es una propiedad muy importante. Lo que nos dice es que, en promedio, de todas las muestras posibles de la población, el estimador es *correcto* en promedio.

Para verlo, recuerde que el estimador MCO puede escribirse en forma matricial como

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{4}$$

Donde $\mathbf{X}$ es la matriz de datos, de dimensión nx(k+1), $\mathbf{Y}$ contiene las observaciones de y para cada $i = 1, ..., n$ y es de dimensión nx1 y $\hat{\boldsymbol{\beta}}$ de dimensión (k+1)x1. La ecuación anterior puede ser escrita como

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) \tag{5}$$

Si tomamos el valor esperado condicional a ambos lados

$$E(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{u}|\mathbf{X}) \tag{6}$$

Si se cumple S4 entonces concluimos que

$$E(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \boldsymbol{\beta}$$

Ahora, la varianza del estimador la podemos calcular a partir de la ecuación (5)

$$Var(\hat{\boldsymbol{\beta}}|\mathbf{X}) = Var(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}) \tag{7}$$

Si suponemos que

- S5: $Var(u_i|\mathbf{X}) = \sigma^2; i = 1, ..., n$ Homocedasticidad

Podemos escribir (7) como

$$Var(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

Bajos los supuestos S1-S5 el estimador MCO es el mejor estimador lineal insesgado.

1.2 Efecto causal

Si tomamos el valor esperado condicional de (1) y usamos S4

$$E(y|\mathbf{x}) = \beta_0+\beta_1x_1+\beta_2x_2+...+\beta_kx_k+E(u|\mathbf{x}) = \beta_0+\beta_1x_1+\beta_2x_2+...+\beta_kx_k \tag{9}$$

A esta ecuación la llamamos la función de regresión poblacional. Nos dice que dado $\mathbf{x}$, el promedio de los errores es cero y que cualquier cambio en x_j no está correlacionado con un cambio en u, luego un cambio en x_j causa un cambio en el valor esperado condicional de y. Es decir

$$\frac{\partial E(y|\mathbf{x})}{\partial x_1} = \beta_1 \tag{10}$$

Es decir que el parámetro indica cuánto cambia el valor esperado condicional de y cuando cambia x_1 en una unidad y todo lo demás está constante. Es decir, el cambio en y atribuible a x_1. Note la relación entre el insesgamiento del parámetro y la noción de causalidad. En investigación se considera que la mejor práctica son los experimentos aleatorios controlados. En este mundo ideal se asignan valores aleatoris de x (tratamiento) y se observa el resultado de y. Si hay una relación sistemática entre los cambios en x y los cambios en y entonces podemos decir que x causa a y. Los demás factores aleatorios, u, que podrían afectar a y son estadísticamente independiente de x. En este sentido, la exogeneidad estricta nos dice que x es tan buena como si hubiese sido asignada aleatoriamente

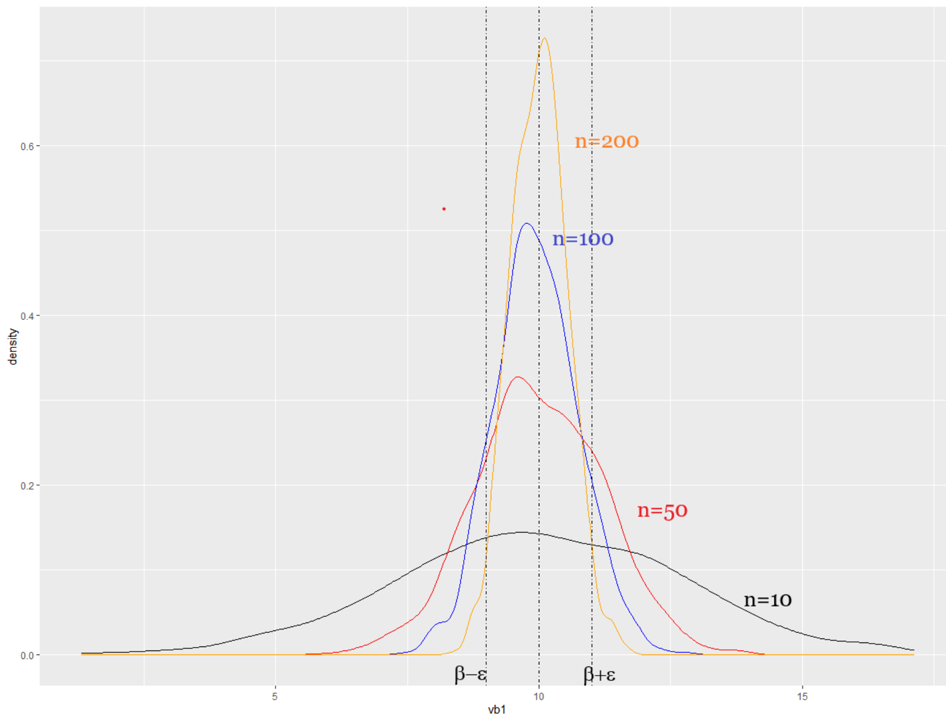

Figure 1: Consistencia

1.3 Propiedades en muestras grandes del estimador MCO

Examinamos las propiedades del estimador cuando la muestra tiende a infinito. En la medida que la muestra es más grande su comportamiento será más parecido, se aproximará al de una muestra que se aproxima a infinito. Estas propiedades se conocen como propiedades asintóticas. Suponga que consideramos satisfactorio obtener $\hat{\beta}_1$ a una distancia ϵ del valor verdadero β_1. La probabilidad de obtener una estimación cercana a β_1 es

$$P(\beta_1 - \epsilon \leq \hat{\beta}_1 \leq \beta_1 + \epsilon) \tag{11}$$

El estimador es consistente si esta probabilidad converge a 1 en la medida que $N \to \infty$

$$\lim_{N \to \infty} P(\beta_1 - \epsilon \leq \hat{\beta}_1 \leq \beta_1 + \epsilon) = 1 \tag{12}$$

Para que el estimador sea consistente es suficiente con que $Cov(x_j, u) = 0$, lo cual es un supuesto más débil que $E(u|x_1) = 0$. Es más débil porque la exogeneidad estricta implica covarianza cero, pero no lo contrario

2 Endogeneidad

¿Qué pasa si alguna de las variables está correlacionada con el termino de error? Esto implica una violación del supuesto S4, y por lo tanto da lugar a estimaciones sesgadas. Es decir, que los parámetros que estimamos por MCO no tendrán interpretación causal. Algunas de las razones por las cuales la variables explicativas pueden estar correlacionadas con el error son las siguiente

1. Omisión de variable relevante: no incluye en la regresión una variable relevante y está variable está correlacionada con una variable explicativa. Si el modelo poblacional es

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

   Pero estima

$$y = \beta_0 + \beta_1 x_1 + \upsilon$$

   Note que en el modelo estimable $\upsilon = u + \beta_2 x_2$. Al tomar el valor esperado condicional de este modelo obtenemos

$$E(y|x_1) = \beta_0 + \beta_1 x_1 + E(\upsilon|x_1)$$

   Donde $E(\upsilon|x_1) = \beta_2 E(x_2|x_1) + E(u|x_2)$. El último término es cero. Ahora, si $\beta_2 \neq 0$ y $E(x_2|x_1) \neq 0$ entonces $E(\upsilon|x_1) \neq 0$ y por lo tanto no se cumple S4. Defina $E(x_2|x_1) = \alpha_0 + \alpha_1 x_1$ y tome la derivada parcial $\partial E(y|x1)/\partial x_1$ ¿Cómo interpreta β_1?

2. Simultaneidad o causalidad en reversa: esto ocurre cuando al menos una de las variables explicativas se determina simultáneamente con y. Por ejemplo, y es el volumen de comercio y x los aranceles, y usted está interesado en estimar el efecto de las reducciones arancelarias sobre el volumen de comercio, así que usa el siguiente modelo

$$comercio = \beta_0 + \beta_1 arancel + \upsilon$$

   Sin embargo, los aranceles también están determinados por el volumen de comercio. A mayor comercio hay más esfuerzo por parte de grupos de presión para ser protegidos de las importaciones, luego los aranceles son más altos.

$$arancel = \gamma_0 + \gamma_1 comercio + \omega$$

   Luego

$$comercio = \beta_0 + \beta_1(\gamma_0 + \gamma_1 comercio + \omega) + \upsilon$$

   El valor de los aranceles depende del valor del comercio, luego choques que afectan el comercio, υ (ej. coronavirus) afectan también a los aranceles, luego υ y *arancel* están correlacionados y por lo tanto β_1 no captura el efecto causal.

3. Error de medición: queremos usar la variable x y pero nuestra medida de la variable es imperfecta, $\tilde{x}$, y usamos esta medida en nuestra regresión. Si el error de medición está correlacionado con el error del modelo entonces tenemos un problema de endogeneidad.

2.0.1 Ejercicio en clase

1. El ahorro personal de un individuio i se basa en su ingreso permanente o de largo plazo. Si y_i es el ahorro anual y x_i^* el ingreso permanente anual, un modelo simple que representa esta relación es

$$y_i = \beta_0 + \beta_1 x_i^* + u_i$$

   En la práctica x_i^* es muy difícil, si no imposible de medir. Se nos ocurre usar el ingreso actual como una medida del ingreso permanente, sin embargo esta es una medida imperfecta

$$x_i = x_i^* + \upsilon_i$$

   Donde υ_i es una perturbación aleatoria con media 0 y varianza σ_υ^2. Asumimos que u_i y υ_i son independientes. Así, se estima el modelo

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

   Explique, y muestre, porque el estimador MCO de β_1 no es consistente

2. Sea p_i el precio de equilibrio y q_i la cantidad de equilibrio. Plantea el modelo

$$q_i = \beta_0 + \beta_1 p_i + e_i$$

   Explique porque la estimación por MCO del parámetro β_1 no es consistente

3. Usted quiere estimar el efecto de la disponibilidad y accesibilidad a Facebook sobre el número de protestas que se presentan en un país en un mes dado. Para ello plantea el siguiente modelo[1]

$$protestas_{ct} = \beta_0 + \beta_1 Face_{ct} + \epsilon_{ct}$$

   Donde *protestas* es el número de protestas en el mes t en el país c y $Face$ mide la proporción de personas que tiene acceso potencial a Facebook. Dado que la protesta es una forma de acción colectiva que requiere coordinación y cooperación, esta puede estar determinada por características sociales, políticas y culturales de los países. Piense en que tipo de variables podrían recoger estas características y analice si estas están relacionadas con la variable $Face$. A partir de lo anterior indique si la estimación del parámetro β_1 sufre de un problema de endogeneidad por variable omitida.

4. Use la base de datos "canastaiv.dta". La base de datos contiene información sobre el estatus nutricional de niños. Una parte de los niños participó en un programa nutricional llamado "canasta", a través del cual se le entregaba un mercado suplementario a una familia. A este grupo lo

[1] Ver Ferguson y Molina, 2019, Facebook causes protests, *Documentos CEDE*, 2019-41, https://economia.uniandes.edu.co/components/com_booklibrary/ebooks/dcede2019-41.pdf

llamaremos “tratamiento”. Los niños que no participaron en el programa son el grupo “control”. La participación en el programa fue voluntaria. Las madres que quisieran inscribir a sus niños y recibir el mercado debían acercarse a una oficina a realizar la inscripción.

Las variables son:
- Ingresos_hogar_jefe: ingresos del jefe de hogar en decenas de miles $
- Personas: número de personas en el hogar
- Educa_jefe: años de educación del jefe de hogar
- Ocupado_jefe: jefe ocupado
- Hombre: niño hombre
- Orden_n: orden de nacimiento del niño
- D: Si el niño hace parte del programa. Tratamiento
- Of_op: oficinas operadoras en el municipio
- Ha_nchs: z-score talla para la edad
- Distancia: distancia en metros a la oficina administradora del programa más cercana

Usted quiere evaluar el programa, para ello plantea el siguiente modelo

$$ha\_nchs = \alpha + \gamma D + \beta_1 personas + \beta_2 orden + \beta_3 ocupado + \beta_4 educajefe + \beta_5 ingresohogarjefe + e$$

(a) Estime el modelo por MCO e interprete los parámetros en términos de magnitud y significancia

(b) Le han dicho que $\hat{\gamma}$ no es un estimador consistente de γ y que por lo tanto no captura el impacto causal del programa sobre el estatus nutricional. Explique esta afirmación

(c) Como la asignación al programa no es aleatoria $E(\hat{\gamma}) \neq \gamma$ ¿por qué podría ser esto? (pista: piénselo como un problema de variable omitida)

3 Variables instrumentales

Feyrer (2009)[2,3] estima el siguiente modelo para estimar el efecto del comercio sobre el crecimiento económico

$$lny_{it} = \alpha + \gamma_i + \gamma_t + \beta lntrade_{it} + \epsilon_{it}$$

La estimación consistente de β requiere que $corr(\epsilon_{it}, lntrade_{ite}) = 0$ ¿Es este supuesto plausible? Feyrer estima en una regresión OLS $\hat{\beta} = 0.3(0.053)$. Sin embargo, este estimador quizás no sea apropiado ¿El comercio aumenta el ingreso o el ingreso aumenta el ingreso? Quizás países más ricos comercian más

[2] Feyrer, J., 2009, Distance, trade and income-The 1967 to 1975 closing of the Suez Canal as a natural experiment, https://www.dartmouth.edu/ jfeyrer/Feyrer_Suez2009_11_22.pdf

[3] Esta entrada resume el paper https://voxeu.org/article/1967-75-suez-canal-closure-lessons-trade

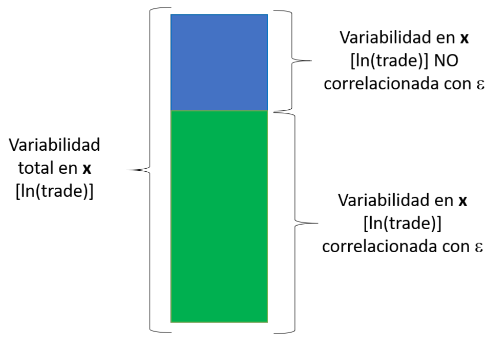

Figure 2: IV

porque prefieren una mayor variedad de productos extranjeros. Causalidad en reversa.

$$lntrade_{it} = \pi + \omega lny_{it} + \upsilon_{it}$$

Usando la ecuación anterior

$$ln(trade_{it}) = \pi + \omega(\alpha + \gamma_i + \gamma_t + \beta lntrade_{it} + \epsilon_{it}) + \upsilon_{it}$$

Es claro que $corr(\epsilon_{it}, lntrade_{ite}) \neq 0$ ¿Cómo podemos solucionarlo?

El objetivo es extraer la variabilidad exógena. Eso lo podemos hacer a través de una variable correlacionada con $lntrade_{it}$ pero no con ϵ_{it}. Generamos variabilidad exógena en $lntrade_{it}$ a partir de una regresión contra la nueva variable. Usamos esta variabilidad exógena en nuestro modelo. Feyrer (2009) usa el cierre del Canal del Suéz entre 1967 y 1975 para generar la variabilidad exógena en $lntrade_{it}$. Si se plantea como un modelo en dos etapas, sería

$$lntrade_{it} = \delta_0 + \theta Suez_t + r_{it}$$

De donde se obtiene: $\widehat{lntrade}_{it} = \hat{\delta}_0 + \hat{\theta} Suez_t$. Que luego se usa en

$$lny_{it} = \alpha + \gamma_i + \gamma_t + \beta \widehat{lntrade}_{it} + \epsilon_{it}$$

Feyrer (2009) estima $\hat{\beta} = 0.23(0.087)$ Recuerde que en OLS había obtenido $\hat{\beta} = 0.3(0.053)$

3.1 Estimador IV

Considere el modelo

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u \tag{13}$$

$$E(u) = 0 \quad Cov(x_j, u = 0) \quad j = 1, 2, ..., k-1 \tag{14}$$

Las variables x_j son exógenas y la variable x_k es endógena. Para usar IV necesitamos una variable z que no está en 13, que satisface 2 condiciones. Primero, que z no esté correlacionada con u

$$Cov(z,u) = 0 \tag{15}$$

Y segundo, que $\theta \neq 0$ en la proyección lineal

$$x_k = \delta_0 + \delta_1 x_1 + ... + \delta_{k-1} x_{k-1} + \theta z + r_k \tag{16}$$

Cuando z satisface estas dos condiciones entonces decimos que es una variable instrumental, o instrumento. Si estas condiciones se cumplen, entonces podemos decir que z resulve el problema de identificación. Es decir, que podemos escribir β_j en términos de los momentos poblacionales. Decimos entonces que el estimador de variables instrumentales es consistente

$$plim \quad \hat{\boldsymbol{\beta}}_{\boldsymbol{IV}} = \boldsymbol{\beta} \tag{17}$$

Una implicación importante del método es la pérdida de información. Si observa la gráfica anterior se dará cuenta que del total de información contenida en la variable solo usaremos una parte, área azul, que es aquella que creemos que es exógena. Esta pérdida de información tiene como consecuencia una menor precisión al la hora de estimar los parámetros. Es decir, el error estándar para el coeficiente de la variable endógena será mayor. Supongamos que tenemos solo una variable del lado derecho, más la constante, (modelo simple). La condición de homocedasticidad es

$$E(u^2|z) = \sigma^2 \tag{18}$$

Se puede mostrar, bajo las condiciones que cumple z que la varianza asintótica de $\hat{\beta}_1$ es

$$\frac{\sigma^2}{n\sigma_x^2 \rho_{x,z}^2} \tag{19}$$

$\rho_{x,z}$ es la correlación entre x y z. Es claro que la varianza del estimador de variables instrumentales es mayor a la varianza del estimador MCO, y que ello depende la intensidad con la que el instrumento y la endógena estén correlacionadas.

Ejercicio

Lea el documento *Forest and conflict in Colombia* escrito por Barry Reilly y Rafael Parra-Peña[4], e identifique y analice lo siguiente

- ¿Cuáles son los problemas de endogeneidad que planetan los autores?
- ¿Cuáles son los instrumentos seleccionados y por qué se argumenta que solucionan los problemas de endogeneidad?

[4] Reilly B., Parra-Peña, R., 2019, Forest and Conflict in Colombia, Archivos de Economía DNP, https://colaboracion.dnp.gov.co/CDT/Estudios%20Econmicos/492.pdf

- Compare las estimaciones de la especificación MCO, tabla 1.5.7, y la especificación con instrumentos, tabla 1.5.6. Compare tanto los coeficientes como los errores estándar ¿Qué pasa con la significancia de la variable de interés?

4 Modelos de variable dependiente binaria

Nos interesa modelar situaciones en las que la variable dependiente recoge la ocurrencia o no de un evento. En algunas ocasiones el evento es una decisión, por ejemplo cuando el consumidor decide cambiar de operador de telefonía móvil, y en otros casos el evento no resulta de un acto deliberado, por ejemplo tener cáncer. En cualquiera de estos casos la variable dependiente se caracteriza de la siguiente manera

$$y = \begin{cases} 1 & \text{Si ocurre el evento} \\ 0 & \text{Si no ocurre el evento} \end{cases} \tag{20}$$

En general podemos decir que estos modelos se usan con dos grandes propósitos: 1) inferencia causal para análisis de política; 2) predicción para clasificación. En el primer caso la discusión de estos modelos es una extensión de lo que hasta ahora hemos estudiado sobre problemas de endogeneidad. El segundo propósito se aprecia cuando lo que nos interesa es identificar quienes están en un mayor riesgo de caer en una situación particular o tomar una decisión específica. Por ejemplo, en un modelo de scoring de crédito nos interesa clasificar a los solicitantes de acuerdo a la probabilidad de impago. Acá lo importante es la capacidad del modelo de clasificar correctamente, mientras que probar alguna hipótesis particular es secundario.

En los problemas de transporte un problema importante es modelar la decisión sobre medio de transporte que toman los individuos. Por ejemplo queremos modelar la probabilidad de llegar en carro particular a la universidad. Definimos $y_i = 1$ si el individuo llega en carro particular y $y_i = 0$ si llega en otro medio de transporte. Por lo pronto decimos que dicha probabilidad es una función de la variable x_i. Suponga que tratamos este problema usando lo que sabemos del modelo de regresión lineal y el estimador MCO. Planteamos el siguiente modelo para estimar

$$y_i = \beta_0 + \beta_1 x_i + u \tag{21}$$

El valor esperado condicional de y_i es

$$E(y_i|x) = 1P(y_i = 1|x) + 0P(y_i = 0|x) \tag{22}$$

Si asumimos $E(u|x) = 0$ entonces

$$E(y_i|x) = P(y_i = 1|x) = \beta_0 + \beta_1 x_i \tag{23}$$

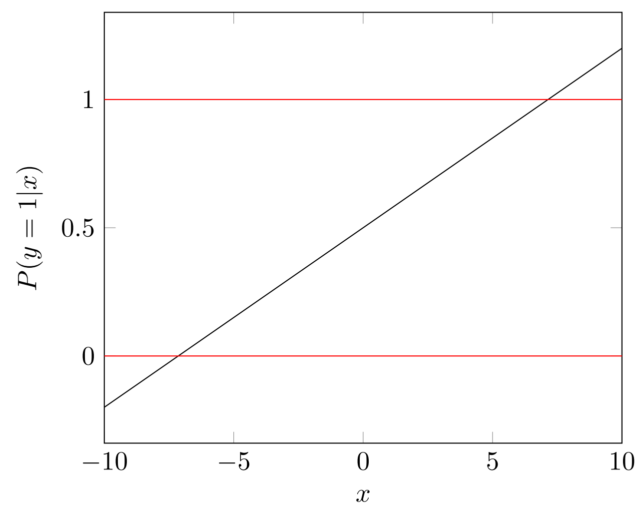

Figure 3: Modelo de probabilidad lineal $E(y|x) = 0.5 + 0.07x$

Note que en este modelo $\beta_0 + \beta_1 x_i$ es una probabilidad. Lo llamamos el modelo de probabilidad lineal. Como se observa en la figura 3 este modelo puede generar probabilidades predichas por fuera del intervalo $[0, 1]$. Adicionalmente, los errores de este modelo son heterocedásticos, veamos

$$u_i = y_i - \beta_0 + \beta_1 x_i \tag{24}$$

Como y_i solo puede tomar 2 valores, entonces debe ser cierto que si $y = 1$ entonces $\beta_0 + \beta_1 x + u = 1$, y si $y = 0$ entonces $\beta_0 + \beta_1 x + u = 0$

$$P(u_i = -\beta_0 - \beta_1 x) = P(y_i = 0|x_i) = 1 - (\beta_0 + \beta_1 x) \tag{25}$$
$$P(u_i = 1 - \beta_0 - \beta_1 x) = P(y_i = 1|x_i) = \beta_0 + \beta_1 x \tag{26}$$

Luego la varianza es

$$V(u_i|x_i) = (\beta_0 + \beta_1 x)[1 - (\beta_0 + \beta_1 x)] \tag{27}$$

Note que la varianza depende del valor que tome las variables independientes, luego es heterocedástica. Una opción es usar errores estándar robustos.

4.1 Modelos logit y probit

La idea es usar una función $G(\mathbf{x}\boldsymbol{\beta})$, donde $\mathbf{x}$ es $1xK$ y $\boldsymbol{\beta}$ es $Kx1$, que lleva a $\mathbf{x}\boldsymbol{\beta}$ a un valor en el intervalo $(0, 1)$. Esto se conoce como un *index model* porque lo que hace es restringir la forma como la probabilidad de respuesta depende de $\mathbf{x}$

$$P(y = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}) \tag{28}$$

La función $G(z)$ sueles ser una función acumulada de distribución (cdf). Si usamos la densidad de la normal estándar tenemos el modelo probit

$$G(z) = \Phi(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} exp(-z^2/2)dz \tag{29}$$

Si usamos la logística estándar tenemos el modelo logit

$$G(z) = \Lambda(z) = \frac{e^z}{1+e^z} \tag{30}$$

Recuerde que el termino de la integral de 29 es la pdf normal ϕz.

4.2 Efectos marginales

Al igual que en los modelos de regresión lineal, nuestro interés consiste en estimar el efecto que tiene una variación en una vairbale x sobre la variabl de resultado y. En los modelos de regresión lineal esta información nos la entrega el parámetro β. En los modelos de respuesta binaria esto no es así. Veamos

Queremos saber el efecto parcial de un cambio en una independiente, continua, sobre la probabilidad de ocurrencia del evento

$$\frac{\partial Pr(y=1|\mathbf{x})}{\partial x_j} = g(\mathbf{x}\boldsymbol{\beta})\beta_j; \quad \text{donde} \quad g(z) = \frac{\partial G(z)}{\partial z} \tag{31}$$

En el caso de probit $g(z) = \phi(z)$ y si es logit $g(z) = \dfrac{e^z}{(1+e^z)^2}$

Note que β_j solo indica la dirección del efecto, su magnitud depende de los valores que tomen las demás variables. En general, se usan las siguientes alternativas.

- PEA: Efecto parcial en la media de las variables

$$\phi(\bar{\mathbf{x}}\hat{\boldsymbol{\beta}})\hat{\beta}_j \tag{32}$$

  Donde $\bar{\mathbf{x}}$ es el promedio de las variables

- APE: Efecto parcial promedio para variables continuas

$$\hat{\beta}_j[N^{-1}\sum_{i=1}^{N}\phi(\mathbf{x}\hat{\beta})] \tag{33}$$

APE es preferido cuando el propósito es evaluar política, pues es el efecto promedio de tratamiento. En estos casos la variable de tratamiento es una dummy que indica si el individuo fue tratado o no (control). En ese caso, cuando x_3 es la dummy de tratamiento el APE es

$$N^{-1}\sum_{i=1}^{N}[G(\hat{\beta}_1 + \hat{\beta}_{i2}x_{i2} + \hat{\beta}_{i3}) - G(\hat{\beta}_1 + \hat{\beta}_{i2}x_{i2})] \tag{34}$$

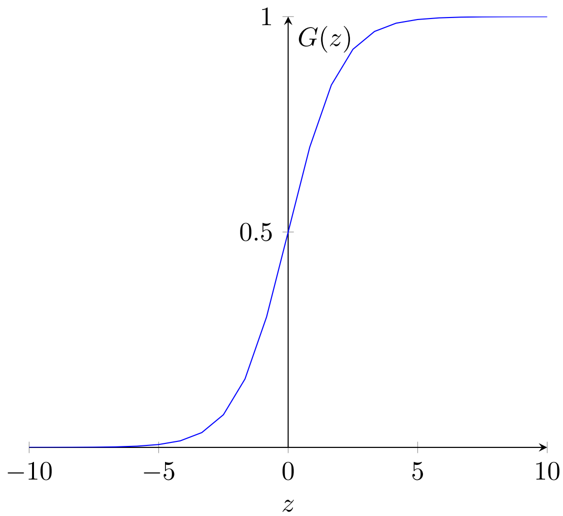

Figure 4: CDF Logistica

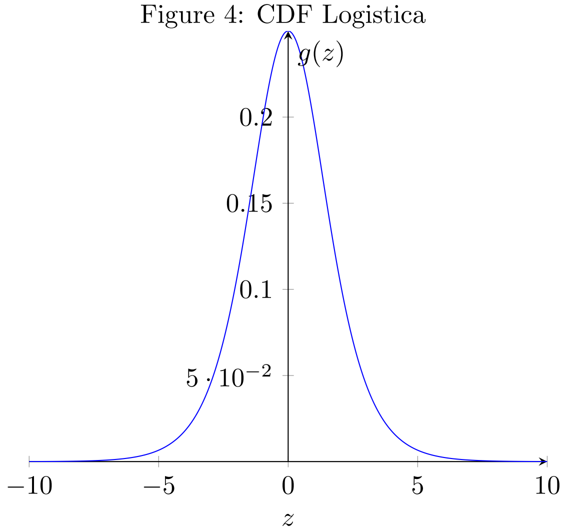

Figure 5: pdf Logistica

ID	Año	lw	esc
1	2018	1.80	12
1	2019	1.81	12
2	2018	1.91	15
2	2019	1.918	15

Table 1: Estructura tipo panel

5 Modelos de paneles de datos

Una estructura de datos tipo panel consiste es aquella en la que observamos a las mismas unidades de análisis durante un periodo de tiempo. Por ejemplo, observamos a los mismos individuos en los años 2019 y 2020. Así, observamos N individuos durante T periodos. Acá, nos concentraremos en paneles donde N es grande y T pequeño.

Una de las ventajas de los paneles de datos es que podemos controlar por las características no observables de los individuos y que son invariantes en el tiempo.

Un modelo simple de panel de datos es como sigue

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \alpha_1 w_{1i} + c_i + u_{it} \tag{35}$$

Acá x_{2it} es la variable x_2 para el individuo i en el momento t. La variable w_1 no tiende sub-índice temporal, luego quiere decir que esta no cambia en el tiempo de observación. c_i es un término que recoge las características individuales no observadas e invariantes en el tiempo, es lo que llamamos la heterogeneidad no observada. Finalmente, u_{it} es el error.

Por ejemplo, si estamos estudiando las ventas de una muestra de firmas y tenemos datos para 2 años, 2019 y 2020, entonces c_i podría contener aspectos como la calidad de la gerencia. Es algo que es importante para el desempeño de la firma, pero que no observamos y que muy probablemente no cambia en ese periodo de tiempo. En los estudios con individuos, c_i puede capturar características como rasgos de personalidad o habilidades cognitivas, por ejemplo.

Como c_i es no observado, entonces podemos escribir el modelo como

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \alpha_1 w_{1i} + v_{it} \tag{36}$$

Donde $v_{it} = c_i + u_{it}$

Ahora, suponga que estimamos el modelo por MCO y resulta que x_{2it} está correlacionada con c_i, en ese caso sabemos que la estimación de (36) no es consistente ¿Qué podemos hacer? Dado que el problema lo ocasiona c_i y como tenemos varios periodos de tiempo, entonces podemos realizar alguna transformación para eliminar c_i. Veamos

5.1 Estimador de primeras diferencias con T=2

Si tenemos 2 periodos de tiempo entonces para el individuo i podemos escribir

$$y_{i1} = \beta_1 + \beta_2 x_{2i1} + \alpha_1 w_{1i} + c_i + u_{i1}$$
$$y_{i2} = \beta_1 + \beta_2 x_{2i2} + \alpha_1 w_{1i} + c_i + u_{i2}$$

Si hacemos $\Delta y_{i2} = y_{i2} - y_{i1}$ obtenemos

$$\Delta y_{i2} = \beta_2(x_{2i2} - x_{2i1}) + (u_{i2} - u_{i1}) \tag{37}$$

Note que en (37) desapareció c_i y por lo tanto podríamos estimar β_2 por MCO de forma consistente. Esto siempre y cuando Δx_{i2} no esté correlacionado con $Deltau_{i2}$.

5.2 Estimador de efectos fijos

La transformación que hacemos acá consiste en restar de cada variable su media temporal para cada individuo, es lo que se conoce como la transformación *within*. Definimos

$$\bar{y_i} = \frac{1}{T}\sum_{t=1}^{T} y_{it}$$
$$\bar{x_{2i}} = \frac{1}{T}\sum_{t=1}^{T} x_{2it}$$

Para w_i es claro que $w_i = \bar{w_i}$, y para toda variable que no cambie en el tiempo. El modelo con la transformación queda como

$$y_{it} - \bar{y_i} = \beta_2(x_{2it} - \bar{x_{2i}}) + (u_{it} - \bar{u_i}) \tag{38}$$

Si estimamos la ecuación (38) por MCO los estimadores son consistentes si $(x_{2it} - \bar{x_{2i}})$ no está correlacionado con $(u_{it} - \bar{u_i})$

6 Series de tiempo

6.1 Introducción

Una serie de tiempo es una secuencia ordenada de valores de una variable que se observa a intervalos de tiempo igualmente espaciados. El propósito es desarrollar modelos simples capaces de pronosticar, interpretar y probar hipótesis referentes a datos económicos.

Por ejemplo, la figura 1 muestra el IPC mensual para Colombia para el periodo enero 2001-mayo 2020. En el panel superior está el nivel de la serie y en el panel inferior la primera diferencia de la serie en logaritmo, es decir la

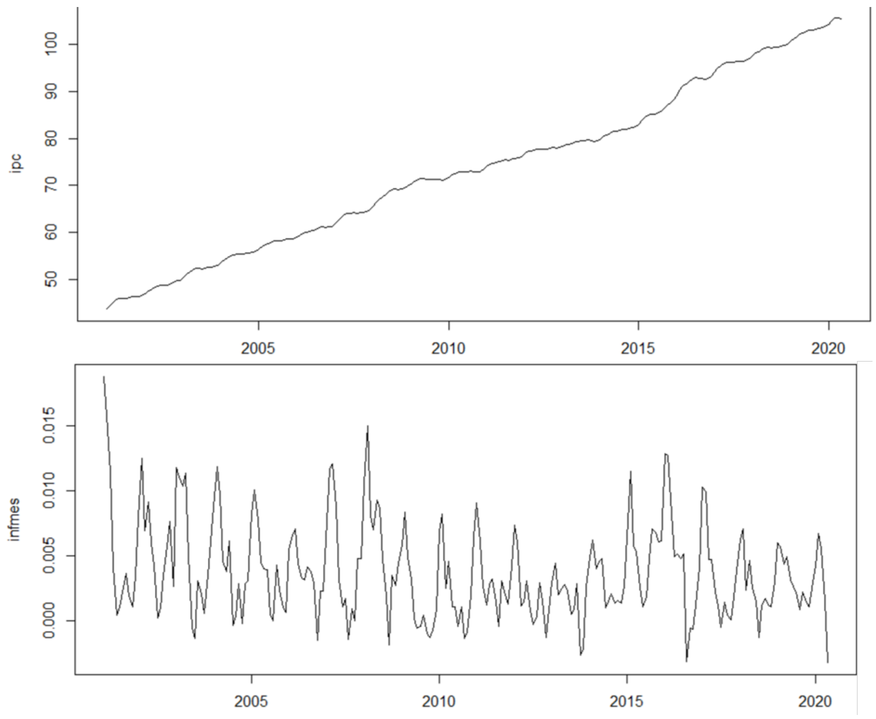

Figure 6: IPC e Inflación mensual

inflación mensual. Dada esta información, usted está interesado en pronosticar la inflación mensual para los siguientes seis meses. Si tomamos el IPC (en logaritmo), vemos que esta puede descomponerse en 3 elementos: tendencia, estacional, irregular

Note que la tendencia cambia la media de la serie, mientras que el componente estacional le da un patrón recurrente con picos a intervalos estables. Tanto la tendencia como el patrón estacional pueden tener elementos estocásticos. Note que el componente irregular, panel inferior, si bien no sigue un patrón, es predecible hasta cierto punto. Valores positivos tienden a estar precedidos por valores positivos, y negativos por negativos. Note también que el promedio de la parte irregular es cero.

La figura 3. muestra el comportamiento de la tasa de cambio peso/dólar. En el panel superior se encuentra la serie en niveles y el panel inferior la variación porcentual diaria. Note que, a diferencia de la serie de inflación, la tasa de cambio no muestra un patrón estacional evidente, ni tampoco es claro que tenga una tendencia ascendente o descente, ni que revierta a una media de largo plazo. La variación porcentual diraira de la tasa de cambio no tiene tendencia, pero si revierte a una media de cero a largo plazo. Note además que esta última presenta periodos más volátiles que otros, es decir que su varianza no es constante.

La tasa de cambio es quizás el precio relativo más importante para una

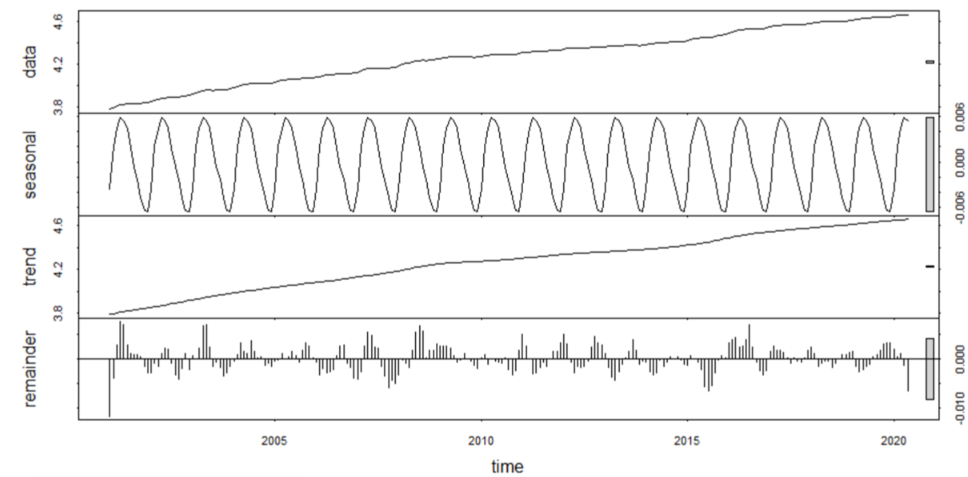

Figure 7: Descomposición IPC log

economía abierta, y por ello hay esfuerzos continuos en pronosticar su comportamiento. Uno de los aspectos que más intriga es la desconexión que parece haber entre los fundamentales económicos y el comportamiento de corto plazo de la tasa de cambio [5]. Es decir que estos son de limitada utilidad a la hora de pronosticarla, lo que sugeriría que la tasa de cambio estaría determinada por factores aleatorios, en otras palabras, la tasa de cambio sigue una caminata aleatoria. Más adelante examinamos esta hipótesis.

Considere ahora el comportamiento del PIB real de Colombia para el periodo 1950-2017, figura 4. [6] De la gráfica es evidente que el PIB tiene una tendencia creciente. La pregunta clave acá es si dicha tendencia es determinística o estocástica. Si la tendencia es determinística entonces cualquier choque que lleve al PIB por debajo de su tendencia de largo plazo no tendrá efectos permanentes. Así, por ejemplo esto implicaría que las crisis, como la de finales de los 90 o la de 2020, solo son de carácter transitorio e inevitablemente retornaremos a la senda de crecimiento que traíamos. Esto lo revisaremos formalmente más adelante.

6.1.1 Modelando relaciones dinámicas

Las series de tiempo son por naturaleza dinámicas. Los valores actuales de las variables están relacionados con los valores pasados de ella misma y de otras variables importantes. Podemos plantear los siguientes tipos de modelos

1. Modelo estático
$$y_t = \alpha + \beta x_t + e_t \tag{39}$$

[5] Rogoff, K., Meese, R., 1983, Empirical exchange rate models of the seventies: do they fit out of sample?, *Journal of International Economics*, vol.14

[6] Datos obtenidos de Penn World Table 9.1 https://www.rug.nl/ggdc/productivity/pwt/?lang=en

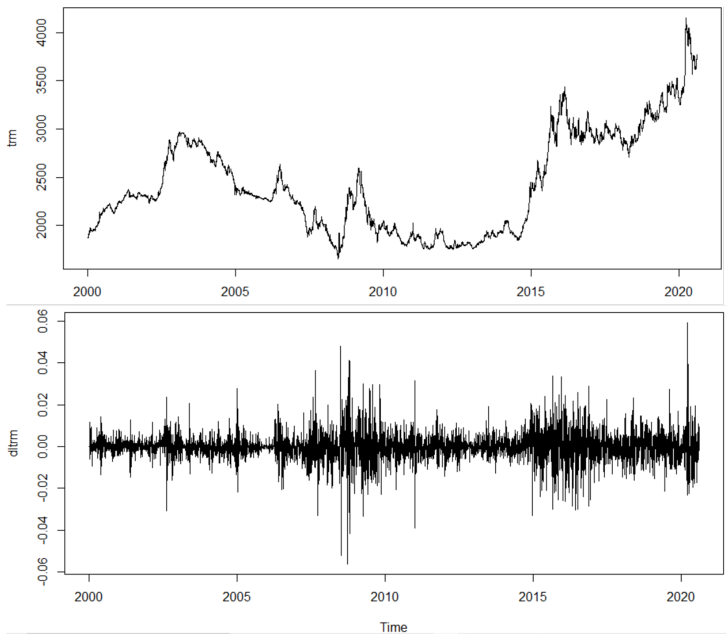

Figure 8: Tasa de cambio USDCOP nivel y variación diaria

2. Modelo de rezagos distribuidos finitos

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_q x_{t-k} + e_t \tag{40}$$

3. Modelo autoregresivo

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + e_t \tag{41}$$

4. Modelo autorregresivo de rezagos distribuidos

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + ... + \theta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + ... + \delta_q x_{t-q} + e_t \tag{42}$$

6.2 Conceptos fundamentales

6.2.1 Variable aleatoria

Uan variable aleatoria es un número $\mathbf{x}(\zeta)$ asignado a cada resultado ζ de un experimento[7].

- Ejemplo: tome un dado. Asigne el valor numérico que se observa en la cara superior del dado cada vez que lo lanza. Lánceló 10 veces, anote los valores y grafique

[7] Se sugiere la lectura de Papoulis, A., and Pillai, U., 2002, *Probability, Random Variables and Stochastic Processes*, 4 ed., McGrawHill

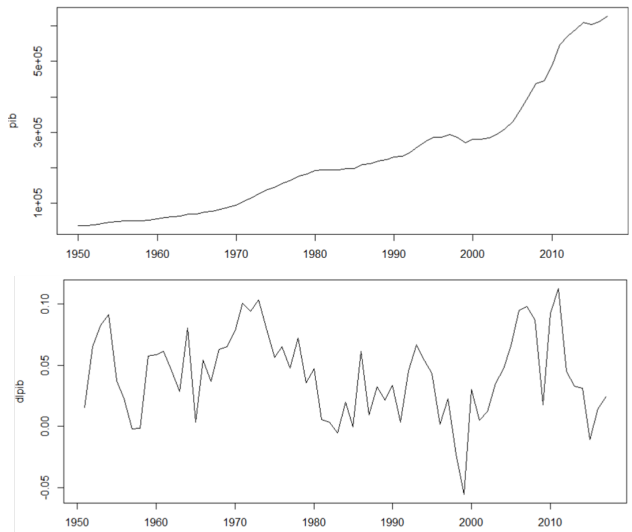

Figure 9: Colombia: PIB real (panel superior) y crecimiento (panel inferior)

- Ejemplo: tome el mismo dado. Asigne el valor de 1 si el la cara superior es un número par, y el valor de 0 si el número es impar. Láncelo 10 veces, anote los valores y grafique
- Ejemplo: Imagine un caminante que esta ubicado en el punto $(0, 0)$ de un plano. Sobre el eje x tiene los lanzamientos y sobre el eje y el punto donde se ubica el caminante en cada lanzamiento. El caminante sigue la siguiente regla. El siguiente paso que dará lo define lanzando una moneda. Si sale cara se mueve una unidad hacia arriba en y, y si sale sello se mueve una unidad hacia abajo en y. Haga 10 lanzamientos, dibuje la trayectoria del caminante

6.2.2 Autocorrelaciones

La correlación mide el grado de asociación lineal entre variables. El coeficiente de correlación para dos variables aleatorias, m y z es

$$\rho_{mz} = \frac{Cov(m, z)}{\sqrt{Var(m)Var(z)}} \tag{43}$$

Si aplicamos estos a la serie de tiempo, x_t podemos preguntarnos por el grado de asociación lineal entre las x que son observadas en diferentes momentos del tiempo, por ejemplo de x_t con x_{t-j}

$$\rho_j = \frac{Cov(x_t, x_{t-j})}{\sqrt{Var(x_t)Var(x_{t-j})}} \tag{44}$$

6.2.3 Ruido blanco

El bloque básico de los procesos que consideraremos es es la secuencia $\{e_t\}_{-\infty}^{\infty}$ cuyos elementos tienen media cero, varianza constante, y no están correlacionados en el tiempo

$$E(e_t) = 0 \tag{45}$$
$$E(e_t^2) = \sigma^2 \tag{46}$$
$$E(e_t e_\tau) = 0 \quad \text{para} \quad t \neq \tau \tag{47}$$

A partir de la definición de ruido blanco definimos series de tiempo autocorrelacionadas de la siguiente forma

$$Y_t = g(Y_{t-1}, Y_{t-1}, ..., e_t, e_{t-1}, ...) \tag{48}$$

Es decir que la observación en el momento t de la variable Y es una función del pasado de dicha variable, así como de choques aleatorios contemporáneos y pasados. Piense por ejemplo en la tasa de crecimiento del PIB trimestral. El crecimiento en un trimestre particular está relacionado con el crecimiento del trimestre anterior así como por choques exógenos que ocurren durante el trimestre, como el estallido de la guerra comercial EE.UU-China o un desplome del precio del petróleo.

6.2.4 Estacionariedad

Una variable es estacionaria si su media y varianza son constantes y las autocovarianzas solo dependen del intervalo de tiempo de separación. Decimos entonces que Y_t es estacionaria si

$$E(Y_t) = \mu \quad \text{para todo} \quad t \tag{49}$$
$$E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j \quad \text{para todo} \quad t \quad \text{y} \quad j \tag{50}$$

Donde γ_j corresponde a $Cov(Y_t, Y_t - j)$ para $j = 0, 1, 2, ...$