Introducción a series de tiempo

Gustavo A. García

ggarci24@eafit.edu.co

Econometría II

Programa de Economía

Universidad EAFIT

Link slides en formato html

Link slides en formato PDF

En este tema

- Introducción
- Procesos estocásticos estacionarios
- Procesos estocásticos no estacionarios
- Procesos estocásticos de raíz unitaria
- Procesos estocásticos integrados
- Regresión espuria
- Pruebas de estacionariedad
- Transformación de las series de tiempo no estacionarias
- En resumen
- Criterios de información para selección de modelos
- Ejercicio aplicado en R: estacionariedad del PIB

Lecturas

- Gujarati, D. y Porter, D. (2010). *Econometría*. 5a edición, Mc Graw Hill. Cap 21
- Enders, W. (2014). *Applied econometric time series*. 4th edition, Wiley. Cap 4, sección 4
- Pfaff, B. (2008). *Analysis integrated and cointegrated series with R*. 2th edition, Springer. Part II

Introducción

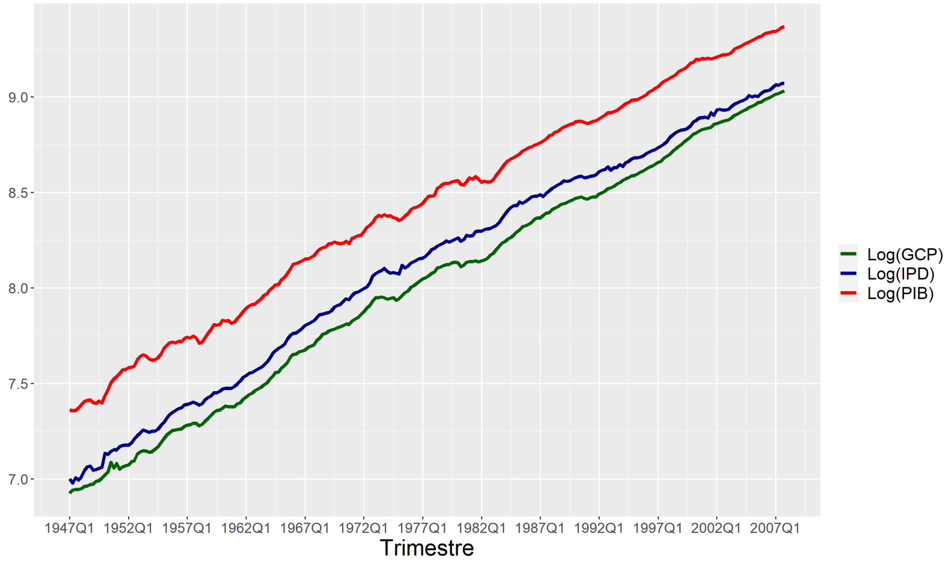
- Una característica obvia de los datos de series de tiempo que los distingue de aquellos de corte transversal es que tienen un orden temporal
- Algunos ejemplos de series temporales:
    - IPD = ingreso personal disponible real (miles de millones de dólares)
    - PIB = producto interno bruto (miles de millones de dólares)
    - GCP = gasto de consumo personal real (miles de millones de dólares)
    - UE = utilidades empresariales (miles de millones de dólares)
    - Dividendo = dividendos (miles de millones de dólares)
- Es práctica común graficar el logaritmo de una serie de tiempo para tener una idea de la tasa de crecimiento de dicha serie.
- Una gráfica de los datos es por lo general el primer paso en el análisis de series de tiempo. Se tienen cifras trimestrales desde 1947-1 a 2007-4

Introducción

```
library(gujarati); library(tidyverse); library(tseries); library(lmtest); library(forecast); library(zoo)

data('Table21_1')
data <- Table21_1 |>
  mutate(time     = as.yearqtr(paste0(Quarter," ",Year), format = "%q %Y"),
         year     = as.numeric(as.character(Year)), quarter = as.numeric(as.character(Quarter)),
         rpd      = as.numeric(as.character(RPD)), pib     = as.numeric(as.character(PIB)),
         dcp      = as.numeric(as.character(DCP)), lc      = as.numeric(as.character(LC))) |>
  select(time, year, quarter, rpd, pib, dcp, lc)

ggplot(data) +
  geom_line(aes(time, log(rpd), color = "Log(IPD)"), linewidth = 1.5) + geom_line(aes(time, log(pib), color = "Log(PIB)"), linewidth = 1.5)  + geom_l
  labs(x="Trimestre", y="") + theme(legend.text = element_text(size = 15), axis.title.x = element_text(size = 20), text = element_text(size=16)) + sca
```



Características de estas series:

- tienden hacia arriba
- presentan algunas fluctuaciones

Introducción

- Si deseáramos especular sobre la forma de estas curvas más allá del período muestral, por ejemplo, para todos los trimestres de 2008, es necesario saber el mecanismos estadístico, o estocástico, o el proceso de generación de datos (PGD) que dio origen a estas series

- Para saber cuál es ese mecanismos es necesario estudiar ciertos conceptos que se utilizan mucho en series de tiempo

Procesos estocásticos

- Un proceso estocástico o aleatorio es una colección de variables aleatorias ordenadas en el tiempo

- Por ejemplo, ¿en qué sentido podemos considerar al PIB un proceso estocástico? Estocástico significa aleatorio, por tanto el PIB es una variable aleatoria o proceso estocástico ya que puede tomar cualquier valor y hay incertidumbre sobre el valor que pueda tomar

Procesos estocásticos estacionarios

- Es un tipo de proceso estocástico

- Ha sido objeto de mucho estudio en las series de tiempo

- Un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos períodos depende sólo de la distancia o rezago entre estos dos períodos, y no del tiempo en el cual se calculó la covarianza

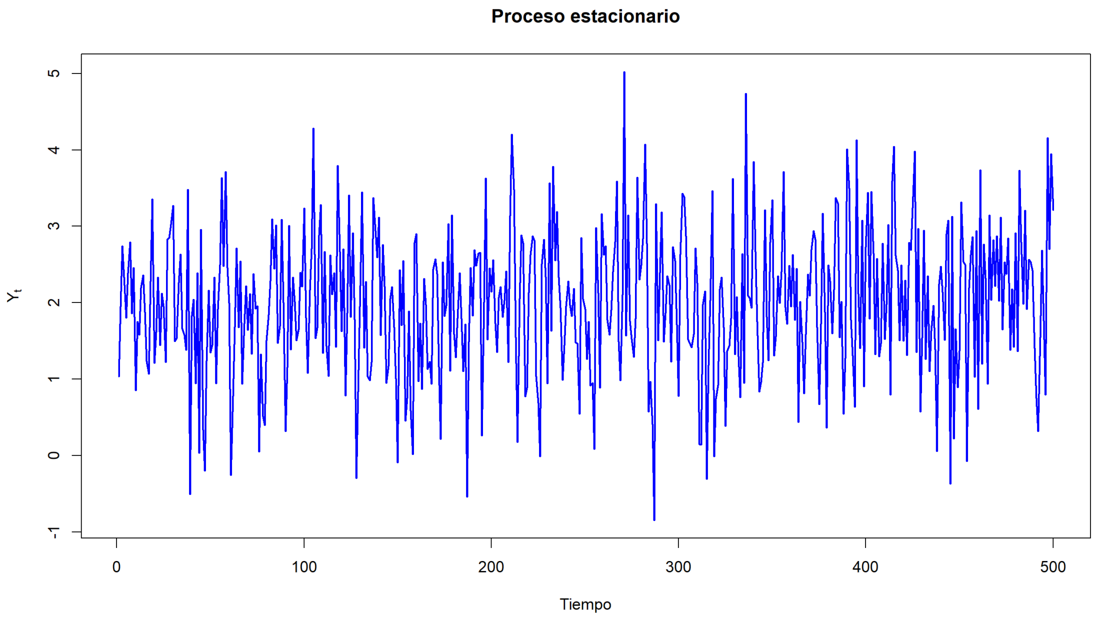
- Una serie de tiempo estacionaria es aquella cuyas propiedades no dependen del tiempo en el cual la serie es observada. Así, series de tiempo con tendencia o con estacionalidades, no son estacionarias $\implies$ la tendencia o estacionalidad afectará el valor de la serie de tiempo en diferentes momentos

- Para explicar la estacionariedad débil, sea Y_t una serie de tiempo con estas propiedades:
  - Media: $E(Y_t) = \mu$
  - Varianza: $Var(Y_t) = E(Y_t - E(Y_t))^2 = E(Y_t - \mu)^2 = \sigma_Y^2$
  - Covarianza: $\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)]$

Procesos estocásticos estacionarios

- Supongamos que el origen de Y se desplaza de Y_t a Y_{t+m} (por ejemplo, de 1947-1 a 1952-1 de los datos del PIB)
- Si esperamos que Y_t sea estacionaria, la media, la varianza y la covarianza de Y_{t+m} deben ser las mismas que la de Y_t
- Un proceso estacionario no se desvía demasiado de su valor medio debido a la varianza finita

```r
eps <- rnorm(500, mean = 0, sd = 1)
mu <- 2
Y_t <- mu + eps
ts.plot(Y_t, main = "Proceso estacionario", xlab = "Tiempo", ylab = expression(Y[t]), col = "blue", lwd = 2)
```



Procesos estocásticos estacionarios

¿Por qué las series de tiempo estacionarias son tan importantes?

- Porque si una serie de tiempo es no estacionaria, sólo podemos estudiar su comportamiento durante el período en consideración
- En consecuencia, no es posible generalizar para otros períodos
- Así, para propósitos de pronóstico, tales series de tiempo (no estacionarias) tienen poco valor práctico

Otro proceso estocástico (o de series de tiempo) es el proceso puramente aleatorio o de ruido blanco: se dice que un proceso es puramente aleatorio si tiene una media igual a cero, una varianza constante σ^2 y no está serialmente correlacionado

Una serie de tiempo ruido blanco es estacionaria, no importa cuando la observes, deberá verse igual en cuaquier momento

Procesos estocásticos no estacionarios

- A menudo nos encontramos con series de tiempo no estacionarias y el ejemplo más clásico es el modelo de caminata aleatoria
- Los precios de valores, como las acciones o las tasas de cambio, siguen una caminata aleatoria, es decir, son no estacionarios
- Hay dos tipos de caminatas aleatorias
  - caminata aleatoria sin deriva o sin desvío (es decir, sin término constante o de intercepto)
  - caminata aleatoria con deriva o con desvío (es decir, hay término constante o de intercepto)

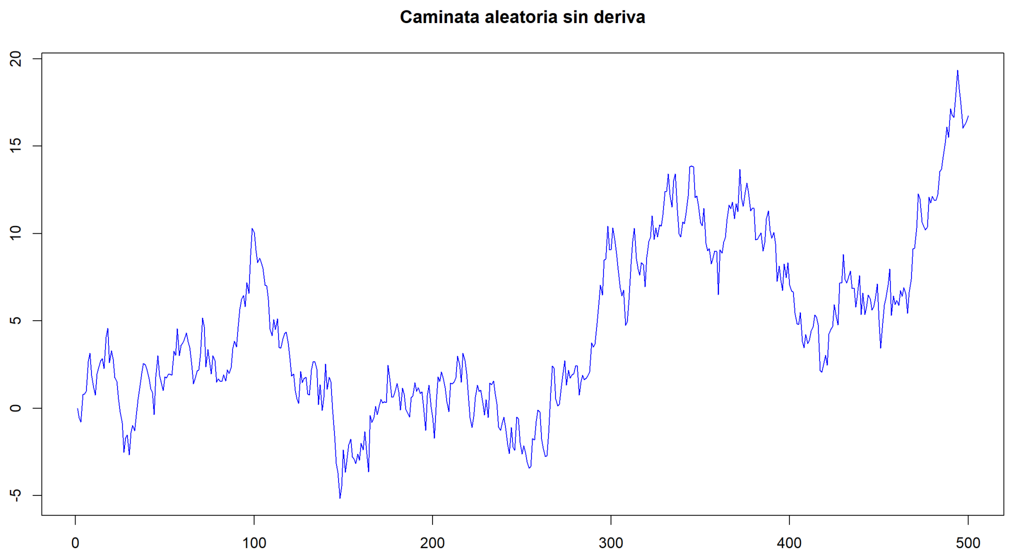
Procesos estocásticos no estacionarios

Caminata aleatoria sin deriva (*Random walk without drift*)
La serie Y_t es una caminata aleatoria sin deriva si

$$Y_t = Y_{t-1} + u_t$$

$$E(Y_t) = Y_0, Var(Y_t) = t\sigma^2$$

```
set.seed(123)
n <- 500
rw <- numeric(n)
for (i in 2:n) {
  rw[i] <- rw[i-1] + rnorm(1)
}
ts.plot(rw,main="Caminata aleatoria sin deriva", col="blue",
        xlab = "", ylab = "")
```

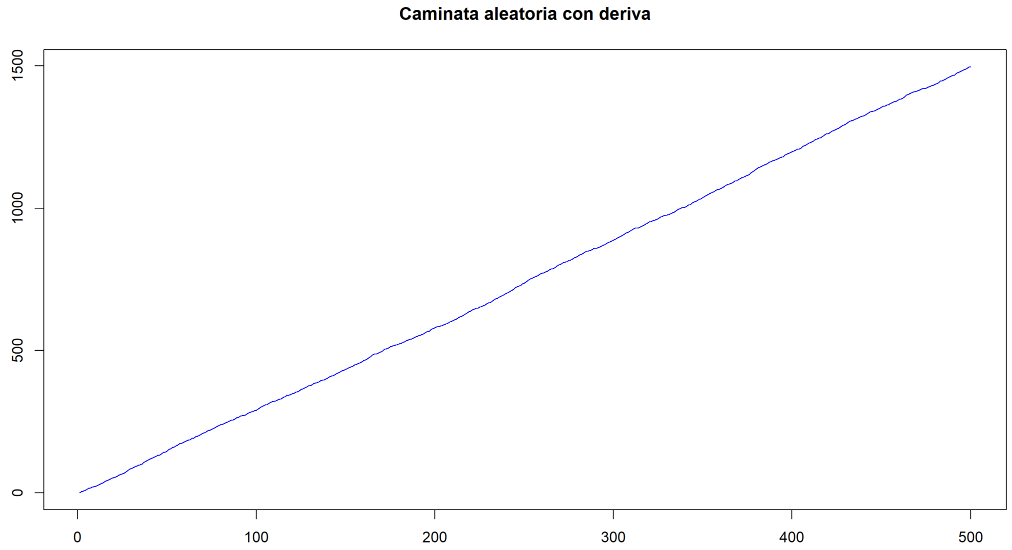


Caminata aleatoria con deriva (*Random walk with drift*)
La serie Y_t es una caminata aleatoria con deriva si

$$Y_t = \delta + Y_{t-1} + u_t$$

$$E(Y_t) = Y_0 + t\delta, Var(Y_t) = t\sigma^2$$

```
rw_drift <- numeric(n)
for (i in 2:n) {
  rw_drift[i] <- 3 + rw_drift[i-1] + rnorm(1)
}
ts.plot(rw_drift, main="Caminata aleatoria con deriva", col = "blue",
        xlab = "", ylab = "")
```



Procesos estocásticos de raíz unitaria

El modelo de caminata aleatoria es un ejemplo de lo que se conoce en la bibliografía como proceso de raíz unitaria. Como este término es ya muy común en las referencias de series de tiempo, a continuación explicaremos lo que es un proceso de raíz unitaria

Escribamos el modelo de caminata aleatoria como:

$$Y_t = \rho Y_{t-1} + u_t$$

- Este modelo es lo que se llama un modelo autorregresivo de primer orden (AR(1))
- Si $\rho = 1$ se convierte en un modelo de caminata aleatoria sin deriva
- Si $\rho = 1$ tenemos lo que se conoce como problema de raíz unitaria, es decir enfrentamos una situación de no estacionariedad
- Los términos no estacionariedad, caminata aleatoria, raiz unitaria y tendencia estocástica se consideran sinónimos
- Si $|\rho| < 1$ se puede demostrar que la serie de tiempo Y_t es estacionaria
- Existen diferentes pruebas para corroborar la existencia de no estacionariedad

Procesos estocásticos integrados

- El modelo de caminata aleatoria sin deriva es no estacionario, pero su serie de primeras diferencias es estacionaria:

$$Y_t = Y_{t-1} + u_t$$

$$\Delta Y_t = Y_t - Y_{t-1} = u_t$$

- El modelo de caminata aleatoria sin deriva se llama proceso integrado de orden 1 y se denota como $I(1)$
- Si una serie de tiempo tiene que diferenciarse dos veces (es decir, se toman primeras diferencias de la serie de primeras diferencias) para hacerla estacionaria, esa serie de tiempo se denomina integrado de orden 2
- En general, si una serie de tiempo (no estacionaria) debe diferenciarse d veces para hacerla estacionaria, decimos que la serie es integrada de orden d
- Y_t es integrada de orden d: $Y_t \sim I(d)$. Si es estacionaria entonces $Y_t \sim I(0)$
- La mayoría de las series de tiempo económicas son $I(1)$; es decir, por lo general se convierten en estacionarias sólo después de tomar sus primeras diferencias

Regresión espuria

Para ver por qué las series de tiempo estacionarias son tan importantes, considere los dos modelos de caminata aleatoria siguientes:

$$Y_t = Y_{t-1} + u_t$$

$$X_t = X_{t-1} + v_t$$

Se asume que u_t y v_t no están serial ni mutuamente correlacionadas. Ya sabemos que ambas series de tiempo son no estacionarias, es decir, son $I(1)$. Supongamos que hacemos la regresión de Y_t sobre X_t. Como estas series son procesos no correlacionados $I(1)$, el R^2 de la regresión debería tender a cero, es decir, no debería haber ninguna relación entre las dos variables. Miremos la regresión

```
Y_t <- numeric(n)
X_t <- numeric(n)
for (i in 2:n) {
  Y_t[i] <- Y_t[i-1] + rnorm(1)
  X_t[i] <- X_t[i-1] + rnorm(1)
}
modelo1 <- lm(Y_t ~ X_t)
summary(modelo1)
```

```
Call:
lm(formula = Y_t ~ X_t)

Residuals:
    Min      1Q  Median      3Q     Max
-7.6186 -3.1349 -0.8393  3.2261 10.4912

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.4209     0.2928   11.68   <2e-16 ***
X_t          -0.2213     0.0133  -16.64   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.099 on 498 degrees of freedom
Multiple R-squared:  0.3574,    Adjusted R-squared:  0.3561
F-statistic:   277 on 1 and 498 DF,  p-value: < 2.2e-16
```

Regresión espuria

- El coeficiente de X_t es muy significativo estadísticamente, y aunque el R^2 es bajo, es distinto de cero
- Con estos resultados se podría estar tentado a decir que hay una relación entre Y_t y X_t. Esto es lo que llamamos regresión espuria o regresión sin sentido
- La correlación espuria puede persistir en las series no estacionarias aunque la muestra sea muy grande
- El DW es muy bajo lo que indica una autocorrelación muy fuerte de primer orden. De acuerdo con Granger y Newbold (1974), $R^2 > DW$ es una buena regla práctica para sospechar que la regresión estimada es espuria[1]

```
dwtest(modelo1)
```

```
	Durbin-Watson test

data:  modelo1
DW = 0.058538, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

- Al estimar la regresión en primeras diferencias $(\Delta Y_t \text{ contra } \Delta X_t)$ se nota que el modelo estimado carece de sentido

```
modelo2 <- lm(diff(Y_t) ~ diff(X_t))
dwtest(modelo2)
```

```
	Durbin-Watson test

data:  modelo2
DW = 2.1533, p-value = 0.9569
alternative hypothesis: true autocorrelation is greater than 0
```

[1] Granger, CWJ y Newbold, P. (1974). "Spurious regressions in econometrics". *Journal of Econometrics*, 2: 111-120.


```
summary(modelo2)
```

Pruebas de estacionariedad

El análisis gráfico

- Antes de efectuar una prueba formal, siempre es aconsejable graficar la series de tiempo en estudio
- Los gráficos proporcionan una pista inicial respeto de la posible naturaleza de las series de tiempo
- Por ejemplo, la serie de tiempo del PIB crece a lo largo de todo el período, es decir que muestra una tendencia ascendente y por lo tanto no es estacionaria
- Esta intuición es el comienzo de una prueba más formal de estacionariedad

Pruebas formales

- Correlograma de la serie
- Pruebas de raices unitarias
  - Dickey-Fuller (aumentado) test $\Longrightarrow H_0$: existe una raíz unitaria, la serie no es estacionaria
  - Phillips-Perron test $\Longrightarrow H_0$: existe una raíz unitaria, la serie no es estacionaria
  - KPSS test $\Longrightarrow H_0$: no existe una raíz unitaria, la serie es estacionaria
  - Elliott-Rothenberg-Stock test
  - Schmidt-Phillips test
  - Zivot-Andrews test

Pruebas de estacionariedad

Qué es raíz unitaria?

- Una raíz unitaria (también llamado proceso de raíz unitaria) es una tendencia estocástica en una seri de tiempo, algunas veces llamado una caminata aleatoria
- Si una seriede tiempo tiene una raiz unitaria, esta muestra un patrón sistemático que es impredecible
- Un modelo simple de series de tiempo es el AR(1):

  $$Y_t = \rho Y_{t-1} + u_t$$

  Si $\rho < 1$ el modelo es estacionario, si $\rho = 1$ el modelo no es estacionario
- Los tests de raices unitarias son pruebas de estacionariedad sobre las series de tiempo. La forma de la estacionariedad es si un cambio en el tiempo no provoca un cambio en la forma de la distribución

Pruebas de estacionariedad

El test de Dickey-Fuller (DF)

- El punto de partida es el proceso (estocástico) de raíz unitaria que vimos anteriormente. Se inicia con:

  $$Y_t = \rho Y_{t-1} + u_t$$

  donde $-1 \leq \rho \leq 1$ y u_t es el término de error ruido blanco

- Sabemos que si $\rho = 1$, es decir, en el caso de la raíz unitaria, la anterior ecuación se convierte en un modelo de caminata aleatoria $\Longrightarrow$ este es la idea general de la prueba de raíz unitaria

- ¿Por qué no simplemente hacer la regresión de Y_t sobre su valor rezagado Y_{t-1} y se averigua si la ρ estimada es estadísticamente igual a 1? $\Longrightarrow$ no podemos estimar por MCO y probar $H_0 : \rho = 1$ por medio de la prueba t acostumbrada, porque esa prueba tiene un sesgo muy marcado en el caso de una raíz unitaria

- Se debe entonces manipular la ecuación, restando Y_{t-1} a ambos lados:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ &= (\rho - 1)Y_{t-1} + u_t \\ \Delta Y_t &= \delta Y_{t-1} + u_t \end{aligned}$$

donde $\delta = (\rho - 1)$ y Δ es el operador de primeras diferencias

Pruebas de estacionariedad

El test de Dickey-Fuller (DF)

- En la práctica entonces se estima por MCO y se prueba $H_0 : \delta = 0$ y la $H_a : \delta < 0$. Si $\delta = 0$ entonces $\rho = 1$, es decir, tenemos una raiz unitaria, lo que significa que la serie de tiempo es no estacionaria
- La interrogante que queda es saber con qué estadístico se hace la prueba ya que no es posible utilizar la prueba t ya que no sigue una distribución normal asintótica
- Dickey y Fuller probaron que según $H_0 : \delta = 0$ el valor estimado t del coeficiente de Y_{t-1} sigue el estadístico $\tau \Longrightarrow$ prueba Dickey-Fuller (DF)
- El procedimiento para aplicar la prueba DF depende del modelo analizado. La prueba se estima en tres diferentes formas dependiendo del proceso asumido:
  - Y_t es una caminata aleatoria $\Longrightarrow \Delta Y_t = \delta Y_{t-1} + u_t$
  - Y_t es una caminata aleatoria con deriva: $\Longrightarrow \Delta Y_t = \beta_1 + \delta Y_{t-1} + u_t$
  - Y_t es una caminata aleatoria con deriva alrededor de una tendencia determinista: $\Longrightarrow \Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t$ donde t es una variable de tiempo o tendencia
- $H_0 : \delta = 0$ (existe una raíz unitaria, la serie no es estacionaria)
  $H_a : \delta < 0$ (la serie es estacionaria)
  Rechazo Ho si $|t| > |\tau|$ o si p-value $< \epsilon$

Pruebas de estacionariedad

El test de Dickey-Fuller Aumentado (DFA)

- En el DFA se asume que el término de error u_t se encuentra correlacionado
- Esta prueba implica aumentar las tres ecuaciones anteriores mediante la adición de los valores rezagados de la variable dependiente ΔY_t. Suponiendo que el proceso es una caminata aleatoria con deriva alrededor de una tendencia determinista, aplicar la prueba DFA implica estimar el siguiente modelo:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta Y_{t-i} + e_t$$

- El número de términos de diferencia rezagados que debemos incluir se determina de manera empírica utilizando los criterios de información de Akaike, Schwarz y otros

Queda de tarea estudiar los otros tests de raices unitarias

Transformación de las series de tiempo no estacionarias

- Ahora que conocemos el problema asociado a las series de tiempo no estacionarias, surge la pregunta de qué hay que hacer
- Para evitar el problema de regresión espuria que pudiese surgir al hacer la regresión de una serie no estacionaria contra otras series no estacionarias, se debe transformar las series de tiempo no estacionarias a estacionarias
- El método de transformación depende de las series sean procesos estacionarios en diferencias (PED) o procesos estacionarios con tendencia (PET)

Transformación de las series de tiempo no estacionarias

Procesos estacionarios en diferencias

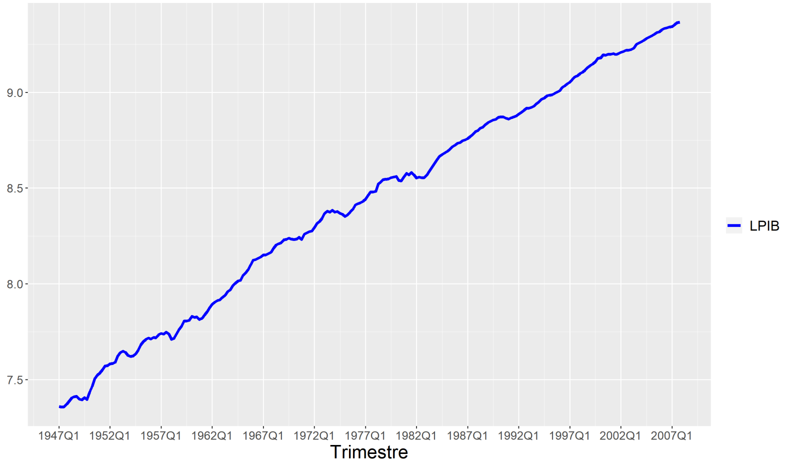
- Si una serie de tiempo tiene una raíz unitaria, las primera diferencia de tal serie es estacionaria
- La solución entonces es tomar las primeras diferencias de la serie de tiempo
- Sea la serie Y_t no estacionaria, entonces las primeras diferencias serán:

$$\Delta Y_t = Y_t - Y_{t-1}$$

$$D_t = \Delta Y_t$$

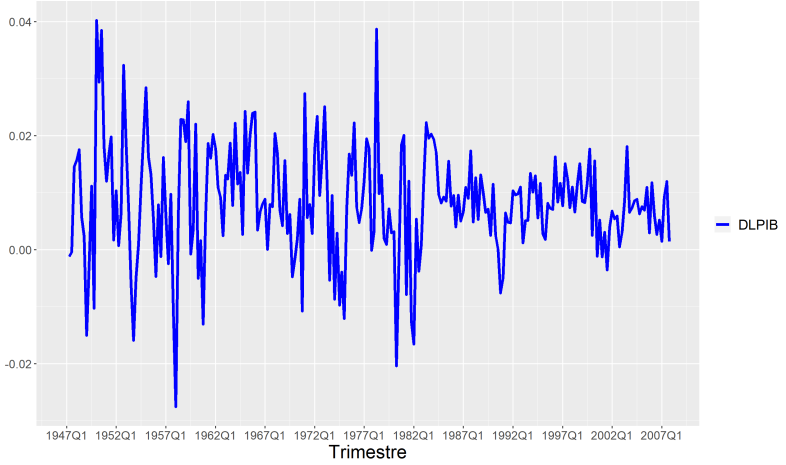
- Veamos el ejemplo de la serie $Log(PIB)$, la cual es no estacionaria:

```
ggplot(data) +
  geom_line(aes(time, log(pib), color = "LPIB"), linewidth = 1.5) +
  scale_color_manual(name = "", values = c("LPIB" = "blue")) + labs(x="Tri
  theme(legend.text = element_text(size = 15), axis.title.x = element_text
```




```
data <- data |> mutate(DLpib = c(NA,diff(log(pib))))

ggplot(data) +
  geom_line(aes(time, DLpib, color = "DLPIB"), linewidth = 1.5) +
  scale_color_manual(name = "", values = c("DLPIB" = "blue")) + labs(x="Tr
  theme(legend.text = element_text(size = 15), axis.title.x = element_text
```



Transformación de las series de tiempo no estacionarias

Procesos estacionarios en tendencia

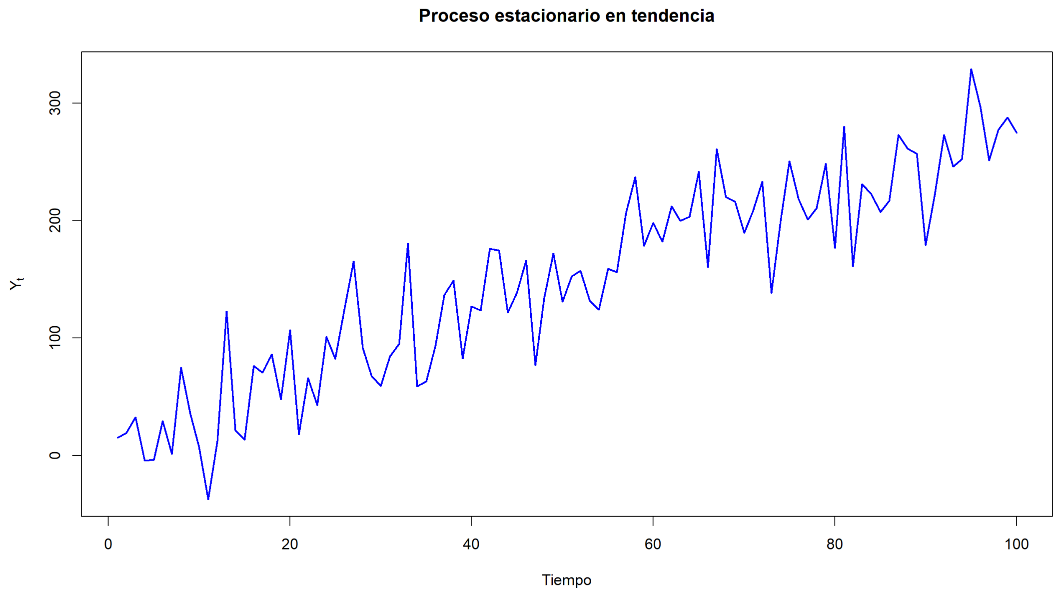
Este tipo de proceso tienen la forma:

$$Y_t = \beta_1 + \beta_2 t + u_t$$

donde t es una tendencia. Aunque la media de Y_t es $\beta_1 + \beta_2 t$ —no constante—, su varianza (= σ^2) sí lo es. Si restamos la media de Y_t de Y_t, la serie resultante será estacionaria; de ahí el nombre de estacionario en tendencia

Simulando un proceso estacionario en tendencia $Y_t = 2 + 3t + u_t$

```
u <- rnorm(100,0,40)
t <- (1:100)
Y_t <- 2 + 3*t + u
ts.plot(Y_t, main = "Proceso estacionario en tendencia", xlab = "Tiempo", ylab = expression(Y[t]), col = "blue", lwd = 2)
```



Transformación de las series de tiempo no estacionarias

Procesos estacionarios en tendencia

La manera más sencilla de convertir este tipo de series en estacionaria es hacer la regresión de ella sobre el tiempo y los residuos de tal regresión serán estacionarios. En otras palabras, realizamos la siguiente regresión:

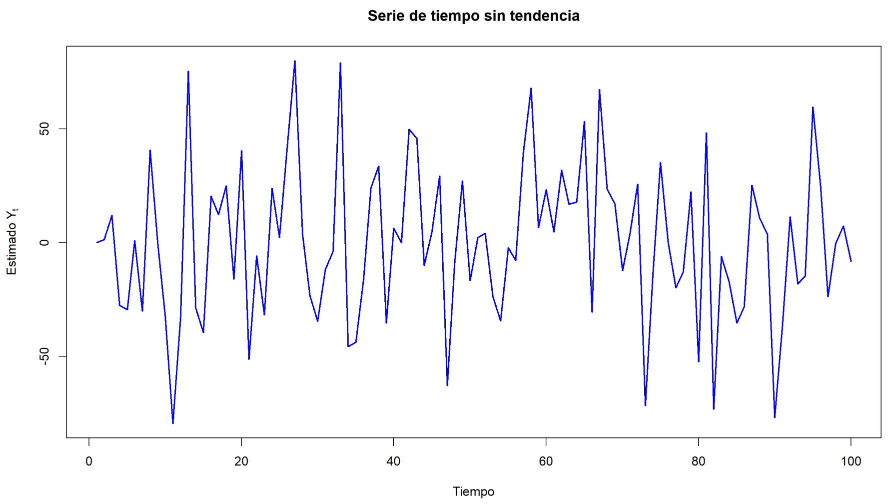
$$Y_t = \beta_1 + \beta_2 t + u_t$$

donde Y_t es la serie de tiempo estudiada y t es la variable de tendencia medida de manera cronológica. Ahora bien

$$\widehat{u}_t = (Y_t - \widehat{\beta}_1 - \widehat{\beta}_2 t)$$

será estacionaria. A $\widehat{u}_t$ se le conoce como serie de tiempo sin tendencia

```
mu_t <- lm(Y_t ~ t)
Ye_t <- Y_t - mu_t[['fitted.values']]
ts.plot(Ye_t, main = "Serie de tiempo sin tendencia", xlab = "Tiempo", ylab = expression('Estimado Y'[t]), col = "blue", lwd = 2)
```



Es importante notar que tal vez la tendencia sea no lineal. Por ejemplo, puede ser $Yt = \beta_1 + \beta_2 t + \beta_3 t^2 + u_t$ que es una serie con tendencia cuadrática. De ser así, los residuos serán ahora una serie (cuadrática) de tiempo sin tendencia

En resumen

- La econometría de series temporales inicia con el análisis de estacionariedad de la serie
- Si es estacionaria se dice que es integrada de orden cero y se denota $Y_t \sim I(0)$
- Si la serie no es estacionaria, se diferencia la serie para ver si se hace estacionaria $\Delta Y_t = Y_t - Y_{t-1}$
- Sea d el orden de integración de la serie el cual corresponde al número de veces que hay que diferenciar una serie para que se haga estacionaria. Cualquier serie es $I(d)$
- Los econométras de esta escuela señala que la inmensa mayoría de series macroeconomicas se pueden representar con un proceso $I(1)$
- Una vez se dispone una serie estacionaria se modela de acuerdo a procesos ARMA(p,q) surgiendo la metodología

$$\underbrace{AR}_{p}\ \underbrace{I}_{I}\ \underbrace{MA}_{q}$$

- Por ejemplo, si una serie sigue un proceso ARIMA(1,1,0) quiere decir que se diferenció una vez y se modelo luego con AR(1) y no se utilizó el MA

Criterios de información para selección de modelos

Criterio de información de Akaike (AIC) y Schwarz (BIC)

$$AIC = -2lnL + 2k$$

$$BIC = -2lnL + klnn$$

Donde L es el valor de la función de máximo verosimilitud, n es el número de observaciones y k es el número de parámetros. El más bajo AIC o BIC indica un mejor ajuste (un modelo más parsimonioso)

Ejercicio aplicado en R: estacionariedad del PIB

Se trabajan con variables macroeconómicas de los Estados Unidos, trimestrales entre 1947 y 2007 (precios constantes de 2000):

- IPD = ingreso personal disponible real (miles de millones de dólares)
- PIB = producto interno bruto (miles de millones de dólares)
- GCP = gasto de consumo personal real (miles de millones de dólares)
- UE = utilidades empresariales (miles de millones de dólares)
- Dividendo = dividendos (miles de millones de dólares)

En esta aplicación estudiamos la serie del PIB y se utilizará el siguiente código:

- Código en R