

UNIVERSIDAD ADOLFO IBÁÑEZ, ECO -TS101

Lecture Note: Econometría de Series de Tiempo

Marcelo Villena Chamorro PhD.

2019 Winter

Contents

1 Tópico I.- Introducción a la Econometría de Series de Tiempo	4
1.1 La naturaleza de los datos de serie de tiempo	4
1.1.1 Ejemplo 1: Cambio climático	5
1.1.2 Ejemplo 2: Series temporales financieras	6
1.2 Modelamiento estadístico de las series de tiempo	11
1.2.1 Ruido blanco - White Noise	11
1.2.2 Caminata Aleatoria - Random Walk	12
1.2.3 Promedios móviles - Moving Averages	15
1.2.4 Autorregresiones - Autoregressions	16
1.3 Descomposición de las series de tiempo	18
1.3.1 Descomposición Tendencia - Filtro Hodrick -Prescott	20
1.3.2 Descomposición componente estacional - Transformaciones de Diferencias	20
1.3.3 Descomposición componente estacional - Variables Dummy	21
1.4 Medidas de dependencia	22
1.5 Estacionaridad	24
1.6 Modelos de regresión lineal múltiple en series de tiempo	25
1.6.1 Propiedades de los estimadores de MC	26
1.6.2 Teorema Gauss-Markov	27
1.6.3 Modelos de regresión lineal múltiple en series de tiempo	27
1.6.4 Evaluación Estadística de Regresiones Estimadas	28
1.6.5 Regresiones sin sentido y regresiones espúrias.	28
1.6.6 Ejemplo 3: Regresiones	28
2 Tópico II.- Univariate Time Series Models	33
2.1 Ejemplo de repaso clase anterior-Detrending global temperature	33
2.2 Sobre la descomposición de una serie	39
2.3 Procesos no estacionarios, integrados y el test de raíz unitaria	40
2.3.1 Procesos Integrados	40
2.3.2 Consecuencias de los Procesos Integrados (Raíz Unitaria)	40
2.3.3 Test de Raíz Unitaria	40
2.4 Modelos ARIMA: modelando el corto plazo	43
2.4.1 Sobre el operador de retroceso - backshift operator	43
2.4.2 Ejemplo Proceso Autoregresivo de Orden 1: AR(1)	44
2.4.3 Simulación modelo AR(1)	45
2.4.4 Identificación modelo AR(1)	46
2.4.5 Media Móvil - MA (q)	49

2.4.6	Identificación modelo ARMA	53
2.4.7	Invertibilidad	54
2.4.8	Estacionaridad e Invertibilidad	54
2.4.9	Modelos ARMA (p,q)	55
2.4.10	Modelos ARIMA (p, i, q)	55
2.4.11	Identificación modelo ARMA	55
2.4.12	Modelos SARIMA (p,q)	56
2.4.13	Sobre la Selección de Modelos	57
2.4.14	Ejemplo IPC en Chile	59
2.4.15	Descomposición de la serie	62
2.5	Metodología de Estimación de un Modelo ARIMA	70
2.5.1	Etapas de Estimación de un Modelo ARIMA	70
2.6	Resumen de los pasos de Box-Jenkins	71
2.7	Tarea 2	72
3	Tópico III.- Vector Autoregressive Models	73
3.1	Vector Autoregressive Models	73
3.1.1	Introducción	73
3.1.2	Notación y Conceptos	74
3.1.3	VAR versus Modelos Estructurales	74
3.1.4	La elección de la longitud del rezago óptimo para un VAR	75
3.1.5	Forma Primitiva versus Forma Estándar de VARs	88
3.1.6	Estabilidad de los procesos VAR	89
3.1.7	Las respuestas de impulso y descomposiciones de la varianza	90
3.1.8	La función de respuesta al impulso y descomposiciones de la varianza: El orden de las variables	91
3.1.9	Significancia de un Block y Test de Causalidad	93
3.1.10	Resumen	95
3.2	Tarea 3	96

1 Tópico I.- Introducción a la Econometría de Series de Tiempo

1.1 La naturaleza de los datos de serie de tiempo

El objetivo principal del análisis de series temporales es desarrollar modelos matemáticos que proporcionen descripciones plausibles para los datos de la muestra.

Existen dos enfoques metodológicos básicos para modelar series de tiempo:

- (i) El enfoque del dominio del tiempo (Time domain approach)
- (ii) El enfoque de dominio de frecuencia (Frequency domain approach)

El **enfoque del dominio del tiempo** generalmente está motivado por la presunción de que la correlación entre los puntos adyacentes en el tiempo de los datos de una serie, se explica mejor en términos de una dependencia del valor actual en valores pasados. De esta forma, éste enfoque se centra en modelar algún valor futuro de una serie temporal como una función paramétrica de los valores actuales y pasados.

En este escenario, por ejemplo, utilizamos modelos de regresiones lineales sobre el valor actual de una serie temporal, utilizando como variables de lado derecho a sus propios valores pasados y los valores pasados de otras series. El seminal trabajo de Box y Jenkins y sus modelos autoregresivos (ARIMA), se encuentran en esta línea, ver [Box and Jenkins \(1970\)](#).

Por otro lado, el **enfoque de dominio de frecuencia** asume que las características principales de interés en el análisis de las series de tiempo se relacionan con variaciones sinusoidales periódicas o sistemáticas que se encuentran naturalmente en la mayoría de los datos.

Estas variaciones periódicas a menudo son causadas por fenómenos biológicos, físicos o ambientales de interés. El estudio de la periodicidad se extiende a la economía y las ciencias sociales, donde uno puede estar interesado en las periodicidades anuales en series como el desempleo mensual o las tasas mensuales de natalidad.

En el análisis espectral, la partición de los diversos tipos de variación periódica en una serie temporal se lleva a cabo evaluando separadamente la varianza asociada con cada tipo de periodicidad.

1.1.1 Ejemplo 1: Cambio climático

Nuestro primer ejemplo de serie de tiempo es la temperatura de la tierra. Observamos una aparente tendencia ascendente en la serie durante la última parte del siglo XX, la que se ha utilizado como argumento para la hipótesis del calentamiento global. Nótese también la tendencia ascendente y bastante pronunciada alrededor de 1970. La cuestión de interés para los defensores del calentamiento global y los oponentes es si la tendencia general es natural, o si por el contrario es causada por el ser humano.

Ejemplo 1: Cambio climático

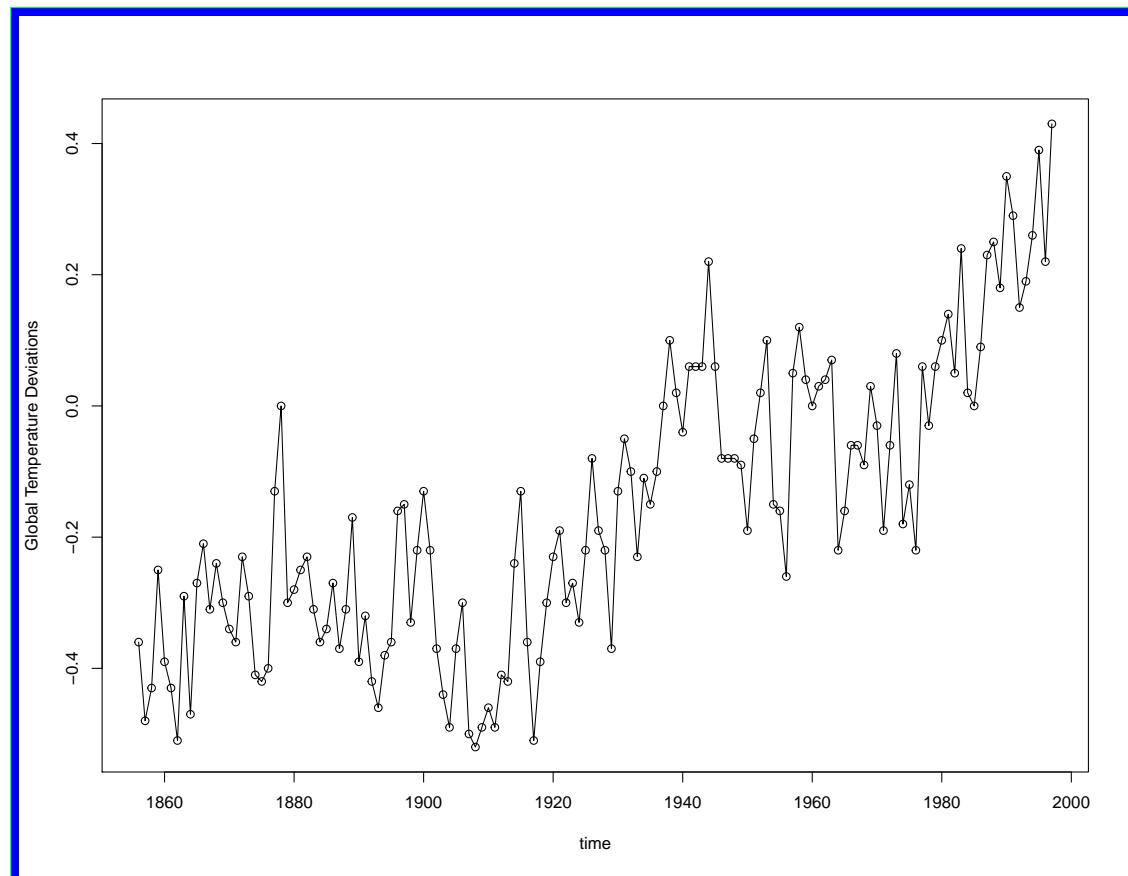


Figure 1.1: Cambio climático: mediciones de temperatura desde 1860

'Código R para ejemplo del cambio climático.'

```
# Set the working directory
setwd("/Users/Desktop/Econometrics/Clase_1._TSE")
# Limpiar_variables
rm(list=ls())

install.packages("tidyverse")
install.packages("dplyr")
install.packages("tseries")

mydata<-read.csv ("gtemp.csv")
plot(mydata, type="o", ylab="Global_Temperature_Deviations")
```

1.1.2 Ejemplo 2: Series temporales financieras

En finanzas siempre es preferible trabajar con el retorno (returns) de los activos, en vez de usar directamente el precio de los activos. Existen dos maneras de convertir el precio en retornos:

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} * 100$$

$$R_t = \ln\left(\frac{p_t}{p_{t-1}}\right) * 100$$

where, R_t denota el retorno al tiempo t , p_t denota el precio del activo al tiempo t , y \ln denota el logaritmo natural. En esta formulación ignoramos los dividendos, o asumimos que las series de precios ya han sido ajustadas por ellos.

Los log-returns tienen la deseable propiedad de poder ser interpretados como rendimientos continuamente compuestos. Además, ellos pueden ser simplemente sumados, de forma de obtener retornos con respecto a períodos más largos:

$$\begin{aligned} r_1 &= \ln\frac{p_1}{p_0} = \ln p_1 - \ln p_0 \\ r_2 &= \ln\frac{p_2}{p_1} = \ln p_2 - \ln p_1 \\ r_3 &= \ln\frac{p_3}{p_2} = \ln p_3 - \ln p_2 \\ r_4 &= \ln\frac{p_4}{p_3} = \ln p_4 - \ln p_3 \\ r_5 &= \ln\frac{p_5}{p_4} = \ln p_5 - \ln p_4 \end{aligned}$$

$$r_1 + r_2 + r_3 + r_4 + r_5 = \ln p_5 - \ln p_0 = \ln\frac{p_5}{p_0}$$

Ejemplo 2: Series temporales financieras

Como segundo ejemplo calcularemos los retornos de la Bolsa de Nueva York, índice "S&P 500", extrayendo data diaria desde el año 2000, del sitio: <https://finance.yahoo.com/>

'Código R precio de la Bolsa de Nueva York.'

```
mydata2<-read.csv("sp.csv")
precio<-mydata2$Adj.Close
plot.ts(precio, type="o", ylab="Precio_Bolsa_de_Nueva_York")
lnprecio <- log10(precio)
Dlnprecio <- diff(lnprecio,1)
plot.ts(Dlnprecio, type="o", ylab="Retorno_Bolsa_de_Nueva_York")
summary (lnprecio)
summary (Dlnprecio)
```

Ejemplo 2: Series temporales financieras

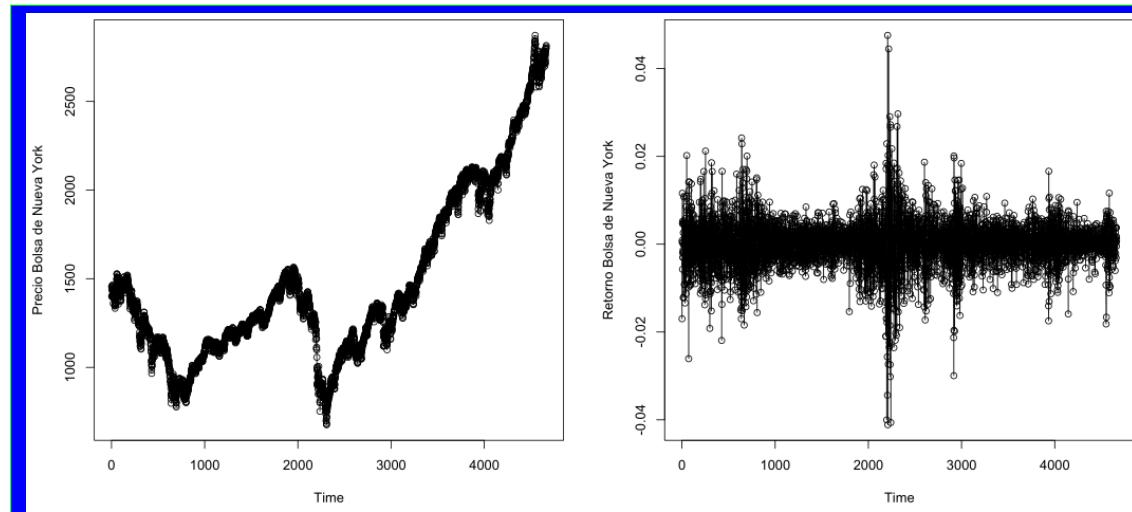


Figure 1.2: Indice Bolsa de Nueva York. Izquierda: precios reales SP500. Derecha: retornos SP500.

Los precios al parecer no son normales, aparentemente son log-normales.

Ejemplo 2: Histograma de los Precios

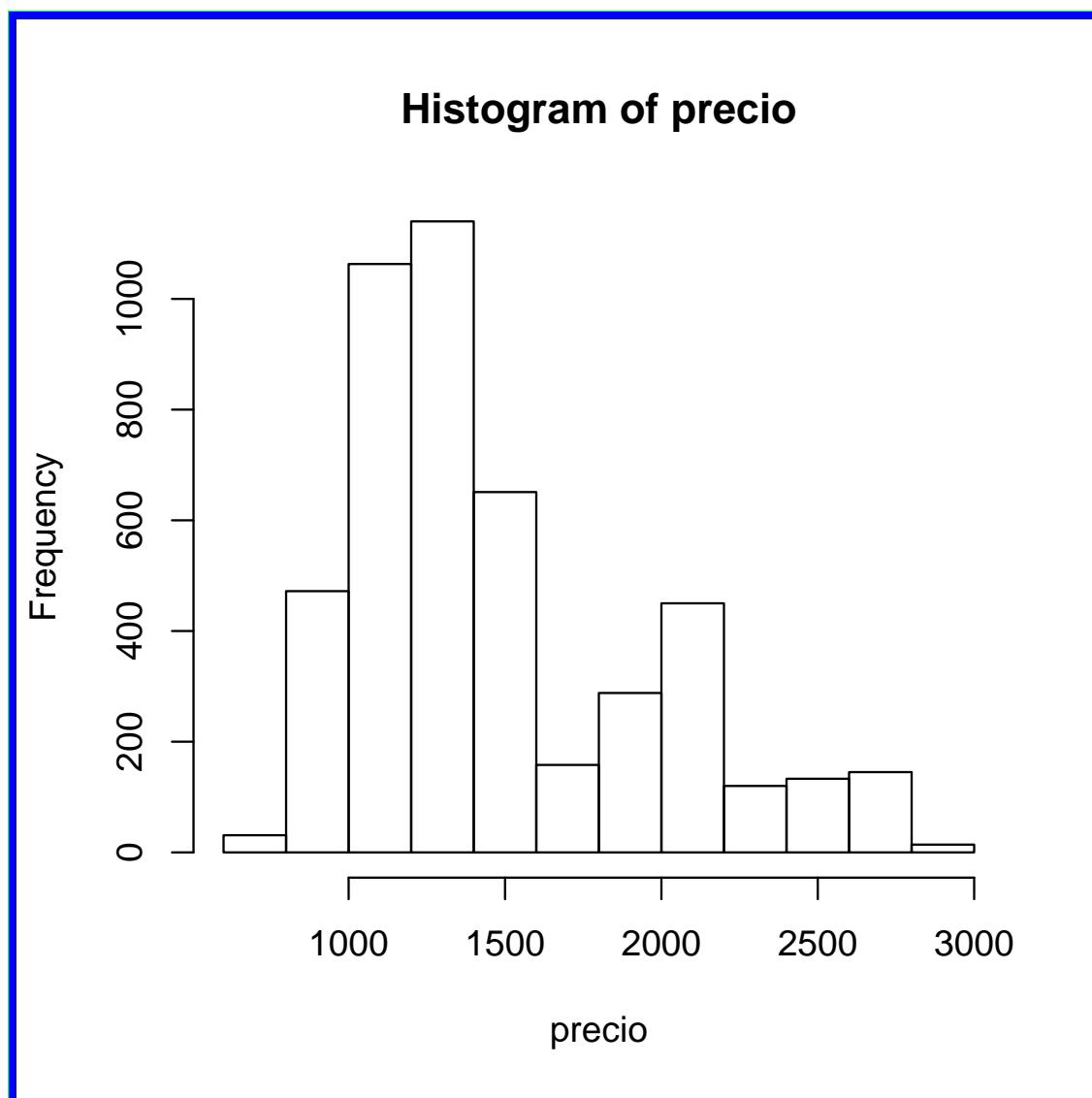


Figure 1.3: Histograma de los precios de la Bolsa de NY

Los retornos de los precios si parecen normales 1.4, lo cual constituye una propiedad deseable para el análisis estadístico.

Ejemplo 2: Histograma de los Precios

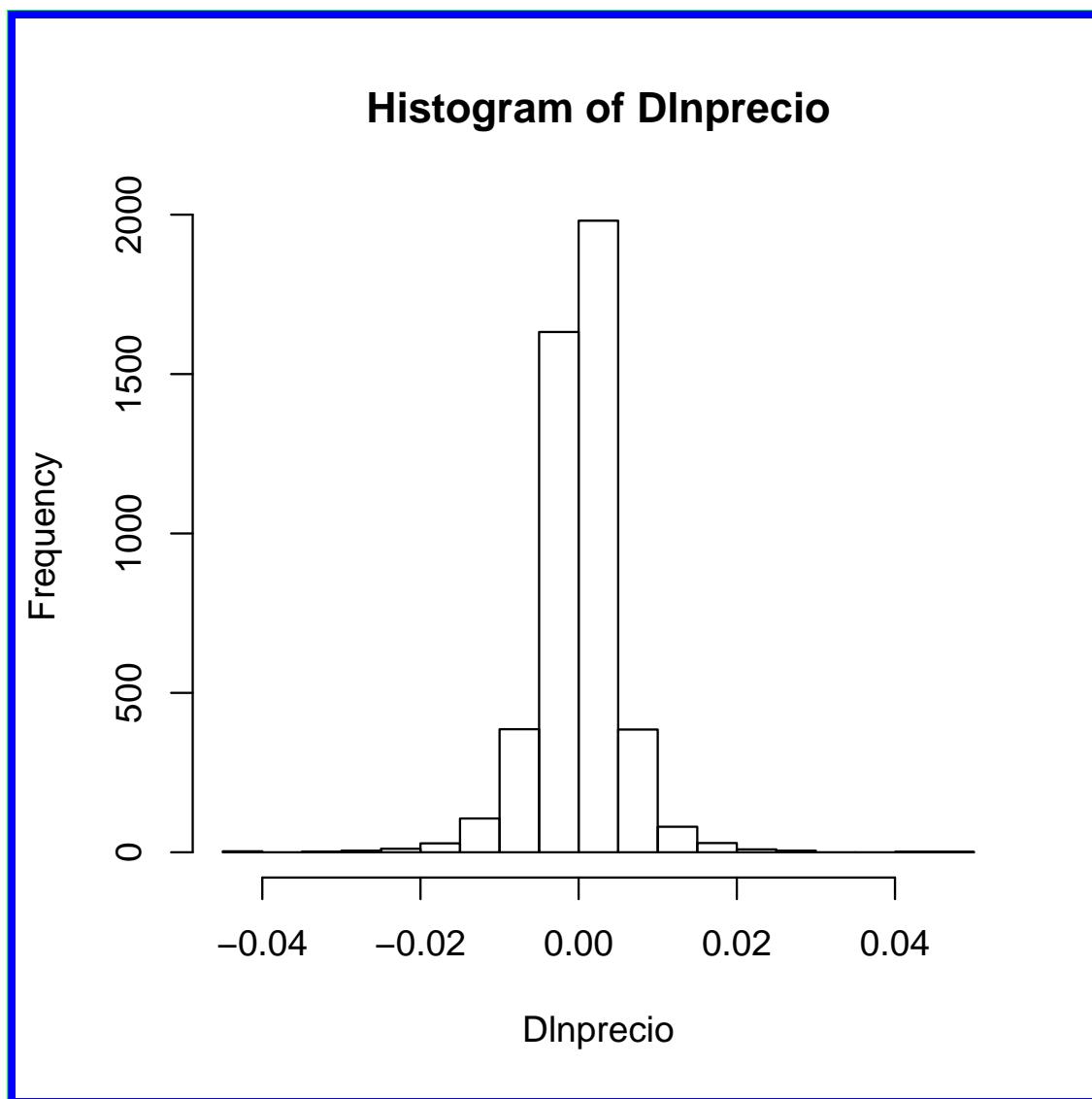


Figure 1.4: Histograma de los retornos de la Bolsa de NY

Ejemplo 2: Histograma de los Precios

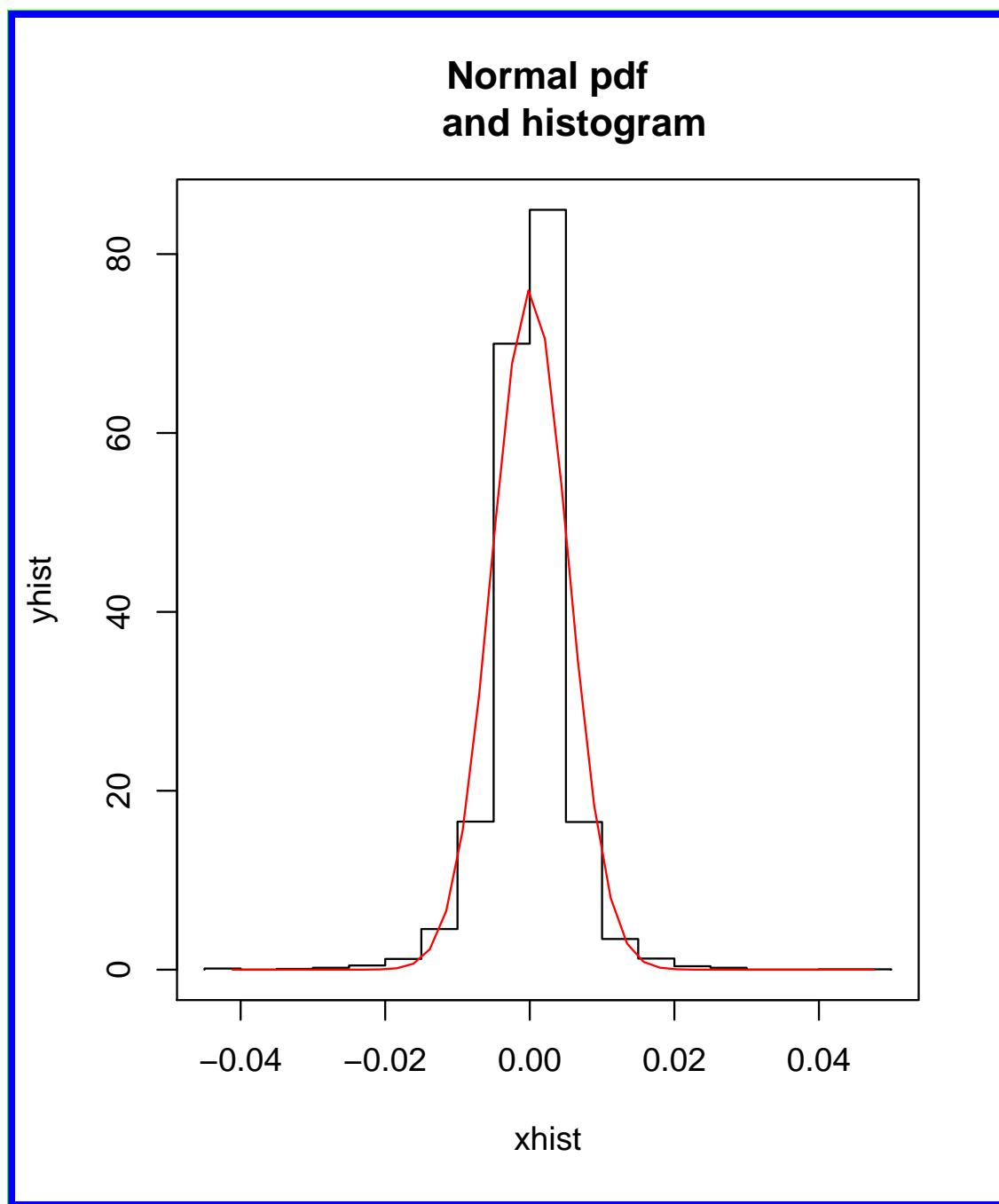


Figure 1.5: Linea negra:Histograma de las diferencias de los log-precios de la Bolsa de NY. Linea roja: ajuste de la distribución normal segúun media y desviación estandar de la muestra. Shapiro-Wilk normality test / data: Dlnprecio / W = 0.90927, p-value < 2.2e-16.

'Código R de histograma y ajuste de curva normal'

```

h <- hist(Dlnprecio ,breaks=15)
xhist <- c(min(h$breaks),h$breaks)
yhist <- c(0,h$density ,0)
xfit <- seq(min(Dlnprecio),max(Dlnprecio),length=40)
yfit <- dnorm(xfit,mean=mean(Dlnprecio),sd=sd(Dlnprecio))
plot(xhist,yhist,type="s",ylim=c(0,max(yhist,yfit)),
      main="Normal_pdf_and_histogram")
lines(xfit,yfit, col="red")
shapiro.test(Dlnprecio)

```

1.2 Modelamiento estadístico de las series de tiempo

Para poder modelar los datos, que aparentemente fluctúan de forma aleatoria a lo largo del tiempo, suponemos que una serie temporal se define como una colección de variables aleatorias. Por ejemplo, podemos modelar una serie temporal como una secuencia de variables aleatorias, x_1, x_2, x_3, \dots , donde la variable aleatoria x_1 denota el valor tomado por la serie en el primer punto de tiempo, la variable x_2 denota el valor para el segundo período de tiempo, y así sucesivamente. En general, una colección de variables aleatorias, $\{x_t\}$, indexada por t se conoce como proceso estocástico. En este texto, t será típicamente discreto y variará sobre los enteros $t = 0, \pm 1, \pm 2, \dots$, o algún subconjunto de los enteros. Los valores observados en un proceso estocástico se conocen como la realización del proceso estocástico.

1.2.1 Ruido blanco - White Noise

Una serie de tiempo muy utilizada es aquella representada por una colección de variables aleatorias no correlacionadas, ϵ_t , con media 0 y varianza finita σ_ϵ^2 . Las series temporales generadas a partir de variables no correlacionadas se utilizan por ejemplo para modelar el ruido en aplicaciones de ingeniería, donde se denomina ruido blanco. A veces denotaremos este proceso como $\epsilon_t \sim \epsilon_n(0, \sigma_\epsilon^2)$. La designación de "blanco" se origina de la analogía con la luz blanca e indica que todas las posibles oscilaciones periódicas están presentes con la misma fuerza.

- En ocasiones, también requeriremos que el ruido sea independiente y tenga una distribución idéntica (iid) de variables aleatorias con media 0 y varianza σ_ϵ^2 . Distinguiremos este caso diciendo ruido blanco independiente, o escribiendo $\epsilon_t \sim iid(0, \sigma_\epsilon^2)$.
- Otra serie de ruido blanco particularmente útil es el ruido blanco gaussiano, en el que las w_t son variables aleatorias normales independientes, con media 0 y varianza σ_ϵ^2 ; o más sucintamente, $\epsilon_t \sim iid N(0, \sigma_\epsilon^2)$.
- La figura siguiente muestra una colección de 500 de estas variables aleatorias, con $\sigma_\epsilon^2 = 1$, trazadas en el orden en que se dibujaron.

'Código R para generar Ruido blanco'

```
set.seed(154)
w = rnorm(200, 0, 1)
plot.ts(w, ylim=c(-3, 3), main="White_Noise")
```

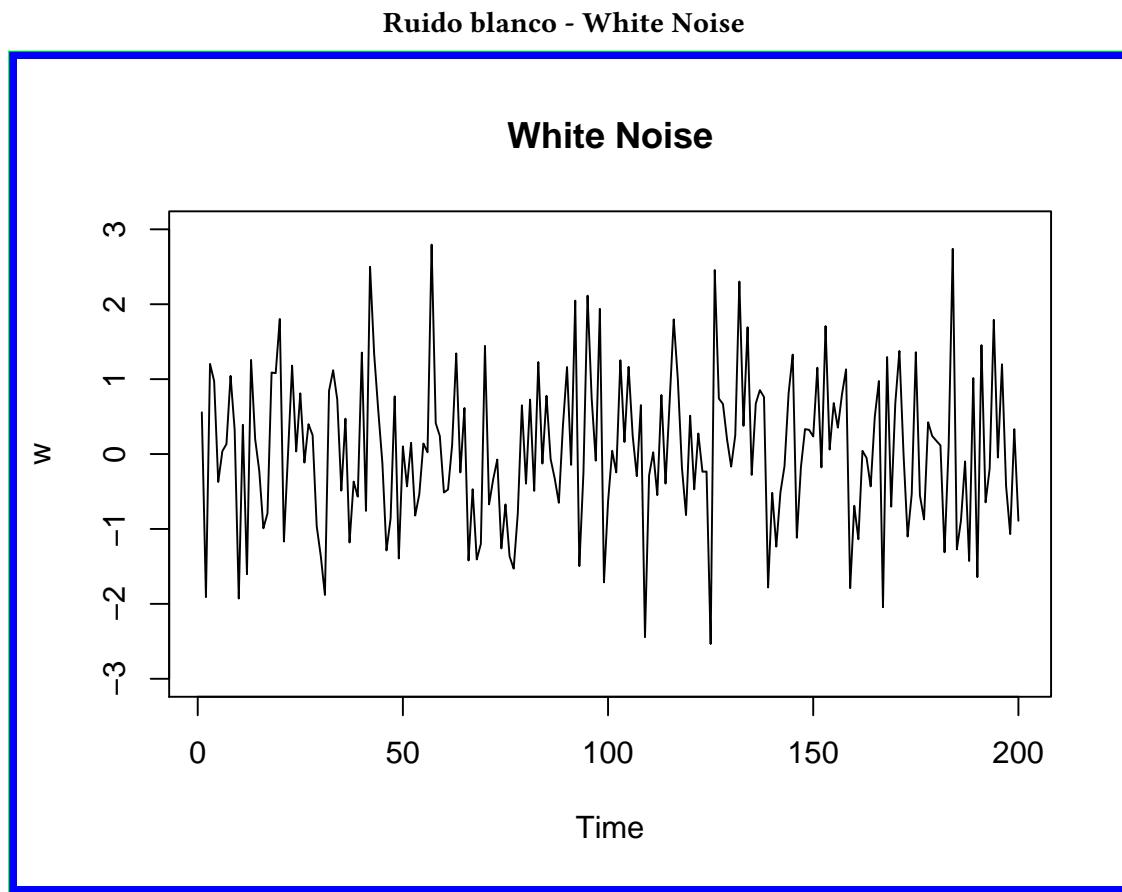


Figure 1.6: Ruido blanco - White Noise

1.2.2 Caminata Aleatoria - Random Walk

Un ejemplo simple para modelar una serie temporal estocástica con tendencia (no estacionaria) es una caminata aleatoria con deriva (Random Walk with drift):

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

para $t = 1, 2, \dots$, con una condición inicial $x_0 = 0$, y donde ϵ_t es ruido blanco. La constante δ se denomina deriva (drift), y cuando $\delta = 0$, se llama simplemente una caminata aleatoria. El

término caminata aleatoria proviene del hecho de que, cuando $\delta = 0$, el valor de la serie de tiempo en el tiempo t , es el valor de la serie en el tiempo $t - 1$ más un movimiento completamente aleatorio determinado por ϵ_t .

Tenga en cuenta que podemos reescribir la ecuación anterior como una suma acumulativa de las variables de ruido blanco. Es decir,

$$x_t = \delta t + \sum_{j=1}^t \epsilon_t$$

'Código R para generar Caminata Aleatoria'

```
set.seed(154)
w = rnorm(200, 0, 1)
x = cumsum(w)
wd = w + 0.2
xd = cumsum(wd)
plot.ts(xd, ylim=c(-5,55), main="random_walk")
lines(x)
lines(0.2*(1:200), lty="dashed")
```

Caminata aleatoria

random walk

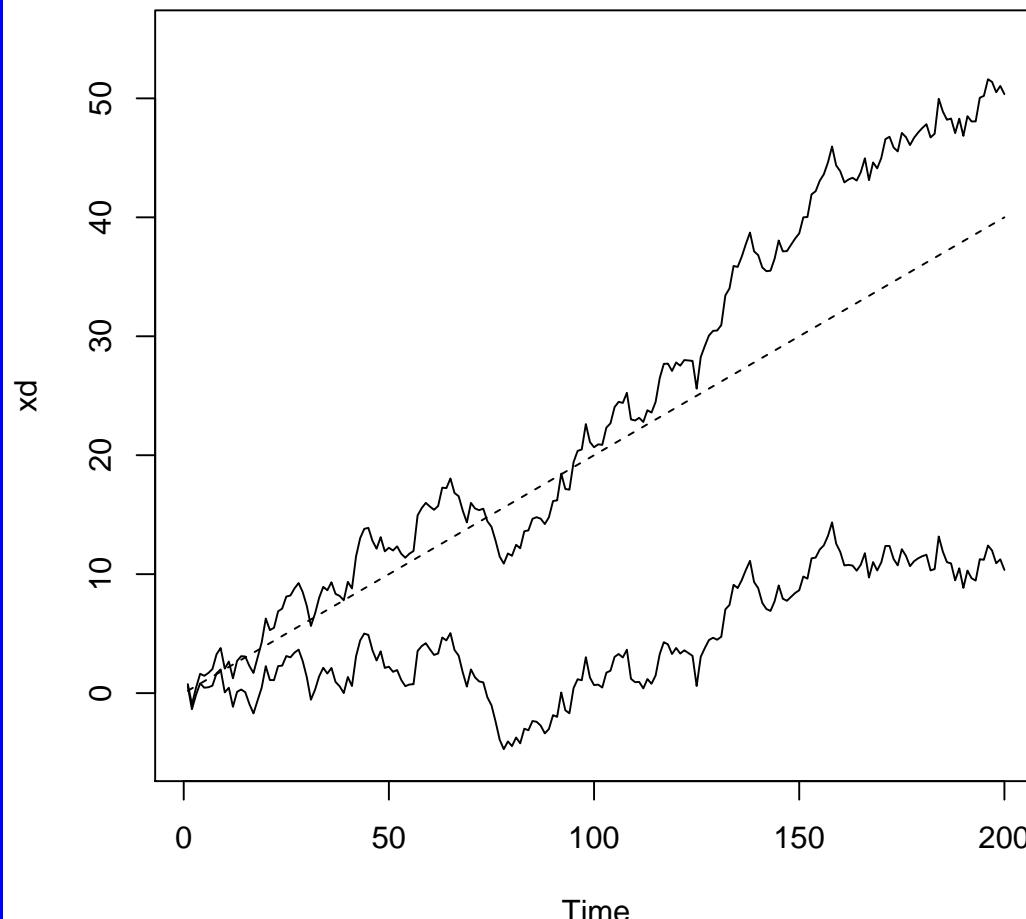


Figure 1.7: Caminata Aleatoria - Random walk, $\sigma_\epsilon = 1$, with drift $\delta = 0.2$ (upper jagged line), without drift, $\delta = 0$ (lower jagged line), and a straight line with slope 0.2 (dashed line).

1.2.3 Promedios móviles - Moving Averages

Podríamos reemplazar la serie de ruido blanco ϵ_t por un promedio móvil que suavice la serie. Por ejemplo, considere reemplazar ϵ_t por un promedio de su valor actual y sus vecinos inmediatos en el pasado y futuro. Es decir:

$$v_t = 1/3(\epsilon_{t-1} + \epsilon_t + \epsilon_{t+1})$$

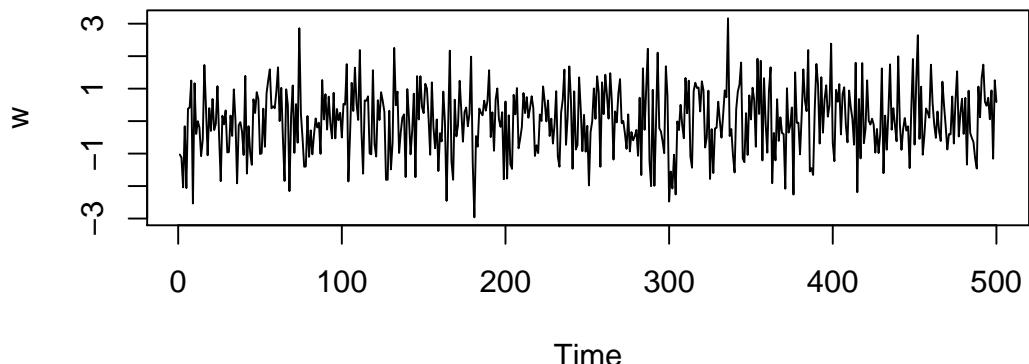
Como veremos en el siguiente ejemplo, los promedios móviles producen una versión más suave que la serie original, lo que refleja el hecho de que las oscilaciones más lentas llegan a ser más evidentes, y se eliminan algunas de las oscilaciones más rápidas.

‘Código R para generar Moving Averages’

```
w = rnorm(500,0,1) ; v = filter(w, sides=2, rep(1/3,3))
par(mfrow=c(2,1))
plot.ts(w, main="white_noise")
plot.ts(v, main="moving_average")
```

Medias móviles

white noise



moving average

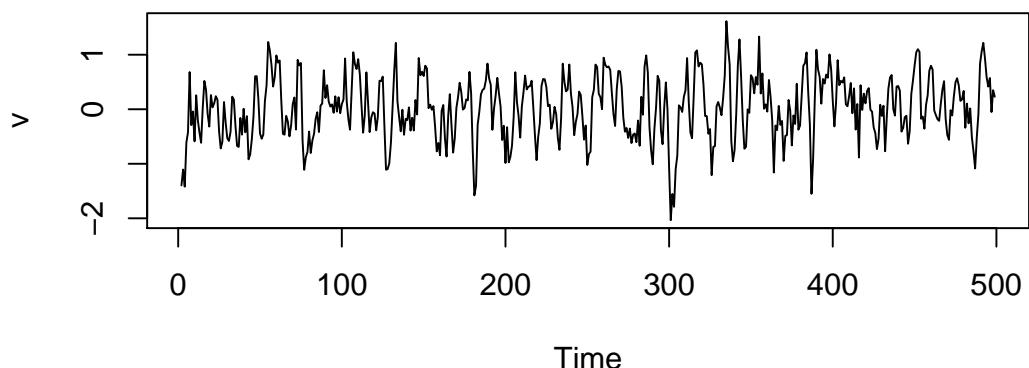


Figure 1.8: Promedios Móviles - Moving Averages

1.2.4 Autorregresiones - Autoregressions

Supongamos nuevamente que consideramos la serie de ruido blanco w_t como entrada, y calculamos la salida usando una ecuación de segundo orden, es decir:

$$x_t = x_{t-1} - 0.9x_{t-2} + \epsilon_t$$

Esta ecuación representa una regresión o predicción del valor actual x_t de una serie temporal en función de los dos valores anteriores de la serie, es por esto que utilizamos el nombre de autoregresión.

'Código R para generar Autoregresiones'

```
w = rnorm(550,0,1) ;  
x = filter(w, filter=c(1,-.9), method="recursive")[-(1:50)]  
plot.ts(x, main= "autoregression")
```

Autorregresiones

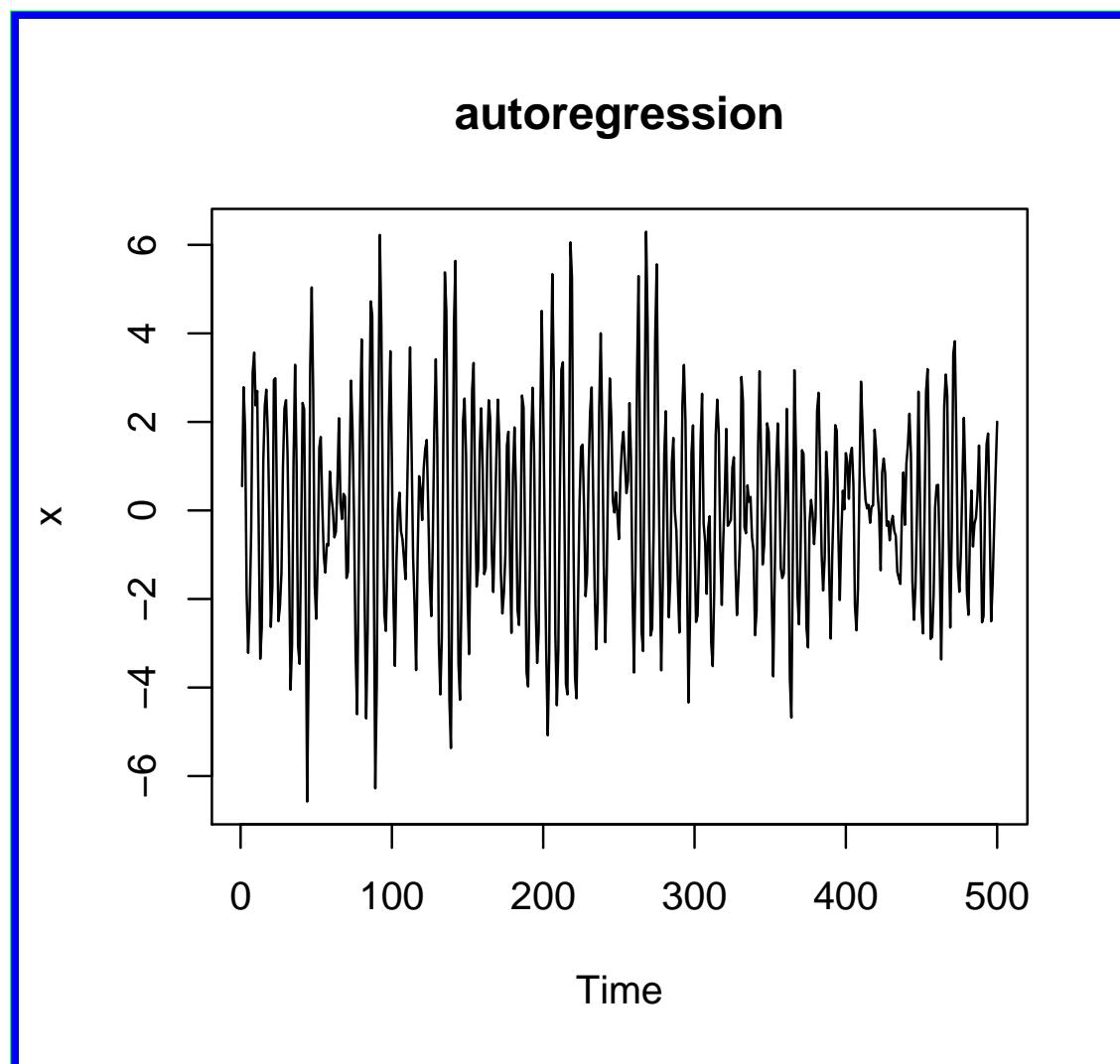


Figure 1.9: Autorregresiones - Autoregressions

1.3 Descomposición de las series de tiempo

Las series de tiempo usualmente se descomponen en:

- Una tendencia (trend) T_t .
- Un componente estacional (seasonal) S_t .
- Un elemento irregular (irregular) I_t .

Por ejemplo:

$$T_t = 2 + 0.1t;$$

$$S_t = 6.5\cos(\pi/60);$$

$$\text{y } I_t \sim N(\mu = 0, \sigma^2 = 0.5).$$

‘Código R para descomponer serie de tiempo’

```
t = 2 + 0.1 * 1:500
s = 6.5 * cos(pi * 1:500/90)
set.seed(154)
i = rnorm(500, 0, 5)
plot.ts(s + t + i)
```

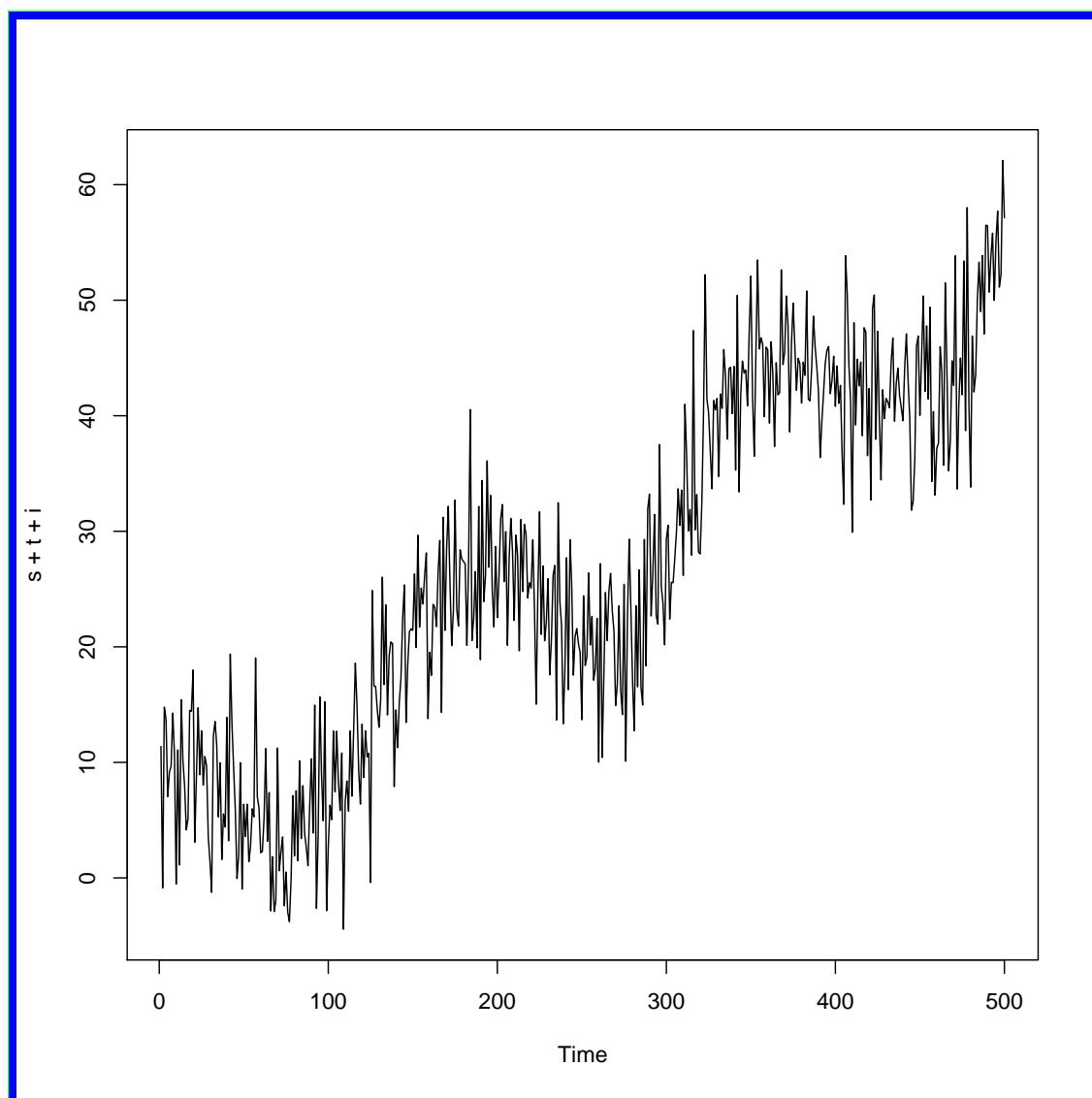
Componentes de una Serie de Tiempo

Figure 1.10: Serie de tiempo con sus componentes T, S e I.

En general, las series de tiempo pueden contener uno o combinación de todos los elementos antes señalados, ya sea de manera aditiva o multiplicativa:

$$x_t = T_t + S_t + I_t$$

$$x_t = T_t * S_t * I_t$$

- La primera especificación se caracteriza por tener cada componente de forma independiente, lo que posibilita descomponer la serie en una suma de los tres factores.
- La segunda, por otra parte, surge cuando la tendencia (T_t), estacionalidad (S_t), y irregularidad (I_t) son dependientes entre si.

En general, la tendencia cambia la media de la serie, mientras que el componente estacional posee un patrón que se repite, por ejemplo de manera mensual o trimestral. El componente irregular a pesar de no tener un patrón bien definido, puede ser pronosticada, de hecho, los pronósticos usan correlaciones con el componente irregular para realizar sus pronósticos. En períodos más largos sin embargo, el componente irregular exhibe una tendencia de reversión a cero.

El pronóstico de series de tiempo pretende entonces predecir cada uno de éstos componentes de manera individual. Como vimos, el pronóstico global de la series de tiempo agrupa de forma aditiva o multiplicativa cada uno de dichos componentes.

1.3.1 Descomposición Tendencia - Filtro Hodrick -Prescott

- En economía, el filtro Hodrick-Prescott (HP) permite separar para x_t los componentes tendencial y cíclico.
- Este método consiste en obtener una serie suavizada S_t a partir de la original x_t , mediante una solución al problema de optimización sugerido en la siguiente ecuación. Una vez resuelto, permite estimar tanto el ciclo como la tendencia de la serie.

$$\min \sum_{t=1}^n (x_t - S_t)^2 + \lambda \sum_{t=2}^{n-1} [(S_{t+1} - S_t) - (S_t - S_{t-1})]^2$$

- Los valores sugeridos para λ dependen de la periodicidad de x_t , y son: 14400 (mensual), 1600 (trimestral) y 100 (anual). Por otra parte, una vez obtenido el componente cíclico ($x_t - S_t$), a partir del filtro HP, puede ser interpretada como la brecha existente entre su valor real x_t y potencial S_t .

1.3.2 Descomposición componente estacional - Transformaciones de Diferencias

La transformaciones de diferencias se usan para capturar el componente estacional de la serie:

Una primera diferencia es definida como:

$$\Delta x_t = (x_t - x_{t-1})$$

Análogamente, la segunda diferencia es definida como:

$$\Delta^2 x_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$$

Anteriormente vimos que la primera diferencia del logaritmo, se podía interpretar como el cambio porcentual de la variable, logrando la simetría del precio de las acciones.

1.3.3 Descomposición componente estacional - Variables Dummy

Una variable dummy, D, es una variable binaria que toma la siguiente forma:

- D=1 si la observación tiene características específicas.
- D=0 si no las tiene.

Por ejemplo:

$$x_t = \beta_0 + \beta_1 z_t + \beta_2 D + \beta_3 Dz_t$$

$$D = 1 \Rightarrow x_t = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)z_t$$

$$D = 0 \Rightarrow x_t = \beta_0 + \beta_1 z_t$$

Las variables dummy pueden ser usadas para cambiar la pendiente y/o intercepto en un modelo lineal, lo cual permite capturar la estacionalidad en la serie, por ejemplo con variables dummies por trimestre o temporada.

1.4 Medidas de dependencia

Como vimos anteriormente, una serie de tiempo puede ser vista como una colección de n variables aleatorias en puntos de tiempo enteros arbitrarios t_1, t_2, \dots, t_n , para cualquier entero positivo se cuenta con una función de distribución conjunta, evaluada como la probabilidad que los valores de las series sean conjuntamente menores que n constantes, c_1, c_2, \dots, c_n , i.e.:

$$F(c_1, c_2, \dots, c_n) = P(x_{t_1} \leq c_1, x_{t_2} \leq c_2, \dots, x_{t_n} \leq c_n)$$

Desafortunadamente, la función de distribución multinomial no puede usualmente ser escrita fácilmente a menos que las variables sean conjuntamente normales, en tal caso la densidad conjunta presenta la forma:

$$f(\mathbf{x}) = (2\pi)^{-n/2} |\Gamma|^{-1/2} \exp\{-1/2(\mathbf{x} - \boldsymbol{\mu})' \Gamma^{-1} (\mathbf{x} - \boldsymbol{\mu})\}$$

donde $|\cdot|$ indica determinante y Γ la matriz de covarianza.

Aunque la función de distribución conjunta permite describir la data completamente, su manipulación es muy compleja, y su despliegue gráfico imposible.

Las funciones de distribución marginales:

$$F_t(x) = P\{x_t \leq x\}$$

o la correspondiente función de densidad marginal:

$$f_t(x) = \frac{\partial F_t(x)}{\partial x}$$

Cuando éstas existen, proveen información valiosa para examinar el comportamiento marginal de la serie.

Si x_t es Gausiana con media μ_t y varianza σ_t^2 , $x_t \sim N(\mu_t, \sigma_t^2)$, la densidad marginal está dada por:

$$f_t(x) = \frac{1}{\sigma_t \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_t^2}(x - \mu_t)^2\right)$$

La función de media, conocida en estadística como el primer momento central, está definida como:

Definition 1. Definición: Función de Media:

$$\mu_{xt} = E(x_t) = \int_{-\infty}^{+\infty} x f_t(x) dx$$

Si existe, E denota el operador de valor esperado.

La función de autocovarianza, conocida en estadística como el segundo momento central, está definida como:

Definition 2. Definición: Función de Autocovarianza:

$$\gamma_x(s, t) = cov(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)]$$

En este caso, $\gamma_x(s, t) = \gamma_x(t, s)$ para todos los puntos de s y t . Si $\gamma(s, t) = 0$ podemos asegurar que x_s y x_t no están linealmente relacionadas. Por otro lado, si x_s y x_t son además normales bivariadas podemos asegurar que son independientes.

Por último, es claro que si $s = t$, la autocovarianza se reduce a la varianza

Definition 3. Definición: Función de Varianza:

$$\gamma_x(t, t) = E[(x_t - \mu_t)^2] = var(x_t)$$

La Función de Autocorrelación, denotada por ACF (autocorrelation function), mide la predictibilidad lineal de la serie en el tiempo t . Es decir, predecimos x_t , utilizando sólo el valor x_s . Suponiendo que ambas series tienen varianzas finitas, tenemos la siguiente definición:

Definition 4. Definición: Función de Autocorrelación:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}$$

Se puede demostrar fácilmente que $-1 \leq \rho(s, t) \leq 1$. Si podemos predecir x_t perfectamente a partir de x_s a través de una relación lineal, $x_t = \beta_0 + \beta_1 x_s$, entonces la correlación será +1 cuando $\beta_1 > 0$, y -1 cuando $\beta_1 < 0$.

Por lo tanto, tenemos una medida aproximada de la capacidad de predecir la serie en el tiempo t desde el valor en el tiempo s .

La función de covarianza cruzada entre dos series, x_t e y_t , viene dada por:

Definition 5. Definición: Función de Covarianza Cruzada:

$$\gamma_{xy}(s, t) = cov(x_s, y_t) = E[(x_s - \mu_{xs})(y_t - \mu_{yt})].$$

La función de correlación cruzada (CCF por su sigla en inglés cross-correlation function) está dada por:

Definition 6. Función de Correlación Cruzada:

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s)\gamma_y(t, t)}}$$

Podemos extender fácilmente las formulaciones anteriores al caso de más de dos series, por ejemplo, $x_{t1}, x_{t2}, \dots, x_{tr}$; es decir, series temporales multivariantes con r componentes. Por ejemplo, la extensión de (1.10) para el caso de covarianza cruzada sería:

$$\gamma_{xy}(j, k) = cov(x_{sj}, y_{sk}) = E[(x_s - \mu_{xs})(y_{tj} - \mu_{y_{tk}})].$$

1.5 Estacionaridad

Definition 7. Definición: Estacionariedad estricta: Una serie de tiempo es estrictamente estacionaria (*strictly stationary*) si el comportamiento probabilístico de cada conjunto de valores $\{x_{t_1}, x_{t_2}, \dots, x_{t_k}\}$ es idéntico al del mismo conjunto desplazado en el tiempo, es decir $\{x_{t_1+h}, x_{t_2+h}, \dots, x_{t_k+h}\}$.

En otras palabras:

$$P(x_{t_1} \leq c_1, \dots, x_{t_k} \leq c_k) = P(x_{t_1+h} \leq c_1, \dots, x_{t_k+h} \leq c_k)$$

para todos los $k = 1, 2, \dots$, todos los períodos t_1, t_2, \dots, t_k , todos los números c_1, c_2, \dots, c_k , y todos los desplazamientos de tiempo $h = 0, \pm 1, \pm 2, \dots$.

Si una serie de tiempo es estrictamente estacionaria, entonces todas las funciones de distribución multivariante para subconjuntos de variables, deben ser iguales con sus contrapartes en el conjunto desplazado. Por ejemplo, cuando $k = 1$:

$$P\{x_s \leq c\} = P\{x_t \leq c\}$$

para cualquier punto de tiempo s y t .

Cuando $k = 2$ tenemos:

$$P\{x_s \leq c_1, x_t \leq c_2\} = P\{x_{s+h} \leq c_1, x_{t+h} \leq c_2\}$$

Para cualquier punto de s, t y h . Si la función de varianza existe, entonces: $\gamma(s, t) = \gamma(s+h, t+h)$

En este contexto, ¿es un proceso random walk with drift estrictamente estacionario?

Definition 8. Definición: Estacionaridad débil: Una serie de tiempo x_t débilmente estacionaria (weakly stationary) es un proceso de varianza finita tal que:

- (i) la función de valor medio, μ_t , es constante y no depende del tiempo t , y
- (ii) la función de autocovarianza, $\gamma(s, t)$, depende de s y t sólo a través de su diferencia $|s - t|$.

De ahora en adelante, usaremos el término estacionario para significar débilmente estacionario; si un proceso es estacionario en sentido estricto, utilizaremos el término estrictamente estacionario. Un caso importante en el que la estacionariedad implica una estacionariedad estricta es si la serie es gaussiana (es decir, todas las distribuciones finitas de la serie son gaussianas).

1.6 Modelos de regresión lineal múltiple en series de tiempo

A continuación introducimos (recordamos) el modelo clásico de regresión lineal.

- Sea \mathbf{X} una matriz de $n \times k$ entradas donde se tienen n observaciones para k independientes variables.
- Sea \mathbf{Y} un vector de n observaciones de la variable dependiente.
- Es posible proponer un modelo de estimación lineal que relaciona las variables \mathbf{X} y la variable \mathbf{Y} :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

El modelo también puede ser escrito de forma compacta como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Vemos que este modelo presenta componentes sistemáticos (determinísticos) ($\mathbf{X}\boldsymbol{\beta}$) y estocásticos ($\boldsymbol{\epsilon}$).

Lo que se busca es determinar los coeficientes β_i que relacionen linealmente a las variables X_i e \mathbf{Y} . Para esto utilizamos el método de mínimos cuadrados ordinarios (MCO, también conocido como OLS por su sigla en inglés: ordinary least square) El criterio de mínimos cuadrados busca minimizar la suma de los cuadrados de los residuos ($\sum e_i^2$).

A continuación revisamos la derivación del método de MCO. Primero, el vector de residuos se puede obtener como:

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$$

donde $\hat{\beta}$ representa el estimador del vector β .

Por lo que la sumatoria del cuadrado de los errores será:

$$\mathbf{e}'\mathbf{e} = [e_1 \ e_2 \ \dots \ e_n] \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} = [e_1^2 + e_2^2 + \dots + e_n^2]$$

Por otro lado, puede escribirse también como:

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \\ &= \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\hat{\beta} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \\ &= \mathbf{Y}'\mathbf{Y} - 2\hat{\beta}'\mathbf{X}'\mathbf{Y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned}$$

Para minimizar el cuadrado de los residuos recurrimos al cálculo diferencial:

$$\frac{\partial(\mathbf{e}'\mathbf{e})}{\partial\hat{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} = 0$$

Luego $\hat{\beta}$ sera un mínimo de $\mathbf{e}'\mathbf{e}$ si la segunda derivada es positiva o equivalentemente \mathbf{X} es definida positiva.

De la expresión anterior obtenemos:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y}$$

Finalmente, multiplicando por $(\mathbf{X}'\mathbf{X})^{-1}$ a ambos lados, obtenemos $\hat{\beta}$:

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

La facilidad de cálculo del MCO ha influido en su popularidad. Los estimadores se obtienen a través de sencillas operaciones de matrices

1.6.1 Propiedades de los estimadores de MC

La estimación de los MCO es la mejor estimación lineal no sesgada (BLUE, best linear unbiased estimators). La prueba de esta proposición es provista por el teorema Gauss-Markov.

1. (i) No sesgado: $E(\hat{\beta}) = \beta$, es decir el valor esperado de el estimador es el verdadero valor de el parámetro desconocido.

2. (ii) Mejor: mínima varianza.

1.6.2 Teorema Gauss-Markov

Theorem 1.1: Supuestos

1. Existe una relación lineal entre X e Y
2. No existe multicolinealidad (X es linealmente independiente)
3. $E(\epsilon|X) = 0$. Equivalentemente $E(Y) = X\beta$
4. $E(\epsilon\epsilon'|X) = \sigma^2 I$. Los errores son homocedásticos y no existe autocorrelación
5. X y ϵ no se encuentran relacionados. $Cov(X\epsilon) = 0$

Usualmente, todos estos supuestos son chequeados en el proceso de pruebas de diagnóstico

Prueba Teorema. MCO es el mejor estimador lineal, insesgado, eficiente y de mínima varianza para β (BLUE)

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= \beta + (X'X)^{-1}X'\epsilon\end{aligned}$$

$$\begin{aligned}E(\hat{\beta}) &= E(\beta + (X'X)^{-1}X'\epsilon) \\ &= E(\beta) + E((X'X)^{-1}X'\epsilon) \\ &= \beta + (X'X)^{-1}X'E(\epsilon) \\ &= \beta\end{aligned}$$

□

1.6.3 Modelos de regresión lineal múltiple en series de tiempo

- $\hat{\beta}$ es un estimador lineal de β .

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + \epsilon) \\ &= \beta + (X'X)^{-1}X'\epsilon \\ &= \beta + A\epsilon\end{aligned}$$

Probar que se trata de un estimador de mínima varianza.

1.6.4 Evaluación Estadística de Regresiones Estimadas

El coeficiente de determinación R^2 es una medida de buen ajuste, el grado para el cual las variables independientes conjuntamente explican la variación en la variable dependiente sobre su media. El R^2 se incrementa cada vez que el número de regresores, k , se aumenta, relativo al tamaño de la muestra, n , independientemente de la justificación teórica de incluir variables adicionales. En el límite, si $n = k + 1$, $R^2 = 1$ pero tal regresión tiene cero poder explicativo.

El R^2 ajustado toma en cuenta el número de regresores relativo al tamaño de la muestra. El

R^2 ajustado es particularmente útil para evaluar el ajuste relativo de un conjunto de regresiones estimadas para la misma variable dependiente pero con un número diferente de variables independientes. Un criterio mecánico de selección de modelos es maximizar el R^2 ajustado.

Test t

Los tests t, son test de hipótesis sobre los parámetros estimados, para determinar si ellos son individualmente significativamente diferentes de cero. Hipótesis nula: $H_0: \beta_j = 0$.

Test estadístico

$$\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - 0}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \sim t(n - k - 1)$$

$SE(\hat{\beta}_j)$ = error estándar de el parámetro estimado

1.6.5 Regresiones sin sentido y regresiones espúrias.

Las regresiones sin sentido, son series de tiempo mutuamente independientes que producen buenos indicadores en la regresión, debido al alto nivel de correlación serial en cada serie. **Las regresiones espúrias** ocurren cuando los datos dependen sobre un tercer factor común, por ejemplo: una tendencia temporal. La relación espuria da la impresión de que existe un vínculo estadístico entre dos variables, el cual es invalidado cuando se examina objetivamente.

1.6.6 Ejemplo 3: Regresiones

Como un nuevo ejemplo, calculamos el beta de un activo financiero. El beta es una medida de riesgo sistemático, el cual se mide con respecto a la relación de los retornos del activo, con los del índice diversificado del mercado, en este caso el índice S&P ya estudiado. En este caso calculamos el beta de otro índice, el Russell 2000 (se denota por RUT) que mide el devenir de las empresas con small cap.

Regresiones

```

Call:
lm(formula = Dlnprecio_sp ~ Dlnprecio_rut)

Residuals:
    Min          1Q   Median       3Q      Max
-0.0193109 -0.0012208  0.0000333  0.0011929  0.0216599

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.041e-05 3.444e-05 -0.593   0.553
Dlnprecio_rut 7.171e-01 5.263e-03 136.237 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.002352 on 4662 degrees of freedom
Multiple R-squared:  0.7992, Adjusted R-squared:  0.7992
F-statistic: 1.856e+04 on 1 and 4662 DF,  p-value: < 2.2e-16

```

'Código R para Regresiones'

```

mydata1 <- read.csv("sp.csv", header=TRUE,
                     stringsAsFactors=FALSE)
precio_sp <- mydata1$"Adj.Close";
lnprecio_sp <- log10(precio_sp)
Dlnprecio_sp <- diff(lnprecio_sp ,1)
mydata2 <- read.csv("rut.csv", header= TRUE,
                     stringsAsFactors = FALSE)
precio_rut <- mydata2$"Adj.Close" ;
lnprecio_rut <- log10(precio_rut)
Dlnprecio_rut <- diff(lnprecio_rut ,1)
reg1 <- lm ( Dlnprecio_sp ~ Dlnprecio_rut)
summary(reg1)

```

Russell explicado por SP500

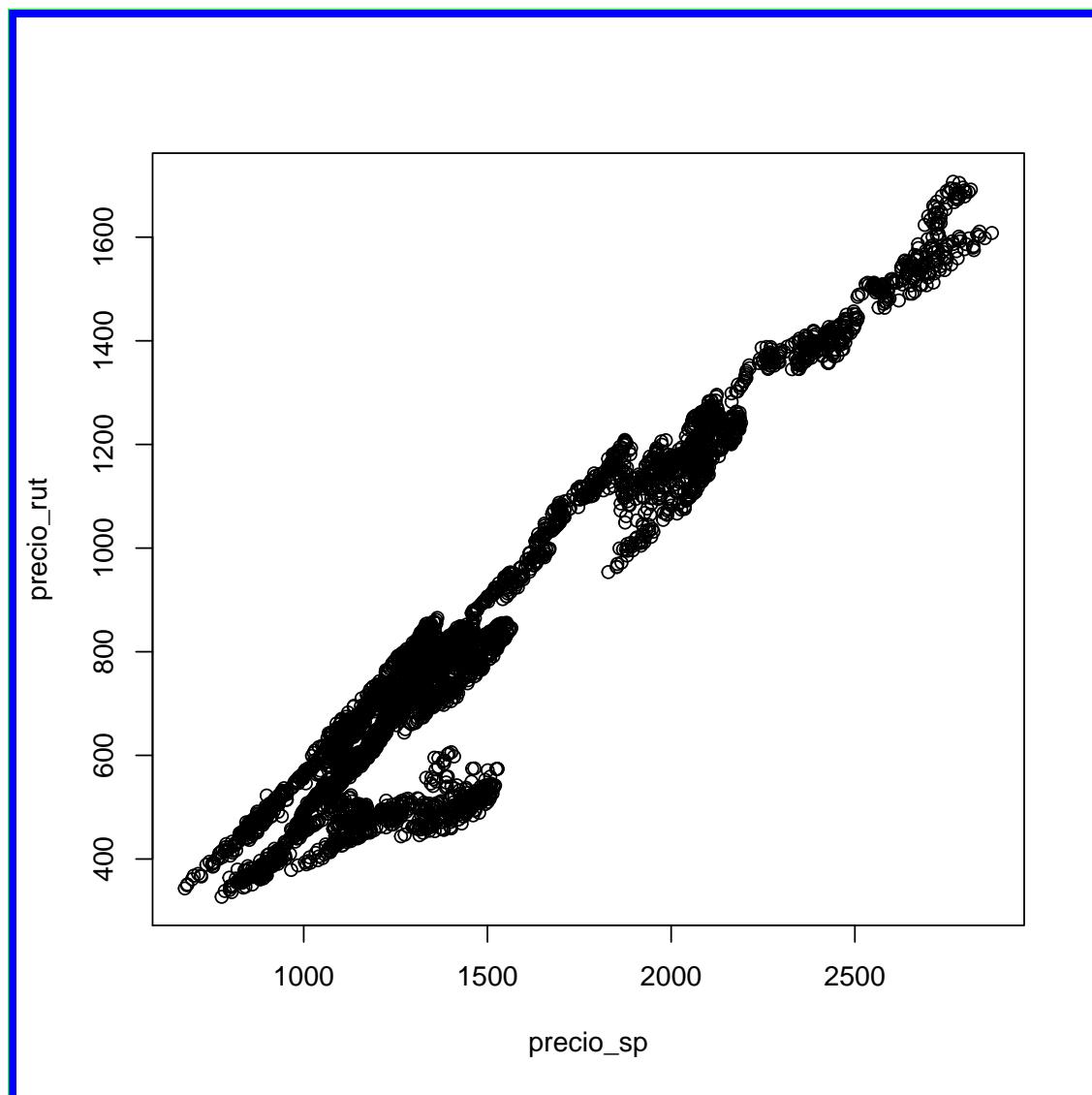


Figure 1.11: Regresión de Russell 2000 en función del S&P500

A partir del siguiente código almacenamos los residuos y analizamos los supuestos que exige una buena regresión.

‘Código R para Residuos de regresión’

```
residuos <- rstandard(reg1)
valores.ajustados <- fitted(reg1)
plot(valores.ajustados, residuos)
qqnorm(residuos)
qqline(residuos)
```

Homocedasticidad

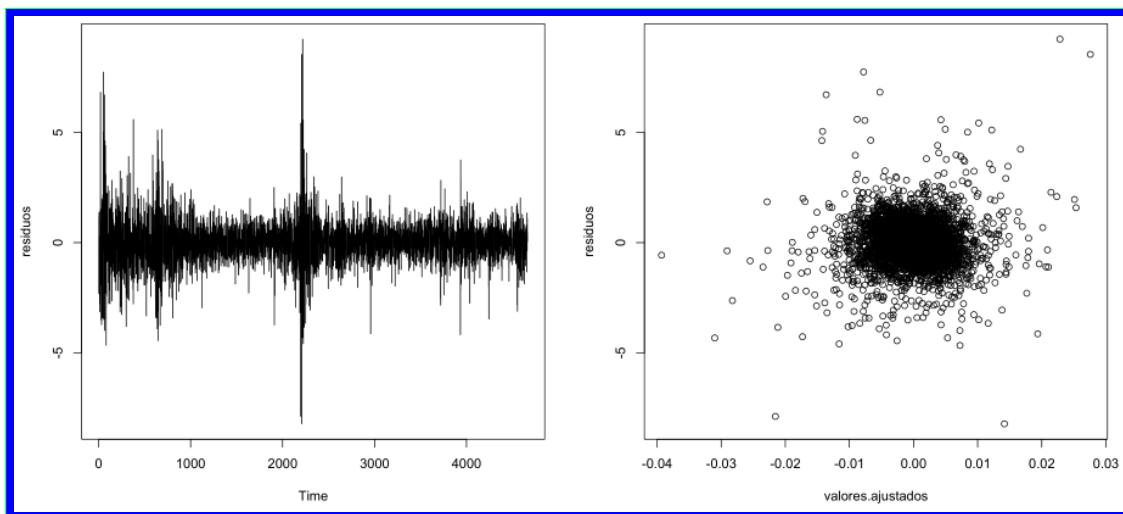


Figure 1.12: Izquierda: Derecha

Normalidad de los residuos

Normal Q–Q Plot

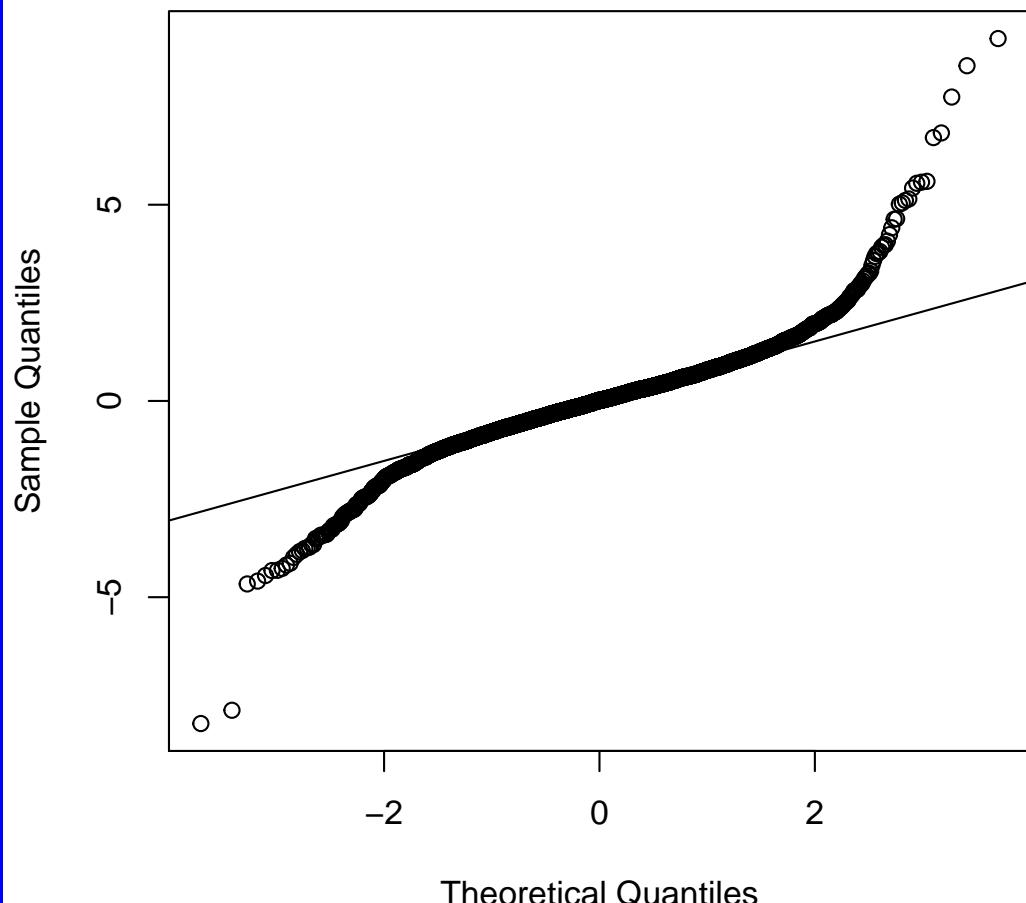


Figure 1.13: Grafico QQ-Plot de los residuos. La linea punteada indica el comportamiento de una distribución normal, mientras los puntos alejados de ella en los extremos dan cuenta de como se alejan los residuos del comportamiento normal.

Instructor: Marcelo Villena Chamorro PhD.

2 Tópico II.- Univariate Time Series Models

2.1 Ejemplo de repaso clase anterior-Detrending global temperature

Como vimos en la clase anterior, la evolución de la temperatura global manifestaba una tendencia lineal, por lo que podemos asumir que esta puede ser escrita como:

$$x_t = \mu_t + y_t$$

Veremos dos maneras de descomponer la serie, “filtrando” la tendencia.

‘Código R: ejemplo 1 Obtención de la componente “tendencia”de la serie temperatura global(gtemp).’

```
rm(list=ls())
mydata<-read.csv ("gtemp.csv")
gtemp<-mydata$"gtemp"
plot(gtemp, type="o", ylab="Global_Temperature_Deviations")
t<-1:142
summary(reg <- lm(gtemp ~ t))
plot(gtemp, type="o", ylab="Global_Temperature_Deviations")
abline(reg)
```

Ejemplo 1: Detrending global temperature

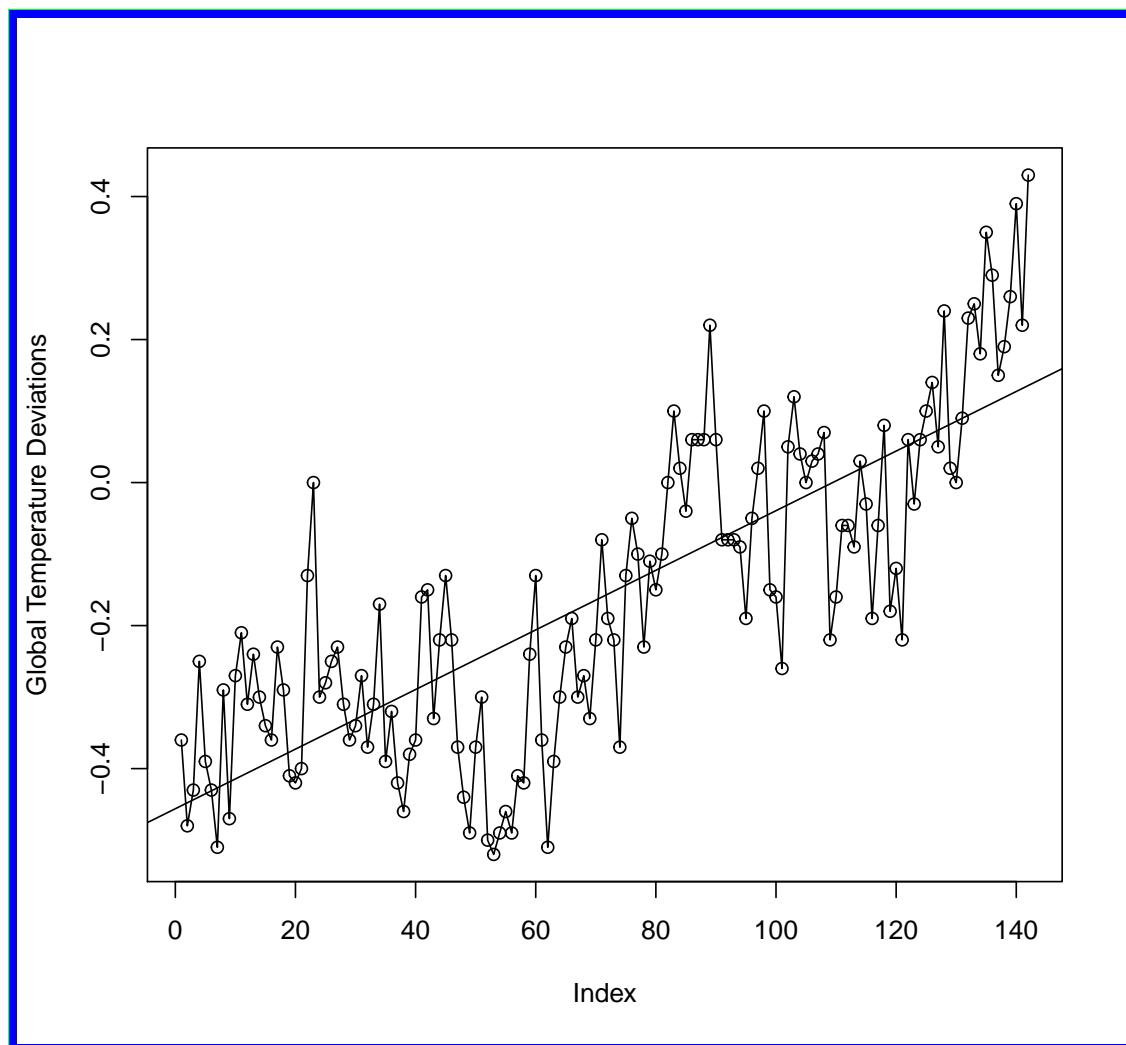


Figure 2.1: La figura muestra en línea punteada la tendencia de la serie de tiempo.

Detrending global temperature

```

Call:
lm(formula = gtemp ~ t)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.31231 -0.08627  0.00681  0.09064  0.36023 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.4560863  0.0227675 -20.03   <2e-16 ***
t             0.0041677  0.0002762   15.09   <2e-16 ***  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1349 on 140 degrees of freedom
Multiple R-squared:  0.6192,    Adjusted R-squared:  0.6164 
F-statistic: 227.6 on 1 and 140 DF,  p-value: < 2.2e-16

```

Figure 2.2: Parametría de la regresión de temperatura en el tiempo.

'Código R: ejemplo 1 Regresando gtemp sobre tiempo.'

```

reg1= lm(gtemp~time(gtemp), na.action=NULL)
par(mfrow=c(2,1))
plot(resid(reg1), type="o", main="detrended")
plot(diff(gtemp), type="o", main="first_difference")

```

Ejemplo 1: Detrending global temperature

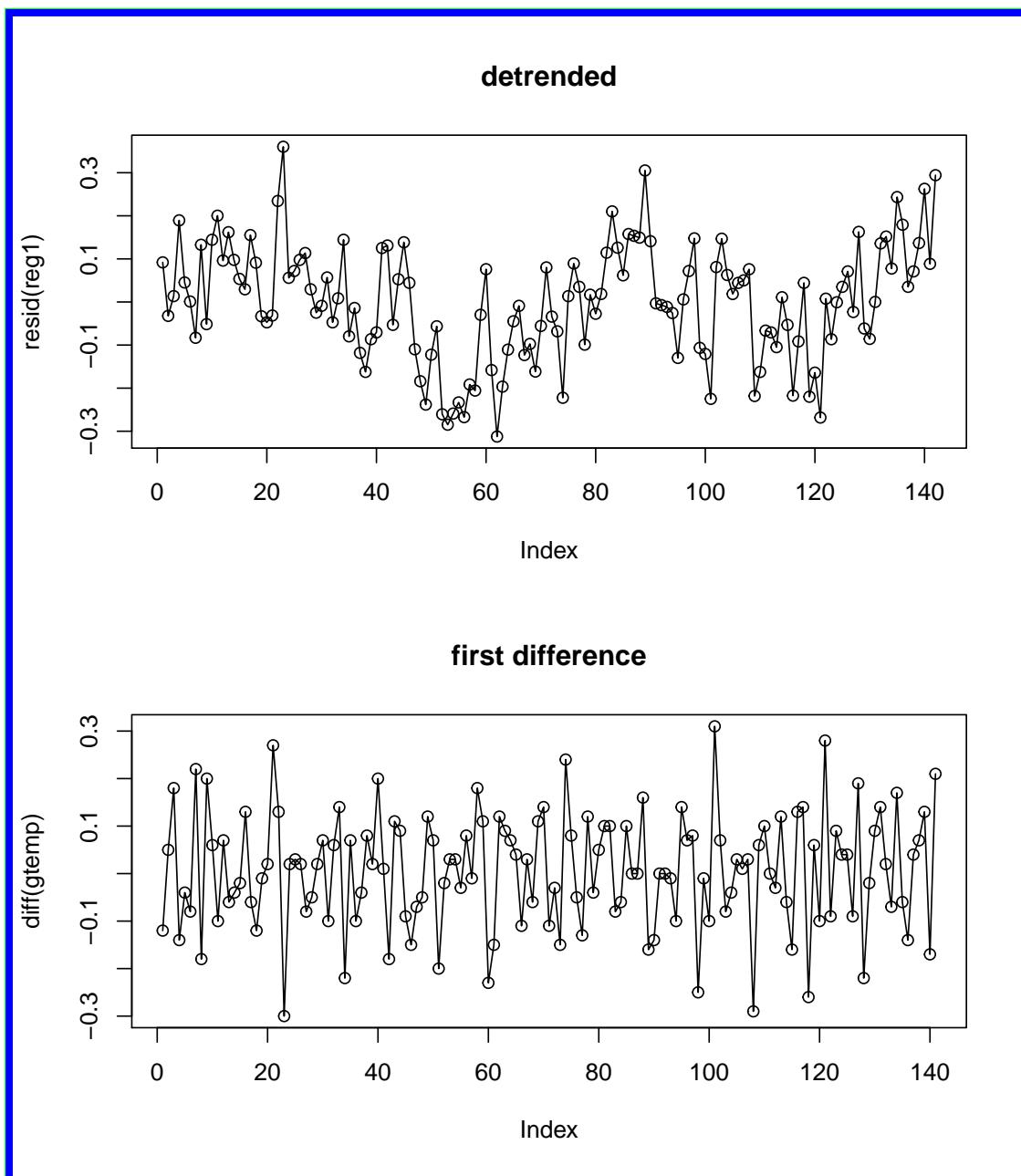


Figure 2.3: Arriba: residuos de la regresión. Abajo:primeras diferencias de la serie original

'Código R: ejemplo 1 Correlogramas de series.'

```
par(mfrow=c(3,1)) # plot ACFs
```

```
acf(gtemp, 48, main="gtemp")
acf(resid(reg), 48, main="detrended")
acf(diff(gtemp), 48, main="first_difference")
```

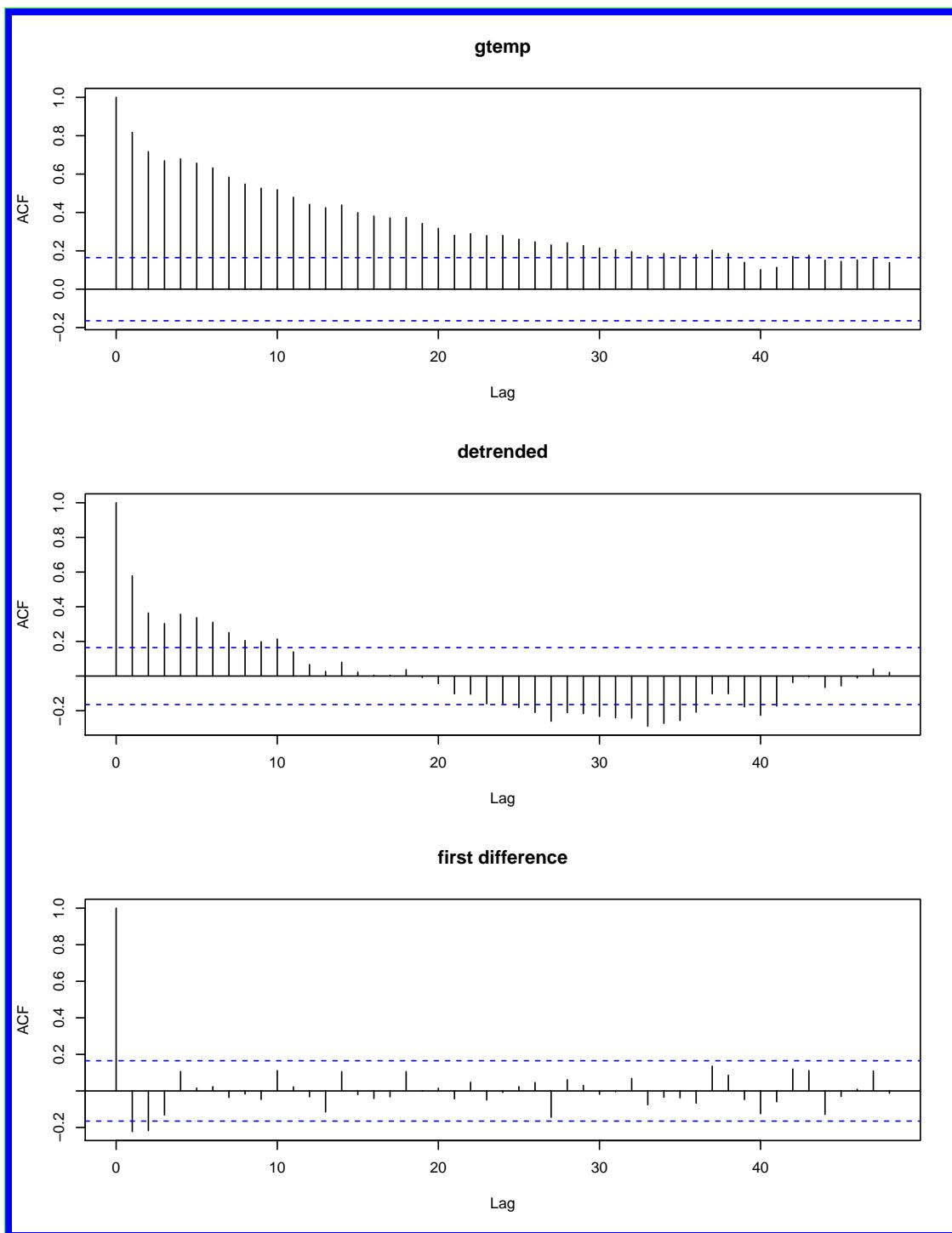
Ejemplo 1: Autocorrelograma

Figure 2.4: Autocorrelogramas para diferentes retardos(lags): Arriba:serie original. Medio: residuos de la regresión. Abajo: primera diferencia de la serie original

2.2 Sobre la descomposición de una serie

En los gráficos podemos apreciar que la primera diferencia de la serie produce resultados diferentes a la eliminación de la tendencia mediante la regresión de la tendencia.

En el caso de los gráficos ACF, el proceso diferenciado muestra una autocorrelación mínima, lo que puede implicar que la serie de temperatura global es similar una caminata aleatoria con deriva.

Es interesante notar que si la serie es una caminata aleatoria con deriva, la media de la serie diferenciada, que es una estimación de la deriva, es aproximadamente ,0066, pero con un gran error estándar:

‘Código R: ejemplo 1: sobre la descomposición de una serie - media y error estandar’

```
mean (diff(gtemp)) #media
sd (diff (gtemp)) / sqrt (length(diff (gtemp))) #error estandar
```

Una ventaja de diferenciar sobre la estimación de una tendencia, para eliminar las tendencias, es que no se estiman parámetros en la operación de diferenciación. Una desventaja, sin embargo, es que la diferenciación no arroja una estimación del proceso estacionario y_t .

De esta forma, si una estimación de y_t es esencial, entonces la estimación de una tendencia puede ser la forma más apropiada para eliminar las tendencias de la serie. Si el objetivo es forzar los datos a la estacionariedad, entonces la diferenciación puede ser más apropiada. La diferenciación también es una herramienta viable si la tendencia es fija.

En EE.UU. el procedimiento oficial de descomposición y ajuste estacional se llama:

X-13-ARIMA (<http://www.census.gov/srd/www/x13as/>)

2.3 Procesos no estacionarios, integrados y el test de raíz unitaria

Recordemos que si una serie de tiempo es estacionaria, su media, su varianza y su autocovarianza (en diferentes rezagos) permanecen iguales sin importar el momento del tiempo en el cual se midan; es decir, son **invariantes respecto al tiempo**.

Por otro lado, hemos visto que la estacionaridad es una característica deseable, por ejemplo, en términos de la normalidad de las variables. Sin embargo, en la práctica nos encontramos con:

- (i) Procesos No-estacionarios: Cuando un proceso estocástico de series de tiempo es dependiente del tiempo.
- (ii) Procesos Integrados: un proceso no-estacionario, el cual puede ser transformado a proceso estacionario diferenciando.

2.3.1 Procesos Integrados

Con respecto a los Procesos Integrados, partimos definiendo:

- La secuencia $\{x_t\}$ es integrada de orden d , $I(d)$, si esta requiere ser diferenciada d veces para llegar a ser estacionaria.
- Todos los **Procesos Integrados son no-estacionarios**, pero no todos los procesos no-estacionarios son integrados.
- Si la secuencia $\{x_t\}$ tiene una raíz unitaria, entonces, es un proceso integrado, y de aquí no-estacionario.

2.3.2 Consecuencias de los Procesos Integrados (Raíz Unitaria)

- Es importante señalar que los test estadísticos estándares no son apropiados cuando los MCO (OLS) son aplicados a procesos integrados, ver por ejemplo [Granger and Newbold \(1974\)](#).
- Si la secuencia $\{x_t\}$ es un proceso de raíz unitaria, entonces, cualquier shock tiene un efecto permanente (que no decae). De aquí, la serie de tiempo es modelada apropiadamente suponiendo una tendencia estocástica. La serie de tiempo entonces puede ser definida como estacionaria diferenciable, y se le deberá sacar la tendencia diferenciando.
- En este contexto, **los términos no-estacionariedad, caminata aleatoria, raíz unitaria y tendencia estocástica se consideran sinónimos**.

2.3.3 Test de Raíz Unitaria

Considere el siguiente proceso autoregresivo:

$$x_t = \alpha_1 x_{t-1} + \epsilon_t \quad (2.1)$$

Si $\alpha_1 = 1$, la secuencia x_t es una raíz unitaria.

El test estándar para probar esta hipótesis, consiste en restar x_{t-1} a la ecuación anterior de forma que:

$$\Delta x_t = \gamma x_{t-1} + \epsilon_t \quad (2.2)$$

donde $\gamma = \alpha_1 - 1$, y $\Delta x_t = x_t - x_{t-1}$. En este contexto, probar la hipótesis que la ecuación (1) tiene una raíz unitaria, $\alpha_1 = 1$, es equivalente a probar la hipótesis de $\gamma = 0$ en ecuación (2). Este es básicamente el enfoque de Dickey-Fuller (DF) para raíces unitarias, ver por ejemplo [Dickey and Fuller \(1981\)](#). Adicionalmente existe el test aumentado de Dickey-Fuller (ADF), y muchos otros tests que se basan en lógicas similares, y que utilizaremos durante el curso.

Ejemplo de Test de Raíz Unitaria - Dickey-Fuller

'Código R: Ejemplo Test de Raíz Unitaria - Dickey-Fuller '

```
library(tseries)
adf.test(gtemp)
adf.test(resid(reg1))
adf.test(diff(gtemp))
```

Ejemplo: Resultados de Test de Raíz Unitaria - Dickey-Fuller

```
Augmented Dickey-Fuller Test
```

```
data: gtemp
```

```
Dickey-Fuller = -2.0624, Lag order = 5, p-value = 0.5505
```

```
alternative hypothesis: stationary
```

```
Augmented Dickey-Fuller Test
```

```
data: resid(reg1)
```

```
Dickey-Fuller = -2.0624, Lag order = 5, p-value = 0.5505
```

```
alternative hypothesis: stationary
```

```
Augmented Dickey-Fuller Test
```

```
data: diff(gtemp)
```

```
Dickey-Fuller = -6.8179, Lag order = 5, p-value = 0.01
```

```
alternative hypothesis: stationary
```

Figure 2.5: Test de raíz unitaria Dickey-Fuller para diferentes series: Arriba:serie original (gtemp).

Medio: residuos de la regresión(resid(reg1)). Abajo: primera diferencia de la serie original (diff(gtemp))

2.4 Modelos ARIMA: modelando el corto plazo

En el año 1970, la metodología propuesta por George Box y Gwilym Jenkins, [Box and Jenkins \(1970\)](#), dos ingenieros con formación estadística sistematizan modelos estadísticos para el análisis de series temporales univariantes, teniendo en cuenta para esto la dependencia existente entre los datos.

Así, cada observación es modelada en función de los valores anteriores, la variable tiempo, por tanto, juega un papel fundamental.

Los modelos de predicción de Box-Jenkins pertenecen a la familia de modelos algébraicos lineales, que consideran que una serie temporal real constituye una probable realización de un determinado proceso estocástico.

Estos modelos se conocen con el nombre genérico de ARIMA (Auto-regresive Integrated Moving Average), el cual deriva de sus tres componentes Autoregresivo (AR), Integrado (I) de Medias Móviles (MA). Modelar una serie temporal supone identificar un modelo ARIMA adecuado que se ajuste a la serie objeto de estudio, debe contener los mínimos elementos necesarios para describir el fenómeno y ser útil para realizar previsiones.

2.4.1 Sobre el operador de retroceso - backshift operator

Definition 9. Definición: operador de retroceso (backshift operator):

$$Bx_t = x_{t-1} \quad (2.3)$$

$$B^2 x_t = B(Bx_t) = Bx_{t-1} = x_{t-2} \quad (2.4)$$

Así:

$$B^k x_t = x_{t-k} \quad (2.5)$$

De esta forma tenemos que la primera diferencia se puede definir en términos de lags, en otras palabras del operador de retroceso:

$$\Delta x_t = x_t - x_{t-1} = (1 - B)x_t \quad (2.6)$$

En general:

$$\Delta^d x_t = (1 - B)^d x_t \quad (2.7)$$

Definition 10. Definición: AR(p): Un modelo autorregresivo de orden p , frecuentemente abreviado como AR(p), tiene la forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t \quad (2.8)$$

donde x_t es una serie estacionaria, y $\phi_1, \phi_2, \dots, \phi_p$ son constantes. Si la media de x_t es μ , entonces podemos reemplazar $x_t - \mu$ en (2.8)

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) + \dots + \phi_p(x_{t-p} - \mu) + \epsilon_t \quad (2.9)$$

$$x_t = \alpha + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t \quad (2.10)$$

donde $\alpha = \mu(1 - \phi_1 - \phi_2 \dots - \phi_p)$

Usando los operadores de retroceso AR(p) queda como:

$$(1 - \phi_1 B + \phi_2 B^2 - \dots - \phi_p B^p) \quad (2.11)$$

o incluso más concisamente

$$\phi(B)x_t = \epsilon_t \quad (2.12)$$

2.4.2 Ejemplo Proceso Autoregresivo de Orden 1: AR(1)

En un procesos AR(1) la variable x_t queda únicamente por su valor pasado x_{t-1} :

$$x_t = \phi x_{t-1} + \epsilon_t \quad (2.13)$$

donde como sabemos ϵ_t es un proceso de ruido blanco con media cero y varianza constante σ^2 , y ϕ es un parámetro. Para verificar que el modelo AR(1) es estacionario debemos probar que es:

(1) Estacionario en media

$$E(x_t) = E(\phi x_{t-1} + \epsilon_t) = \phi E(x_{t-1}) \quad (2.14)$$

Para que el proceso sea estacionario, la media debe ser constante y finita en el tiempo, lo que implica:

$$E(x_t) = \phi E(x_{t-1}) = \frac{0}{1 - \phi} = 0 \quad (2.15)$$

Por lo tanto, para que el proceso sea estacionario el parámetro $\phi \neq 0$.

(2) Estacionario en covarianza

Para verificar que el modelo AR(1) sea estacionario, la varianza debe ser constante y finita en el tiempo:

$$\gamma = E(x_t - E(x_t))^2 = E(\phi x_{t-1} + \epsilon_t - 0)^2 = \phi^2 var(x_{t-1}) + \sigma^2 \quad (2.16)$$

Asumiendo que el proceso es estacionario:

$$E(x_t)^2 = var(x_{t-1}) = var(x_t) = \gamma \quad (2.17)$$

De aquí tenemos que:

$$\gamma = \phi^2 \gamma + \sigma^2 \quad (2.18)$$

Por lo que:

$$\gamma = \frac{\sigma^2}{1 - \phi^2} \quad (2.19)$$

Para que un proceso sea estacionario, varianza constante y finita, es necesario que $|\phi| < 1$.

Si se cumple que $|\phi| < 1$, entonces podemos representar el modelo AR(1) como un proceso lineal dado por:

$$x_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j} \quad (2.20)$$

La ecuación (2.20) se llama **solución estacionaria causal del modelo**. El término causal se refiere al hecho de que x_t no depende del futuro. De hecho, por simple sustitución,

$$\underbrace{\sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}}_{x_t} = \phi \underbrace{\left(\sum_{k=0}^{\infty} \phi^k \epsilon_{t-1-k} \right)}_{x_{t-1}} + \epsilon_t \quad (2.21)$$

2.4.3 Simulación modelo AR(1)

‘Código R: Simulación modelo AR(1)’

```
par(mar=c(1,1,1,1))
par(mfrow=c(2,1))
plot(arima.sim(list(order=c(1,0,0), ar=.9), n=100), ylab="x",
main=(expression(AR(1)~~~phi==+.9)))

plot(arima.sim(list(order=c(1,0,0), ar=-.9), n=100), ylab="x",
main=(expression(AR(1)~~~phi==-.9)))
```

Simulación AR(1)

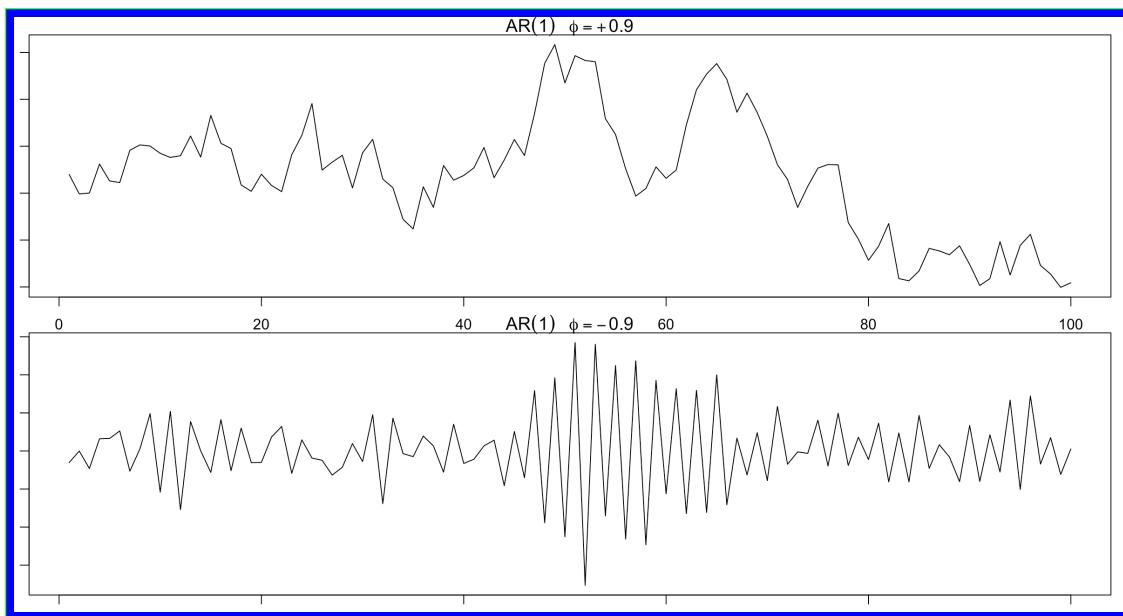


Figure 2.6: Simulaciones procesos AR(1): Arriba: $\phi = +0.9$. Abajo: $\phi = -0.9$

2.4.4 Identificación modelo AR(1)

Para el caso de un proceso del tipo AR, el correlograma, representación gráfica de la función de autocorrelación, tendrá un comportamiento amortiguado hacia cero con todos los valores positivos, en caso de que $\theta > 0$, o bien alternando el signo, comenzando con negativo, si $\theta < 0$.

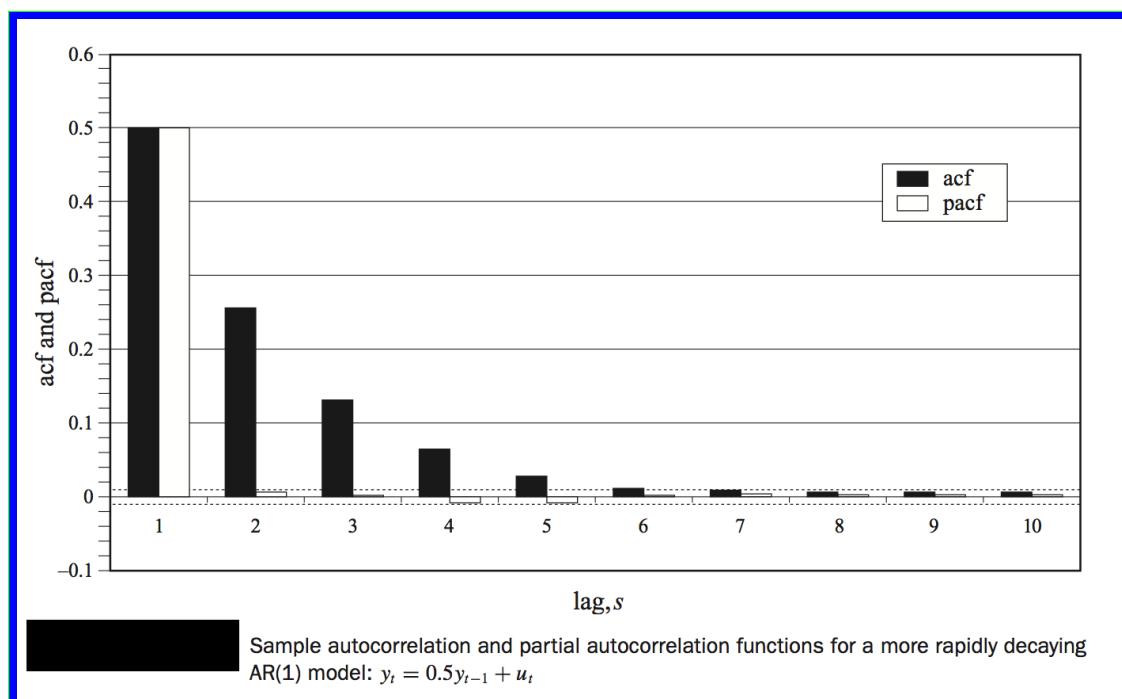
Proceso Autoregresivo: AR(1)

Figure 2.7: Autocorrelogramas para autocorrelación(acf) y autocorrelación parcial(pacf)

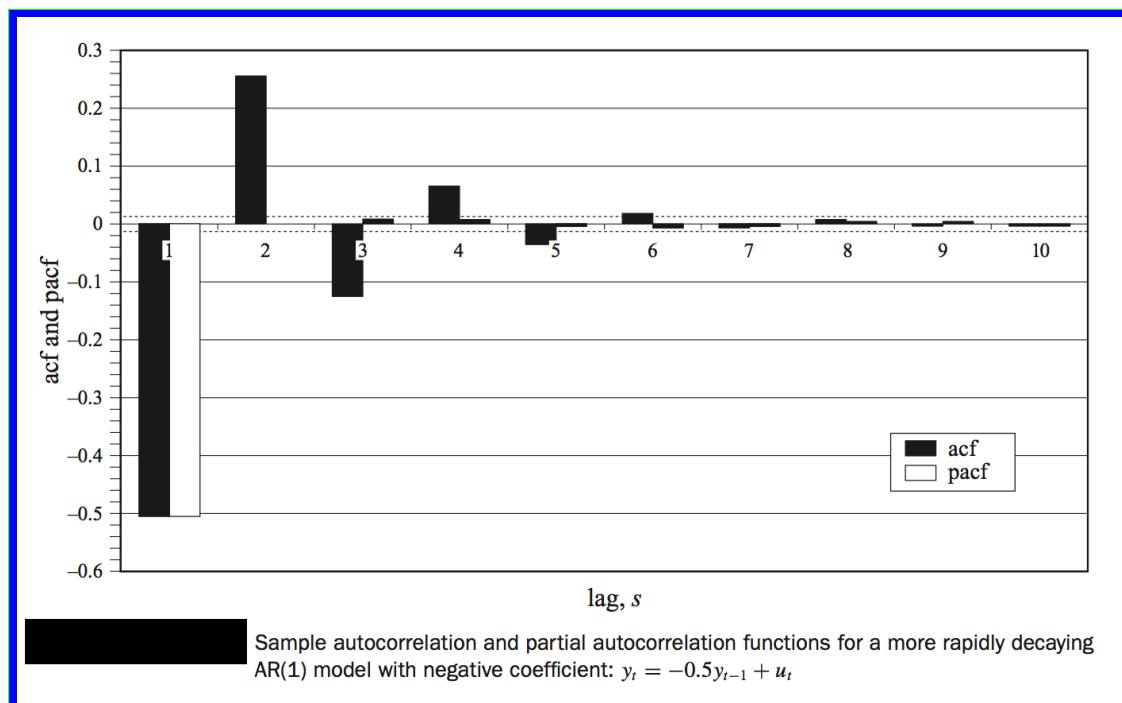
Proceso Autoregresivo: AR(1)

Figure 2.8: Autocorrelogramas para autocorrelación(acf) y autocorrelación parcial(pacf)

2.4.5 Media Móvil - MA (q)

Como una alternativa a la representación autorregresiva en la que se supone que el x_t en el lado izquierdo de la ecuación se combina linealmente, el modelo de promedio móvil de orden q, abreviado como $MA(q)$, asume que el ruido blanco ϵ_t usualmente a la mano derecha de la ecuación, se combinan linealmente para modelar los datos observados.

Definition 11. Definición: Media Móvil - $MA(q)$:

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q} \quad (2.22)$$

donde hay q rezagos de la media móvil ϵ_t y $\theta_1 + \theta_2 + \dots + \theta_q$ son parámetros.
Aunque no es necesario, suponemos que ϵ_t es una serie de ruido blanco.

Definition 12. Definición: Media Móvil - MA(q): También podemos escribir el proceso MA(q) en la forma equivalente:

$$x_t = \theta_t(B)\epsilon_t \quad (2.23)$$

donde θ_t es el operador de promedio móvil definido como:

$$\theta(B) = 1 + \theta_1B + \theta_2B^2 + \dots + \theta_qB^q \quad (2.24)$$

A diferencia del proceso autorregresivo, el proceso de promedio móvil es estacionario para cualquier valor de los parámetros $\theta_1 + \theta_2 + \dots + \theta_q$.

Interpretación del modelo de media móvil - MA (q)

Así como un modelo autorregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta frecuentemente no intuitivo. ¿Qué significa que una variable aleatoria se explique en función de los errores cometidos en períodos anteriores?, ¿De dónde vienen esos errores?, ¿Cuál es la justificación de un modelo de este tipo?. En realidad, un modelo de medias móviles puede obtenerse a partir de un modelo autorregresivo a partir de la realización de sucesivas sustituciones.

Interpretación del modelo de media móvil - MA (q)

Supongamos un modelo AR(1), sin término independiente:

$$x_t = \phi x_{t-1} + \epsilon_t \quad (2.25)$$

si consideramos $t - 1$ en lugar de t el modelo sería en este caso:

$$x_{t-1} = \phi x_{t-2} + \epsilon_{t-1} \quad (2.26)$$

sustituyendo:

$$x_t = \phi^2 x_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \quad (2.27)$$

si ahora sustituimos x_{t-2} por su expresión autorregresiva y así sucesivamente llegamos a un modelo del tipo:

$$x_t = \epsilon_t + \theta\epsilon_{t-1} + \theta^2\epsilon_{t-2} + \dots + \theta^q\epsilon_{t-q} \quad (2.28)$$

que es la expresión, sin término independiente, de un modelo de medias móviles como el planteado anteriormente. En realidad, de forma estricta, el paso de un modelo a otro debería realizarse al contrario, de un MA a un AR, utilizando el teorema general de descomposición de Wold.

‘Código R: ejemplo Simulación modelo MA(1)’

```
par(mfrow = c(2,1))
plot(arima.sim(list(order=c(0,0,1), ma=.5), n=100), ylab="x",
main=expression(MA(1), theta==+.5)))
plot(arima.sim(list(order=c(0,0,1), ma=-.5), n=100), ylab="x",
main=expression(MA(1), theta==-.5)))
```

Ejemplo 1: Simulación modelo MA(1)

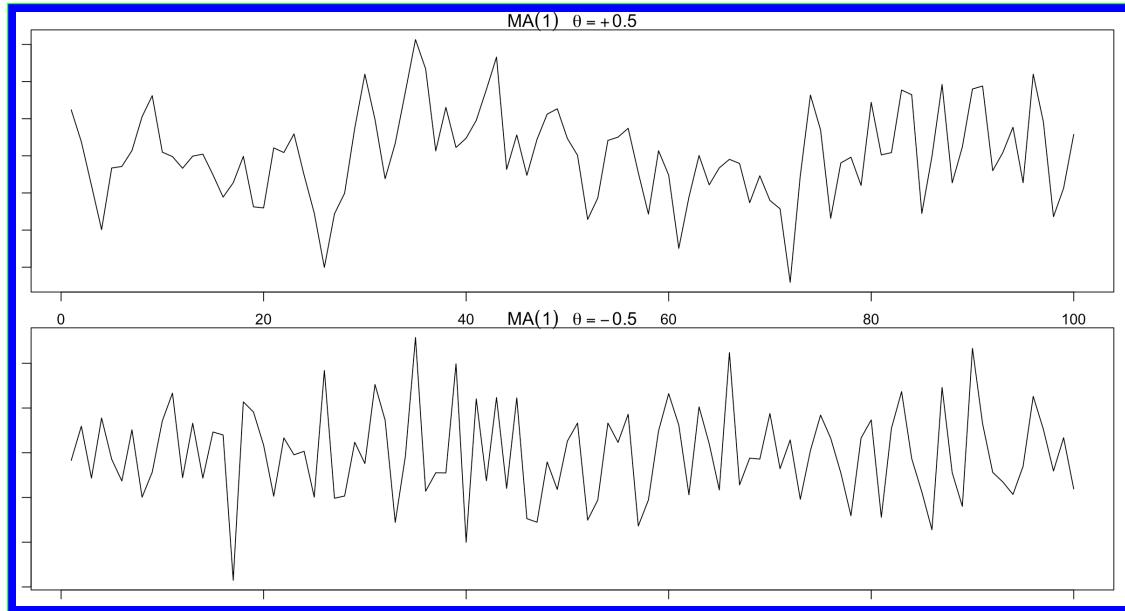


Figure 2.9: Procesos de medias móviles. Arriba: parámetro $\theta = +0.5$. Abajo: parámetro $\theta = -0.5$

Identificación modelo MA

Para la identificación de todos los componentes del modelo MA, tal como vimos para el modelo AR, se utiliza la función de autocorrelación (AFC) y la función de autocorrelación parcial (PAFC), y así se procede a la identificación de los componentes, en base a los gráficos de los distintos modelos teóricos.

Autocorrelograma de proceso de media móvil MA(1), $\theta = +0.5$

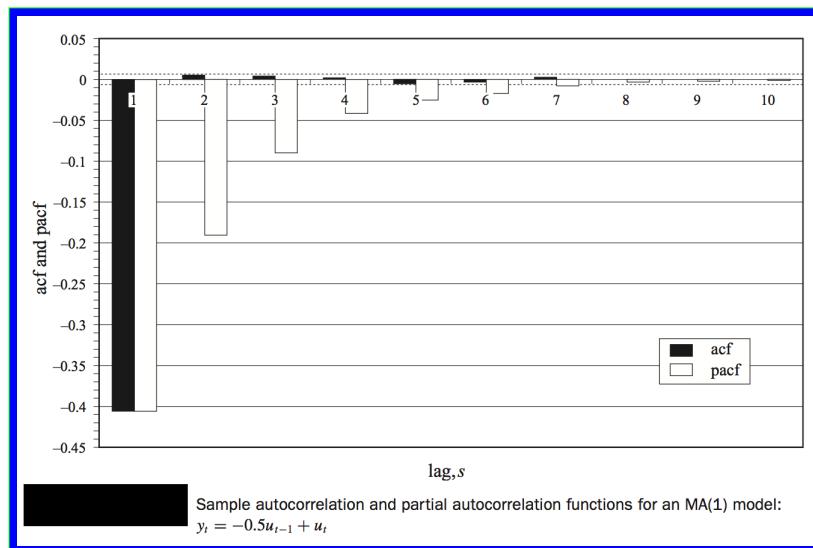


Figure 2.10: Autocorrelogramas para diferentes retardos de proceso de media móvil con $\theta = +0.5$ (lags). ACF(barras negras), PACF(barras blancas).

2.4.6 Identificación modelo ARMA

Autocorrelograma de proceso de media móvil MA(1), $\theta = -0.5$

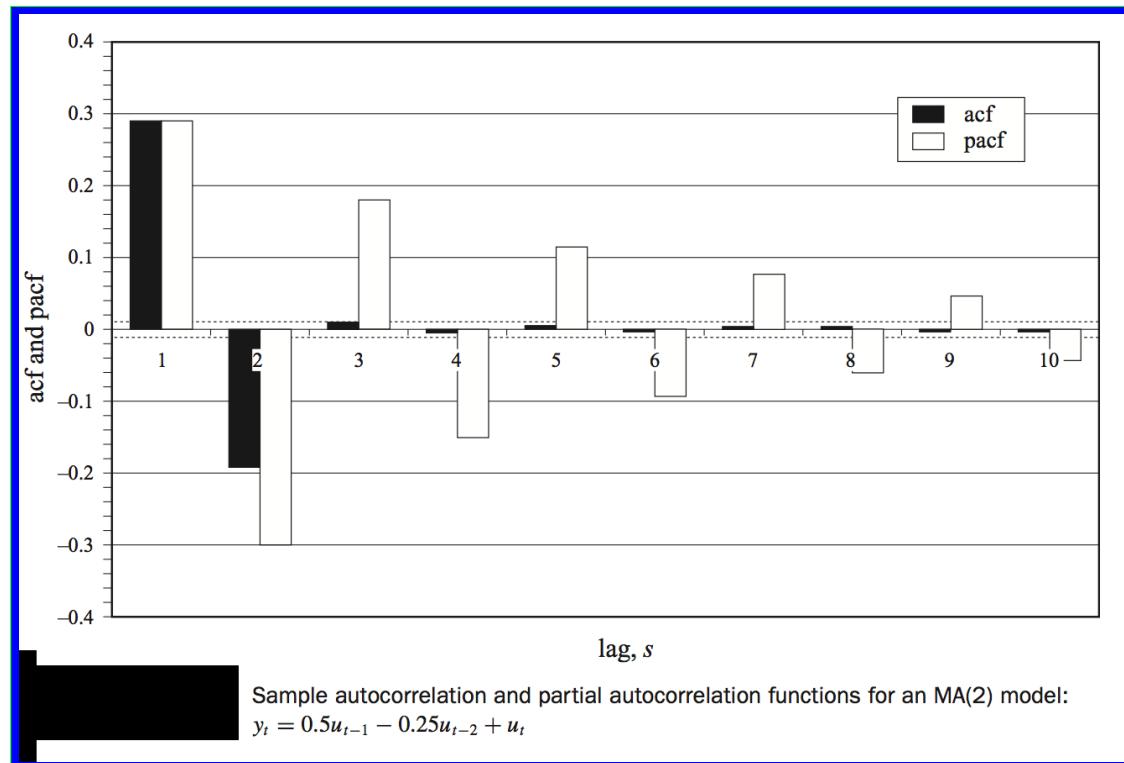


Figure 2.11: Autocorrelogramas para diferentes retardos de proceso de media móvil con $\theta = -0.5$ (lags). ACF(barras negras), PACF(barras blancas).

Definition 13. Definición: *ModeloARMA – ARMA(r, q):*

Una serie de tiempo $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$ es un proceso ARMA(p, q), si es estacionario y

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (2.29)$$

Los parámetros p y q se llaman órdenes autoregresivas y promedios móviles, respectivamente.

Si x_t tiene una media distinta de cero μ , establecemos que $\alpha = \mu(1 - \theta_1 - \dots - \theta_q)$ y podemos re-escribirnos el modelo como:

$$x_t = \alpha + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + w_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}. \quad (2.30)$$

2.4.7 Invertibilidad

Una serie temporal es invertible si los errores se pueden invertir en una representación de observaciones pasadas. Así por ejemplo, como ya vimos, el modelo AR es siempre invertible. En el caso del modelo ARMA, las raíces de las siguientes ecuaciones deben ser analizadas para garantizar invertibilidad.

$$\phi(z) = 1 + \phi_1 z + \phi_2 z^2 + \dots + \phi_p z^p \quad (2.31)$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q \quad (2.32)$$

En particular el modelo ARMA será invertible si y solo si $\theta(z) \neq 0$ para $|z| \leq 1$. En general, los valores propios son la solución del $\det(A - \lambda I) = 0$, vemos que este es casi el polinomio característico de las ecuaciones que definimos arriba. Por lo tanto, vemos que los valores propios de A son el inverso de las raíces del polinomio característico, y esa convergencia de la iteración hacia atrás ocurre cuando las raíces del polinomio característico se encuentran fuera del círculo unitario.

2.4.8 Estacionaridad e Invertibilidad

Wold demostró que todos los procesos estocásticos estacionarios de covarianza podrían descomponerse como la suma de procesos determinísticos y linealmente indeterministas los cuales no estaban correlacionados con todos los rezagos; es decir, si y_t es la covarianza estacionaria, entonces:

$$y_t = x_t + z_t \quad (2.33)$$

donde x_t es un proceso determinista estacionario en covarianza y z_t es linealmente indeterminista, con $\text{Cov}(x_t, z_s) = 0$ para todas los t y s . Este resultado proporciona una base teórica para la propuesta de Box y Jenkins para modelar procesos estacionarios de covarianza escalar (desestacionalizados) como son los procesos ARMA.

2.4.9 Modelos ARMA (p,q)

Como se indicó anteriormente, cuando $q = 0$, el modelo se denomina modelo autoregresivo de orden p , $AR(p)$, y cuando $p = 0$, el modelo se denomina modelo de promedio móvil de orden q , $MA(q)$.

Es útil escribir los modelos ARIMA usando el operador AR y el operador MA descritos anteriormente. En particular, el modelo $ARMA(p, q)$ puede escribirse entonces en forma concisa como:

$$\phi(B)x_t = \theta(B)\epsilon_t. \quad (2.34)$$

2.4.10 Modelos ARIMA (p, i, q)

El modelo ARMA gana su I y se convierte en ARIMA cuando debe ser integrado para lograr estacionaridad. El índice I será entonces el numero de veces que debe ser diferenciado.

2.4.11 Identificación modelo ARMA

Para la identificación de todos los componentes del modelo ARMA se utiliza la función de autocorrelación (AFC) y la función de autocorrelación parcial (PAFC), y así se procede a la identificación de los componentes estacional y no estacional por separado, en base a los gráficos de los distintos modelos teóricos.

Proceso autoregresivo de media móvil de orden (1,1): ARMA(1,1)

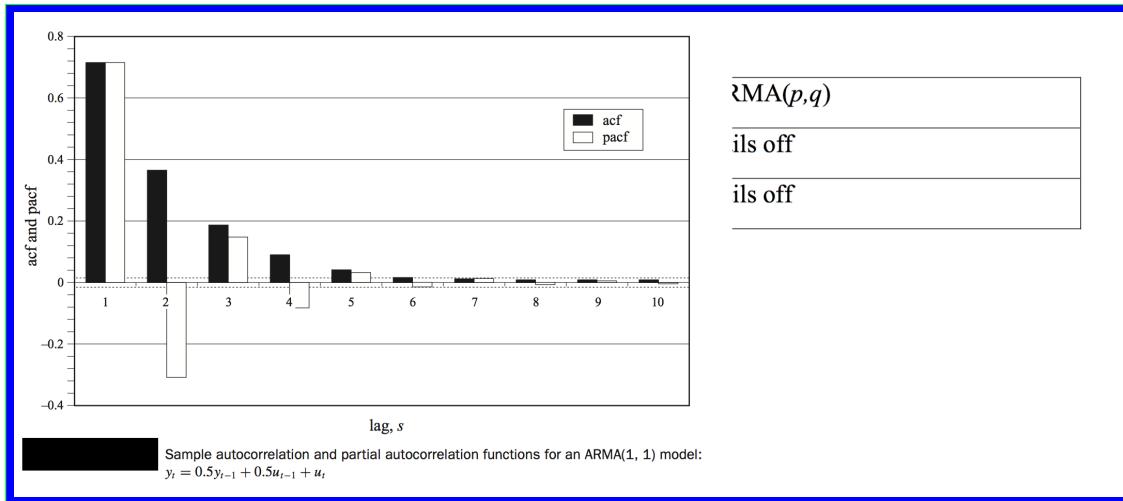


Figure 2.12: Autocorrelogramas para diferentes retardos(lags) para un proceso ARMA(1,1)

En resumen tendremos:

En resumen tendremos:

ACF and PACF properties

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Figure 2.13: Tabla resumen de los modelos AR(p),MA(q) y ARMA(p, q), con sus respectivos resultados esperados para las funciones de autocorrelación(ACF) y autocorrelación parcial(PACF)

2.4.12 Modelos SARIMA (p,q)

Los modelos ARIMA también son capaces de modelar una amplia gama de datos estacionales. Los llamados modelos SARIMA, Seasonal ARIMA models, se obtienen al incluir términos estacionales adicionales en los modelos ARIMA que hemos visto hasta ahora, de la siguiente manera:

$$ARIMA(p, d, q)(P, D, Q)m \quad (2.35)$$

donde m = número de períodos por temporada.

Usamos la notación en mayúsculas para las partes estacionales del modelo y la notación en minúsculas para las partes no estacionales del modelo.

La parte estacional del modelo consiste en términos que son muy similares a los componentes no estacionales del modelo, pero implican retrocesos del período estacional.

Evaluación estadística de un Modelo ARIMA

Se debe evaluar:

- A **Significancia estadística de los parámetros** Los coeficientes obtenidos en la estimación que no sean significativamente distintos de cero, a un nivel de significancia del 5%, no son necesarios, por lo que deben eliminarse.
- B **Estacionariedad e invertibilidad del modelo estimado.** Para valores de los coeficientes estimados próximos a la frontera de la no-estacionariedad, es conveniente llevar a cabo un test de raíces unitarias.
- C **Estabilidad del modelo estimado.** Aunque los parámetros sean significativos, el modelo puede ser rechazado si existe una fuerte correlación entre los parámetros del modelo. Esto ocurre cuando el coeficiente de correlación tiene un valor absoluto superior a 0,7, entonces es conveniente probar con modelos alternativos.

2.4.13 Sobre la Selección de Modelos

Puede ocurrir que varios modelos describan satisfactoriamente la serie temporal, por lo que sea necesario seleccionar el modelo más adecuado. Este proceso de selección puede ser sencillo o un poco más complejo, por lo que es necesario recurrir a criterios de selección de modelos.

Los criterios más comunes en la selección de modelos son el AIC (Akaike Information Criterion) y el BIC (Bayesian Information Criterion) que es una extensión bayesiana del primero.

Criterios de Información

Definition 14. Definición: Akaike Information Criterion

$$AIC = \log \hat{\sigma}_k^2 + \frac{n + 2k}{n}$$

Donde $\hat{\sigma}_k^2 = \frac{SSE_k}{n}$, donde k es el número de parámetros del modelo, n el tamaño de la muestra, y SSE_k equivale a la suma de los residuos al cuadrado bajo el modelo k ($SSE_k = \sum_{t=1}^n (x_t - \bar{x})^2$). El valor de k que produce el mínimo AIC representa el mejor modelo. La idea es que minimizar $\hat{\sigma}_k^2$ representa un objetivo razonable, excepto que disminuye monótonamente a medida que k aumenta. Por lo tanto, debemos penalizar la varianza del error por un término proporcional al número de parámetros.

Definition 15. Definición: Bias Corrected

$$AICc = \log \hat{\sigma}_k^2 + \frac{n + k}{n - k - 2}$$

Definition 16. Definición: Bayesian Information Criterion - BIC

$$AICc = \log \hat{\sigma}_k^2 + \frac{k \log n}{n}$$

BIC también se conoce como el **Schwarz Information Criterion (SIC)**. Varios estudios de simulación han verificado que BIC es adecuado para obtener el orden correcto en muestras grandes, mientras que AICc tiende a ser superior en muestras más pequeñas donde el número relativo de parámetros es grande.

En último término un modelo es mejor que otro si su predicción es mejor. Por otro lado, diremos que **una predicción, es mejor que otra, cuando comete un menor error extra-muestral**. Así, la precisión de los métodos utilizados para pronosticar se pueden medir por ejemplo a través de la función de pérdida: **Error cuadrático medio - Mean Square Error (MSE)**, con el fin de comprender qué modelo proporciona un mejor pronóstico extra-muestral sobre otro. Esto es:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (x_t - \hat{x}_t)^2 \quad (2.36)$$

donde x_t corresponde al valor real de la serie en el tiempo t y \hat{x}_t corresponde al valor pronosticado por el modelo propuesto en el mismo instante.

Otros criterios de selección de modelos que consideran el error extra-muestral son:

- i) el **Error Absoluto Medio (EAM) - mean absolute deviation (MAD)**,

$$\text{MAD} = \frac{1}{T} \sum_{t=1}^T |x_t - \hat{x}_t| \quad (2.37)$$

y

- ii) **Error Absoluto Porcentual Medio (EAPM) - mean absolute percentage error (MAPE)**

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| 1 - \frac{x_t}{\hat{x}_t} \right| \quad (2.38)$$

2.4.14 Ejemplo IPC en Chile

Considerando data mensual del IPC desde enero del 2013 a la fecha en Chile, obtenida de la página del Banco Central, intetnaremos predecir el IPC (serie original).

‘Código R: Ejemplo IPC en Chile’

```
rm(list=ls())
data<-read.csv ("ipc.csv")
ipc <- ts(data[,2], start = c(2013,1), end=c(2018, 6),
           frequency = 12)
plot.ts(ipc,xlab='Years',ylab='Indice_de_Precios_al_Consumidor')
```

Ejemplo IPC en Chile.

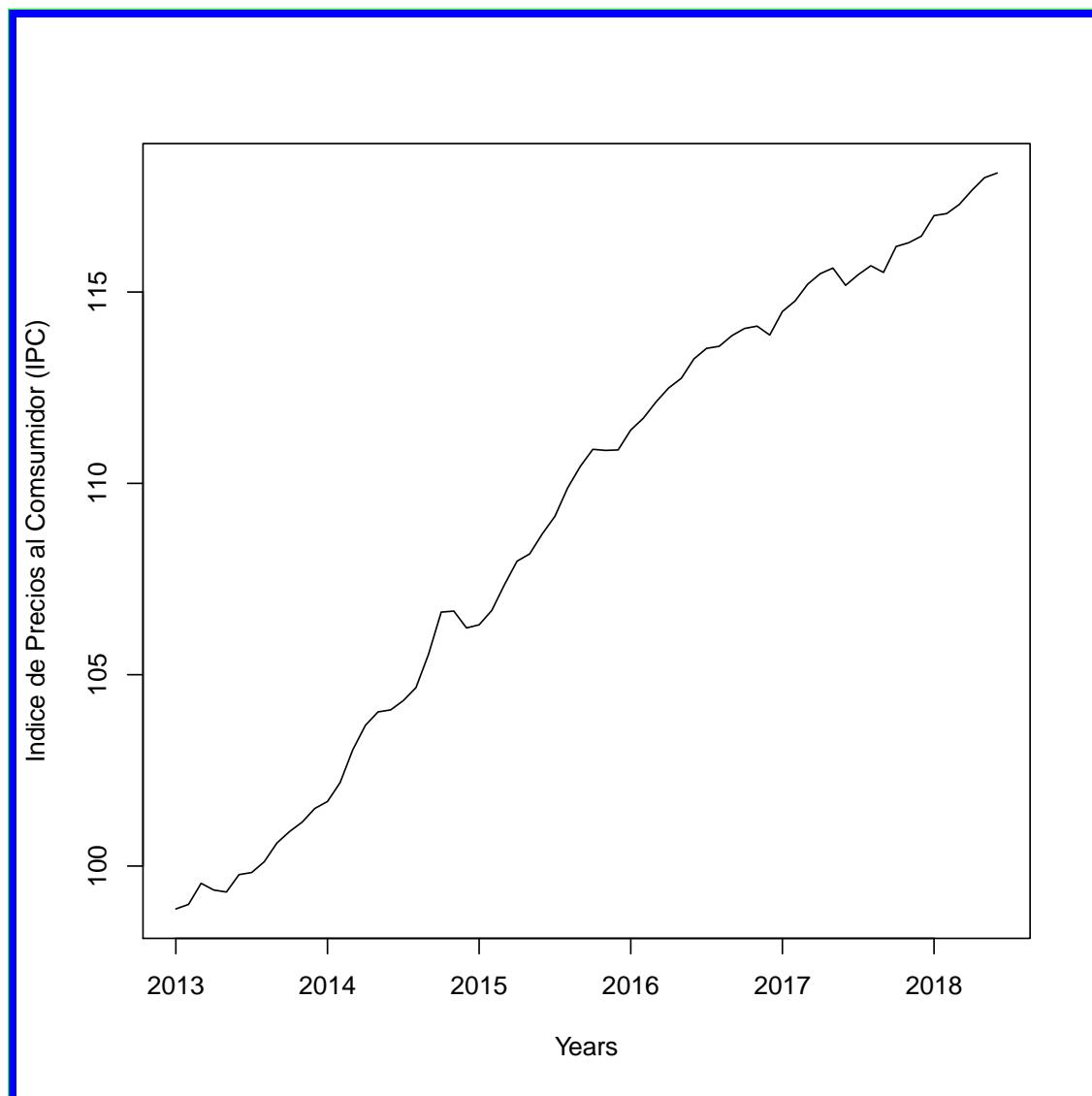


Figure 2.14: Serie de tiempo del índice de precios al consumidor (IPC) en Chile.

Encontrando el orden del modelo. Tendencia, estacionaridad, autocorrelación.

'Código R: Tendencia, estacionaridad, autocorrelación.'

```
#Descomposicion
fit <- stl(ipc, s.window="period")
plot(fit)
library("tseries")
#TestRaizUnitaria
adf.test(ipc)
adf.test(diff(ipc))
#FuncionAutocorrelacion_yAutocorrelacionParcial
acf(diff(ipc),lag=36,lwd=3)
pacf(diff(ipc),lag=36,lwd=3)
```

2.4.15 Descomposición de la serie

Ejemplo IPC: Descomposición de la serie de IPC

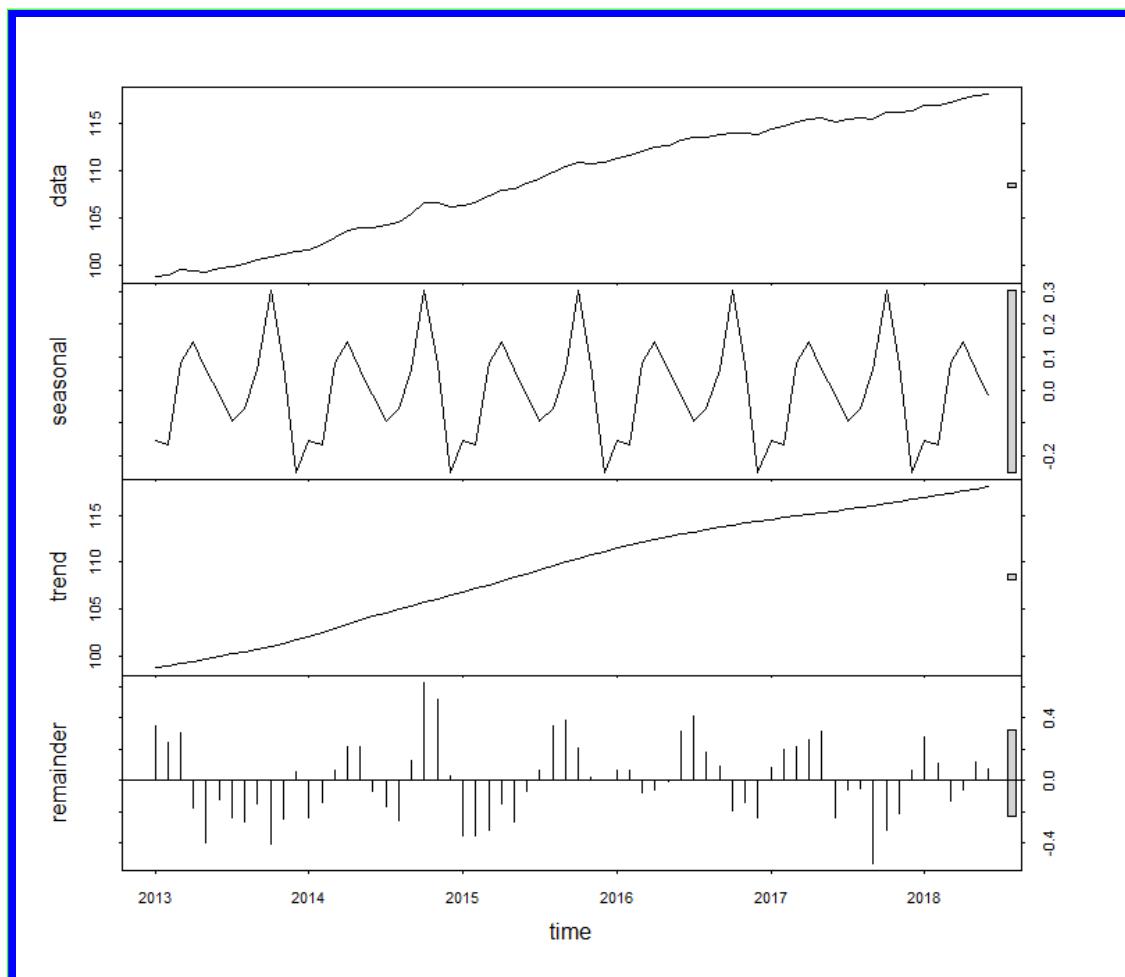


Figure 2.15: Diferentes componentes de la serie: data: serie original de IPC; seasonal: componente estacional; trend: tendencia de la serie; remainder: componentes irregulares.

Test de raíz unitaria

Augmented Dickey-Fuller Test

data: ipc

Dickey-Fuller = -0.11148, Lag order = 4, p-value = 0.99

alternative hypothesis: stationary

Augmented Dickey-Fuller Test

data: diff(ipc)

Dickey-Fuller = -5.8024, Lag order = 3, p-value = 0.01

alternative hypothesis: stationary

Figure 2.16: Test de Augmented Dickey-Fuller sobre data original(ipc) y sobre primera diferencia de la data de ipc(diff(ipc))

Función de autocorrelación (AFC) y autocorrelación parcial (PAFC)

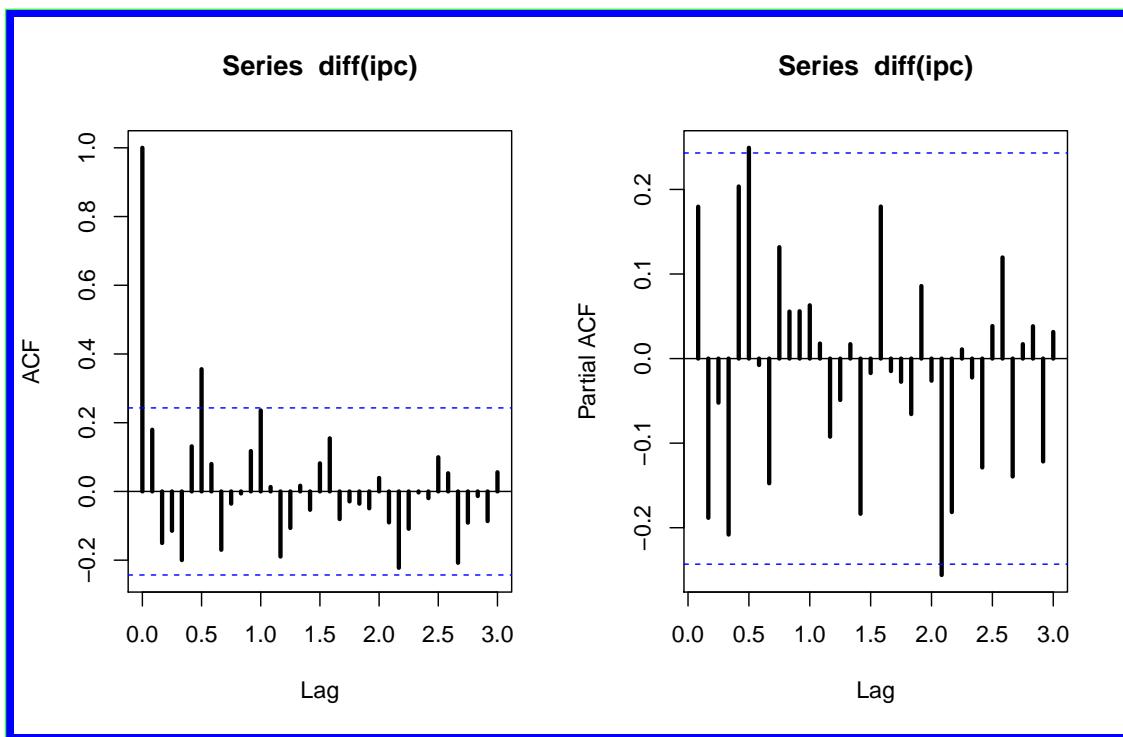


Figure 2.17: Izquierda: autocorrelacion de serie de ipc diferenciada. Derecha: autocorrelación parcial de serie de IPC diferenciada.

'Código R: REVISAR: Diapo 59(125)'

```
train_series=ipc[1:44]
test_series=ipc[45:62]
arimaModel=arima(train_series, order=c(0,1,1))
forecast=predict(arimaModel, length(test_series))
mse <- sum((forecast$pred-test_series)^2)/length(test_series)
mad <- sum(abs(forecast$pred-test_series))/length(test_series)
mape <- sum(abs(1-test_series/forecast$pred))/length(test_series)
fit <- auto.arima(ipc)
summary(fit)
plot(fit)

mape <- 100*sum(abs(1-test_series/f[[ "mean"]]))/length(test_series)
accuracy(fit)
```

Resultado-Parámetros ajuste de modelo ARIMA(0,1,1) y Error de Forecast(MSE)

output ARIMA (0, 1, 1)

Call:

```
arima(x = train.series, order = c(0, 1, 1))
```

Coefficients:

ma1	
0,8205	
s.e.	0,0906

σ^2 estimated as 0,1029 : $\text{loglikelihood} = -12,68$, $\text{aic} = 29,37$

forecast ARIMA (0, 1, 1)

```
mse [1] 69.80031
```

Figure 2.18: Resultados del ajuste del modelo ARIMA(0,1,1) y su error de pronóstico

Pronósticos de modelo ARIMA(0,1,1)

```
$pred
Time Series:
Start = 45
End = 54
Frequency = 1
[1] 113,6141 113,6253 113,6292 113,6307 113,6311 113,6313
[7] 113,6314 113,6314 113,6314 113,6314
$se
Time Series:
Start = 45
End = 54
Frequency = 1
[1] 0,3128668 0,6841962 0,9882296 1,2406559 1,4565974
[6] 1,6465783 1,8174943 1,9738906 2,1188485 2,2545301
```

Figure 2.19: Pronóstico ARIMA(0,1,1)

Pronósticos de modelo AutoARIMA

output auto.arima

Series: ipc

ARIMA(0,1,1)(0,0,1)[12] with drift

Coefficients:

	ma1	sma1	drift
	0,2329	0,2483	0,2909
s.e.	0,1443	0,1396	0,0500

σ^2 estimated as 0,07771 : *loglikelihood* = -8,01

AIC = 24,02 ICc = 24,69 BIC = 32,72

Training set error measures:

	ME	RMSE	MAE	MPE
Training set	0,00467571	0,2701877	0,2012356	0,005434612
	MAPE	MASE	ACF1	
Training set	0,185618	0,05414794	-0,03368001	

Figure 2.20: Pronóstico auto ARIMA

Inverse MA roots - auto.arima

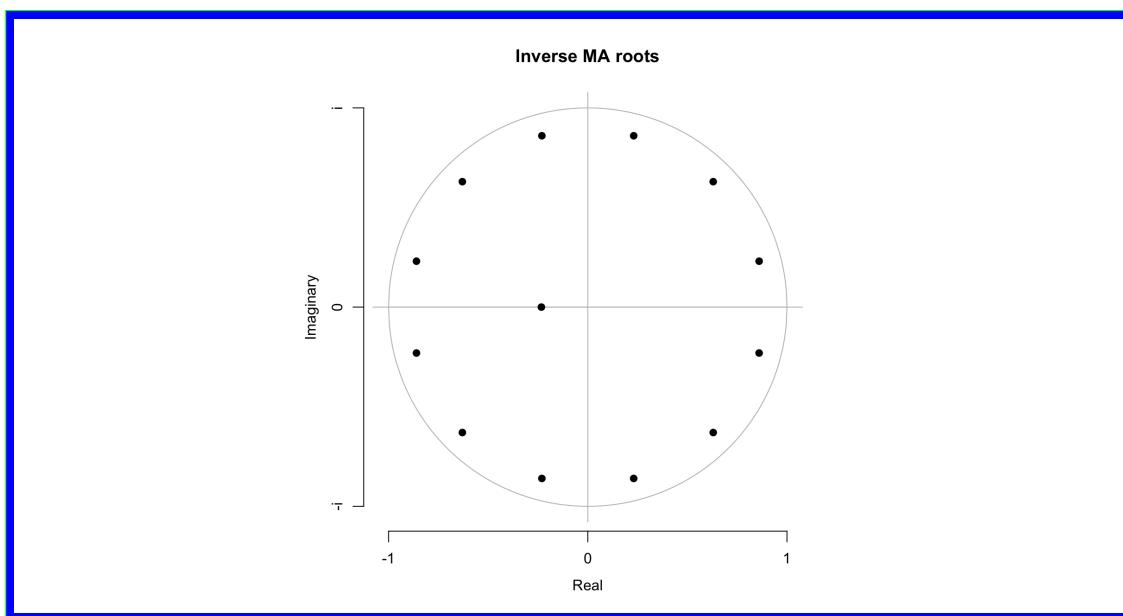


Figure 2.21: Soluciones de ecuación característica - Representación de circulo unitario

Forecasts from ARIMA(0,1,1)(0,0,1)[12] with drift

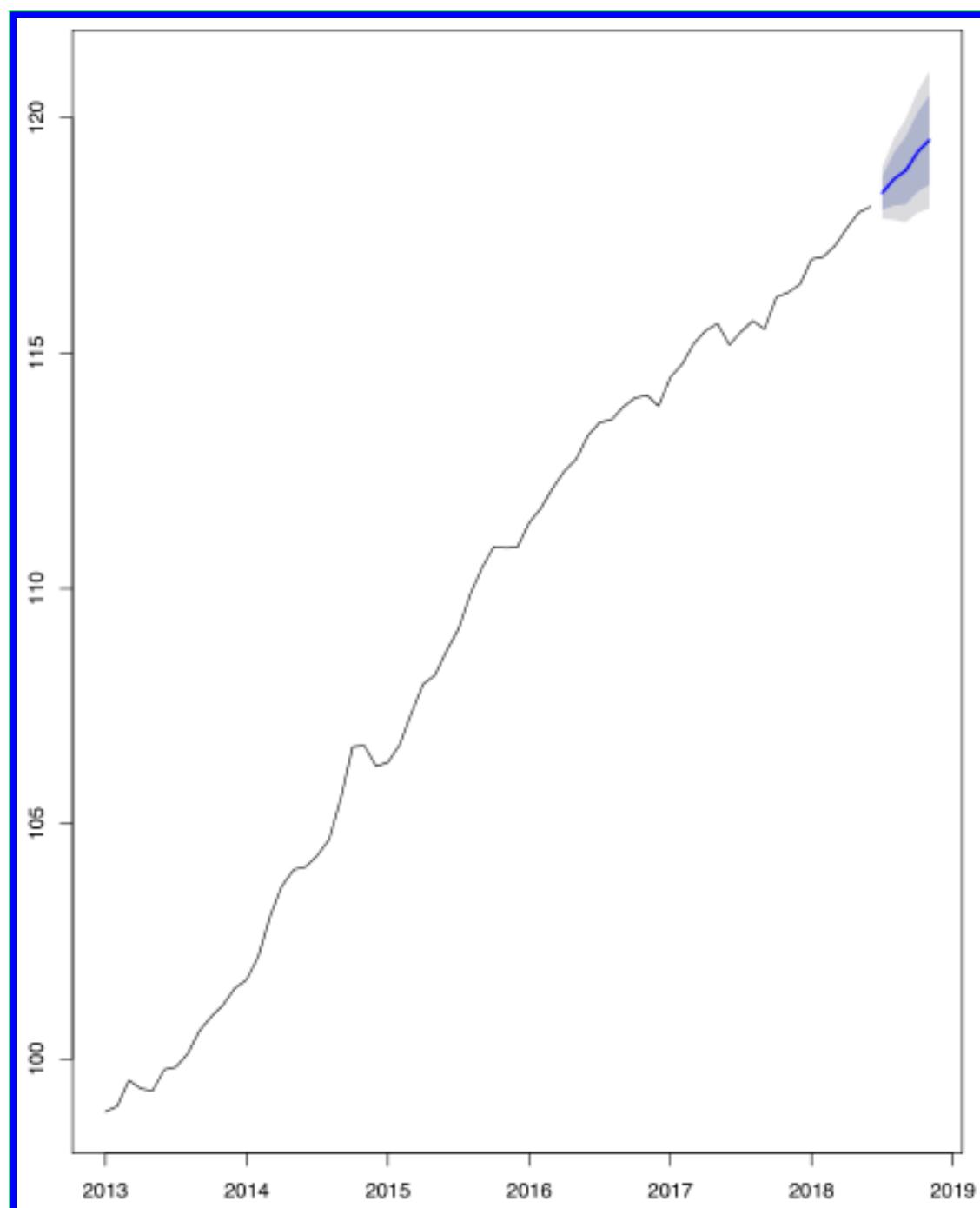


Figure 2.22: Ejemplo IPC en Chile: Forecast de modelo auto.arima

2.5 Metodología de Estimación de un Modelo ARIMA

2.5.1 Etapas de Estimación de un Modelo ARIMA

- 1 **Recolección de datos:** Es recomendable disponer de al menos 50 datos, y en el caso de series mensuales, es conveniente trabajar con entre seis y diez años de datos.
- 2 **Representación gráfica de la serie:** Resulta de gran utilidad disponer de diversos gráficos de la serie y sus transformaciones para decidir sobre la estacionariedad de la misma.
- 3 **Transformación de la serie:** La transformación de la serie es muchas veces necesaria en caso de encontrarnos con no-estacionaridad.
- 4 **Eliminación de la tendencia:** Al comprobarse gráficamente la existencia de una tendencia, esta debe ser eliminada usando como vimos primeras diferencias, e incluso dos diferencias para una tendencia no lineal.
- 5 **Identificación del modelo:** Se debe determinar el tipo de modelo más adecuado, es decir, el orden de los procesos autorregresivos y de medias móviles de las componentes regular y estacional. Se suelen seleccionar varios modelos alternativos, estimarlos, y contrastarlos, antes de modelar definitivamente la serie.
- 6 **Estimación de los coeficientes del modelo:** A partir del modelo elegido se procede a la estimación de sus parámetros.
- 7 **Contraste de validez conjunta del modelo:** Se utilizan los diversos criterios y procedimientos vistos anteriormente para valorar el modelo o modelos seleccionados: test de significancia de parámetros, criterios de información, covarianzas entre estimadores, coeficiente de correlación, R^2 , i.e. suma de cuadrados de errores, etc.
- 8 **Análisis detallado de los errores:** Los errores extra-muestrales del modelo son determinantes para una valoración final del modelo. Las diferencias entre valores reales y estimados por el modelo son determinantes para una evaluación final del modelo.
- 9 **Selección del modelo:** Analizando los resultados de las fases anteriores se decidirá sobre el modelo adoptado. Si ninguno de los modelos estudiados nos proporciona resultados suficientemente satisfactorios se vuelve a la etapa 3, revisando todas las decisiones adoptadas.
- 10 **Predicción:** Se tomará el modelo válido como fórmula inicial de predicción. Será necesario comparar las predicciones con los valores ya conocidos y, posteriormente, analizar los errores extramuestrales.

2.6 Resumen de los pasos de Box-Jenkins

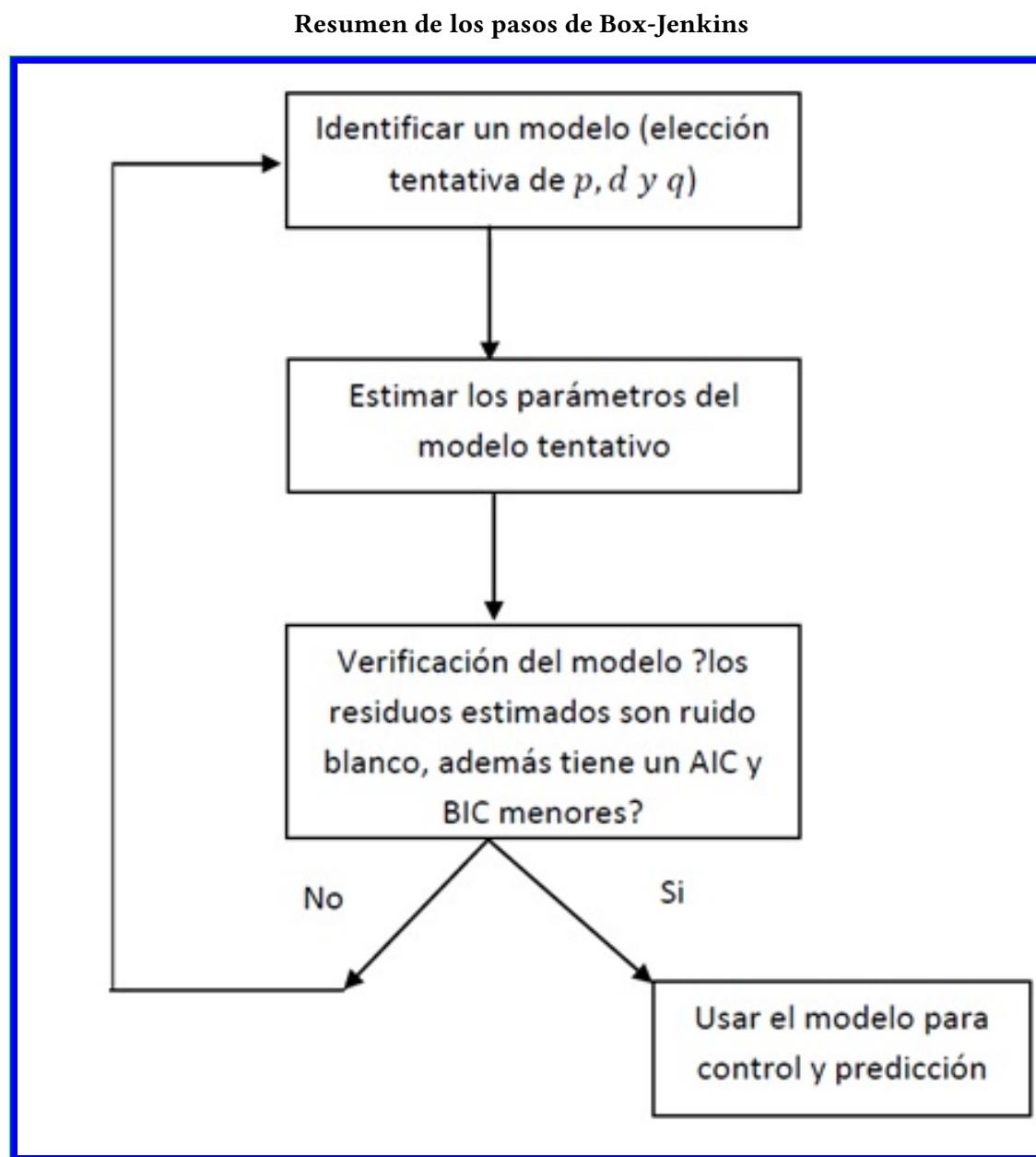


Figure 2.23: Metodología de Box-Jenkins

2.7 Tarea 2

Calibrar y evaluar los siguientes modelos para el precio de un commodity (o activo en último caso) a su elección:

1. Camino aleatorio sin drift.,
2. Camino aleatorio con drift.
3. Promedio de los últimos 5 años.
4. Promedio de los últimos 10 años.
5. ARIMA(1,1,0).
6. ARIMA(0,1,1),,
7. ARIMA(1,1,1),,
8. AR(1),,
9. AR(2),,
10. AR(3).
11. α constante, $\psi = 1$ y δ sigue un camino aleatorio.
12. $\psi=1, \delta=0$ y ?sigue un camino aleatorio.
13. α constante, δ y ψ siguen caminos aleatorios con innovaciones independientes.
14. $\delta = 0$, α y ψ siguen caminos aleatorios con innovaciones independientes.
15. α constante, $\delta = 0$ y ψ sigue un camino aleatorio.
16. α , δ y ψ siguen caminos aleatorios con innovaciones independientes.
17. α constante, $\delta = 0$ y ψ sigue un AR(1).
18. α y δ constantes, ψ sigue un AR(1).

Basarse en paper anexo.

Instructor: Marcelo Villena Chamorro PhD.

3 Tópico III.- Vector Autoregressive Models

3.1 Vector Autoregressive Models

3.1.1 Introducción

Desde la crítica de Sims [Sims \(1980\)](#) a principios de los años ochenta del siglo pasado, el análisis multivariado de datos, en el contexto de los modelos autorregresivos vectoriales (en adelante, VAR por su sigla en inglés **Vector Autoregressive Models**) se ha transformado en un instrumento estándar en econometría.

Debido a que las pruebas estadísticas se utilizan con frecuencia para determinar interdependencias y relaciones dinámicas entre variables, esta metodología pronto se enriqueció al incorporar información antes no considerada.

Los modelos VAR explican las variables endógenas únicamente por su propia historia, además de los regresores determinísticos. En contraste, los modelos VAR estructurales (en adelante, SVAR, por Structural VAR) permiten el modelado explícito de la interdependencia contemporánea entre las variables del lado izquierdo. Por lo tanto, este tipo de modelos intenta eludir las deficiencias de los modelos VAR.

Un VAR es un sistema de dos o más series de tiempo que se modela considerando rezagos de las variables y la interacción dinámica que pudiera existir entre ellas. Consiste fundamentalmente de dos dimensiones, el número de variables (g) y el número de rezagos (k). El caso más simple es un VAR bivariado:

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \dots + \beta_{1k}y_{1t-k} + \alpha_{11}y_{2t-1} + \dots + \alpha_{1k}y_{2t-k} + \mu_1 t$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \dots + \beta_{2k}y_{2t-k} + \alpha_{21}y_{1t-1} + \dots + \alpha_{2k}y_{1t-k} + \mu_2 t$$

donde u_{it} es un término de error con $E(u_{it}) = 0$, $i = 1, 2$; $E(u_{1t}u_{2t}) = 0$. El supuesto de independencia de los errores puede relajarse como veremos más adelante.

El análisis podría extenderse por ejemplo a un modelo VAR (g), donde tenemos g variables y g ecuaciones.

3.1.2 Notación y Conceptos

Una característica importante de los modelos VAR es la sencillez de su notación. Por ejemplo, considere el caso de arriba, donde $k = 1$. Podemos escribir este modelo como:

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + \mu_{1t}$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + \mu_{2t}$$

ó

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix} \quad (3.1)$$

o incluso de forma más compacta como

$$\begin{matrix} y_t = \beta_0 + \beta_1 y_{t-1} + \mu_{1t} \\ g_x 1 g_x 1 g_x g \quad g_x 1 \quad g_x 1 \end{matrix} \quad (3.2)$$

Este modelo se puede extender al caso en que hay k retardos de cada variable en cada ecuación:

$$\begin{matrix} y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \mu_{1t} \\ g_x 1 g_x 1 g_x g \quad g_x 1 \quad g_x g \quad g_x 1 \quad \dots \quad g_x g \quad g_x 1 \quad g_x 1 \end{matrix} \quad (3.3)$$

También podemos extender este al caso en que el modelo incluye términos de primera diferencia y relaciones de cointegración (un VECM que veremos más adelante).

3.1.3 VAR versus Modelos Estructurales

Vector Autoregressive Models Compared with Structural Equations Models

Ventajas del modelo VAR:

- No es necesario especificar qué variables son endógenas o exógenas - todas son endógenas.
- Permite que el valor de una variable dependa de algo más que sus propios rezagos o combinaciones de términos de ruido blanco, por lo que son más generales que nuestro ya conocido modelo ARIMA.
- Siempre que no tengamos términos contemporáneos sobre el lado derecho de las ecuaciones, podremos utilizar OLS por separado en cada ecuación

- Los pronósticos son a menudo mejores que los modelos "estructurales tradicionales".

Problemas con el VAR:

- Los modelos VAR son ateóricos (al igual que los modelos ARIMA).
- ¿Cómo se decide la longitud del rezago apropiado?
- Son tantos parámetros! Si tenemos g ecuaciones para las g variables y tenemos k rezagos de cada una de las variables en cada ecuación, tenemos que estimar $(g + kg^2)$ parámetros. Por ejemplo si $g = 3$, y $k = 3$, tendremos 30 parámetros!!!
- ¿Tenemos que asegurar que todos los componentes del modelo VAR sean estacionarios?
- ¿Cómo interpretar los coeficientes?

3.1.4 La elección de la longitud del rezago óptimo para un VAR

Existen dos enfoques posibles:

- i) restricciones de ecuaciones cruzadas,
- ii) los criterios de información

Restricciones de ecuaciones cruzadas

En el espíritu de un modelado VAR (sin restricciones), cada ecuación debe tener la misma longitud de rezago

Supongamos que un VAR bivariado(8) se estimó utilizando datos trimestrales con 8 rezagos para las dos variables en cada ecuación, y queremos examinar la restricción de que los coeficientes de los rezagos 5 a 8 son conjuntamente cero. Esto se puede hacer utilizando una prueba de razón de verosimilitud.

Denotamos la matriz de varianza-covarianza de los residuales (dada por $\hat{\mu}\hat{\mu}'/T$), como $\hat{\Sigma}$. La prueba de razón de verosimilitud de esta hipótesis conjunta está dada por:

$$LR = T \left[\log |\hat{\Sigma}_r| - \log |\hat{\Sigma}_u| \right]$$

donde $\hat{\Sigma}_r$ es la matriz de varianza-covarianza de los residuos para el modelo restringido (con 4 rezagos), $\hat{\Sigma}_u$ es la matriz de varianza-covarianza de los residuos para el modelo VAR sin restricciones (con 8 rezagos), y T es el tamaño de la muestra.

El estadístico de prueba se distribuye asintóticamente como una χ^2 con grados de libertad igual al número total de restricciones.

En el caso del VAR anterior, estamos restringiendo 4 rezagos de dos variables en cada una de las dos ecuaciones, un total de $4 * 2 * 2 = 16$ restricciones.

En el caso general en el que tenemos un VAR con p ecuaciones, y queremos imponer la restricción

de que los últimos q rezagos tienen coeficientes cero, habría un total de $p2q$ restricciones.
Desventajas: La realización de la prueba LR es complicada y requiere una supuesto de normalidad de las perturbaciones.

Criterios de información para la Selección de Lags del VAR

Se requieren versiones multivariadas de los criterios de información. Estos pueden definirse como:

Akaike information criterion

$$MAIC = \ln| + 2k'/T$$

Bayesian or Schwarz criterion

$$MSBIC = \ln| + k'/T \ln(T)$$

$$MHQIC = \ln| + 2k/T \ln(\ln(T))$$

donde toda la notación se mantiene y k' es el número total de regresores en todas las ecuaciones, que será igual a $g2k + g$ para las g ecuaciones, cada una con k retardos para las g variables, más un término constante en cada ecuación. Los valores de los criterios de información se construyen para $0, 1, \dots$ retardos (hasta algún máximo pre-especificado de \bar{k}).

Ejemplo Modelo VAR Lütkepohl and Krätsig (2004) utilizaron las siguientes series: productividad laboral (prod) definida como la diferencia logarítmica entre el PIB y el empleo, el logaritmo del empleo (e), la tasa de desempleo (U) y los salarios reales (rw), definidos como el logaritmo del índice del salario real. Los datos provienen de la base de datos de la OCDE y abarcan desde el primer trimestre de 1980 hasta el cuarto trimestre de 2004.

'Código R: ejemplo Modelo VAR

```
rm(list=ls())
install.packages(vars)
library(vars)
data("Canada")
summary(Canada)
plot(Canada, nc = 2, xlab = "")
adf2 <- summary(ur.df(Canada[, "prod"], type="drift", lags=1))
```

Ejemplo Modelo VAR: Gráficos de las series de tiempo de cada variable

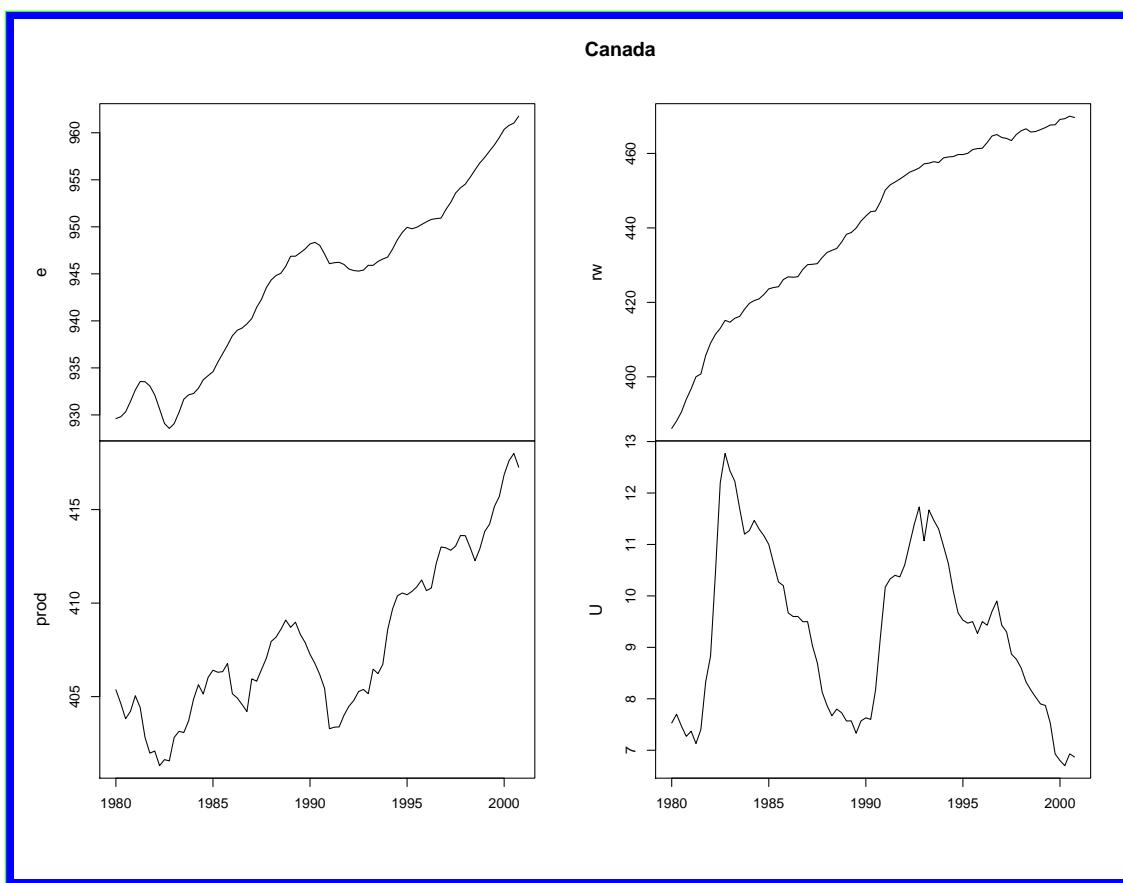


Figure 3.1: e: logaritmo del empleo, prod: productividad definida como la log diferencia entre GDP y empleo,rw: salarios reales , U: tasa de desempleo.

Modelo VAR: resumen de variables

e	prod	rw	U
Min. :929	Min. :401	Min. :386	Min. : 6.70
1st Qu.:935	1st Qu.:405	1st Qu.:424	1st Qu.: 7.78
Median :946	Median :406	Median :444	Median : 9.45
Mean :944	Mean :408	Mean :441	Mean : 9.32
3rd Qu.:950	3rd Qu.:411	3rd Qu.:461	3rd Qu.:10.61
Max. :962	Max. :418	Max. :470	Max. :12.77

Figure 3.2: Estadística resumen para variables del modelo VAR.

Ejemplo VAR: Augmented DF Test Unit Root Test

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.0512 -0.3953  0.0782  0.4111  1.7513 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.1153    0.0803   1.44    0.15    
z.lag.1     -0.6889    0.1335  -5.16  1.8e-06 ***  
z.diff.lag   -0.0427    0.1127  -0.38    0.71    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.697 on 78 degrees of freedom
Multiple R-squared: 0.361,          Adjusted R-squared: 0.345 
F-statistic: 22.1 on 2 and 78 DF,  p-value: 2.53e-08
```

Figure 3.3: Test de Dickey-Füller Aumentado para evaluar la presencia de raiz unitaria en la serie “prod”. El modelo es considerado con “drift” y un rezago(lags=1).

'Código R: ejemplo Modelo VAR Mercado laboral canadiense

```
VARselect(Canada,lag.max = 8,type = "both")
Canada<Canada[, c("prod", "e", "U", "rw")]

p1ct<-VAR(Canada, p = 1, type = "both")
p1ct
summary(p1ct, equation = "e")
plot(p1ct, names = "e")

ser11<-serial.test(p1ct, lags.pt = 16, type = "PT.asymptotic")
ser11$serial
norm1<-normality.test(p1ct)
norm1$jb.mul
prd<-predict(p1ct, n.ahead = 10, ci = 0.95, dumvar = NULL)
print(prd)
plot(prd, "single")
```

Ejemplo Modelo VAR: selección de modelos: instrucción VARselect

\$selection			
AIC(n)	HQ(n)	SC(n)	FPE(n)
3	2	1	3
\$criteria			
	1	2	3
AIC(n)	-6.272579064	-6.636669705	-6.771176872
HQ(n)	-5.978429449	-6.146420347	-6.084827770
SC(n)	-5.536558009	-5.409967947	-5.053794411
FPE(n)	0.001889842	0.001319462	0.001166019
	4	5	6
AIC(n)	-6.634609210	-6.398132246	-6.307704843
HQ(n)	-5.752160366	-5.319583658	-5.033056512
SC(n)	-4.426546046	-3.699388378	-3.118280272
FPE(n)	0.001363175	0.001782055	0.002044202
	7	8	
AIC(n)	-6.070727259	-6.06159685	
HQ(n)	-4.599979185	-4.39474903	
SC(n)	-2.390621985	-1.89081087	
FPE(n)	0.002768551	0.00306012	

Figure 3.4: Cada uno de los modelos contiene mediciones de AIC, HQ, SC y FPE

Ejemplo Modelo VAR: instrucción VAR

```
VAR Estimation Results:  
=====  
  
Estimated coefficients for equation prod:  
=====  
  
Call:  
prod = prod.l1 + e.l1 + U.l1 + rw.l1 + const + trend  
  
prod.l1          e.l1          U.l1  
0.96313671  0.01291155  0.21108918  
      rw.l1          const        trend  
-0.03909399 16.24340747  0.04613085
```

Figure 3.5: Coeficientes de la ecuación para variable “prod”.

Ejemplo Modelo VAR: instrucción VAR

```
Estimated coefficients for equation rw:
=====
Call:
rw = prod.l1 + e.l1 + U.l1 + rw.l1 + const + trend

      prod.l1          e.l1          U.l1
-0.22308744 -0.05104397 -0.36863956
      rw.l1          const         trend
 0.94890946 163.02453066  0.07142229
```

Figure 3.6: Coeficientes de la ecuación para variable “rw”.

Ejemplo Modelo VAR: instrucción VAR

```
Estimated coefficients for equation U:
=====
Call:
U = prod.l1 + e.l1 + U.l1 + rw.l1 + const + trend

      prod.l1          e.l1          U.l1
-0.12319201 -0.24844234  0.39158002
      rw.l1          const         trend
 0.06580819 259.98200967  0.03451663
```

Figure 3.7: Coeficientes de la ecuación para variable “U”.

Ejemplo Modelo VAR: instrucción VAR

```
Estimated coefficients for equation e:  
=====  
Call:  
e = prod.l1 + e.l1 + U.l1 + rw.l1 + const + trend  
  
prod.l1           e.l1           U.l1  
0.19465028      1.23892283    0.62301475  
rw.l1            const          trend  
-0.06776277   -278.76121138   -0.04066045
```

Figure 3.8: Coeficientes de la ecuación para variable “e”.

Ejemplo Modelo VAR - instrucción VAR para e

```
> summary(p1ct, equation = "e")
```

VAR Estimation Results:
=====

Endogenous variables: prod, e, U, rw
Deterministic variables: both
Sample size: 83
Log Likelihood: -207.525
Roots of the characteristic polynomial:
0.9504 0.9504 0.9045 0.7513
Call:
VAR(y = Canada, p = 1, type = "both")

Ejemplo Modelo VAR - instrucción VAR para e

```
Estimation results for equation e:
=====
e = prod.l1 + e.l1 + U.l1 + rw.l1 + const + trend

            Estimate Std. Error t value
prod.l1      0.19465   0.03612   5.389
e.l1         1.23892   0.08632  14.353
U.l1         0.62301   0.16927   3.681
rw.l1        -0.06776   0.02828  -2.396
const       -278.76121  75.18295  -3.708
trend        -0.04066   0.01970  -2.064

            Pr(>|t|)
prod.l1 7.49e-07 ***
e.l1     < 2e-16 ***
U.l1    0.000430 ***
rw.l1    0.018991 *
const    0.000392 ***
trend    0.042378 *

---
Signif. codes:
0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3.9: Evaluación de los parámetros en la ecuación del VAR para variable “e”.

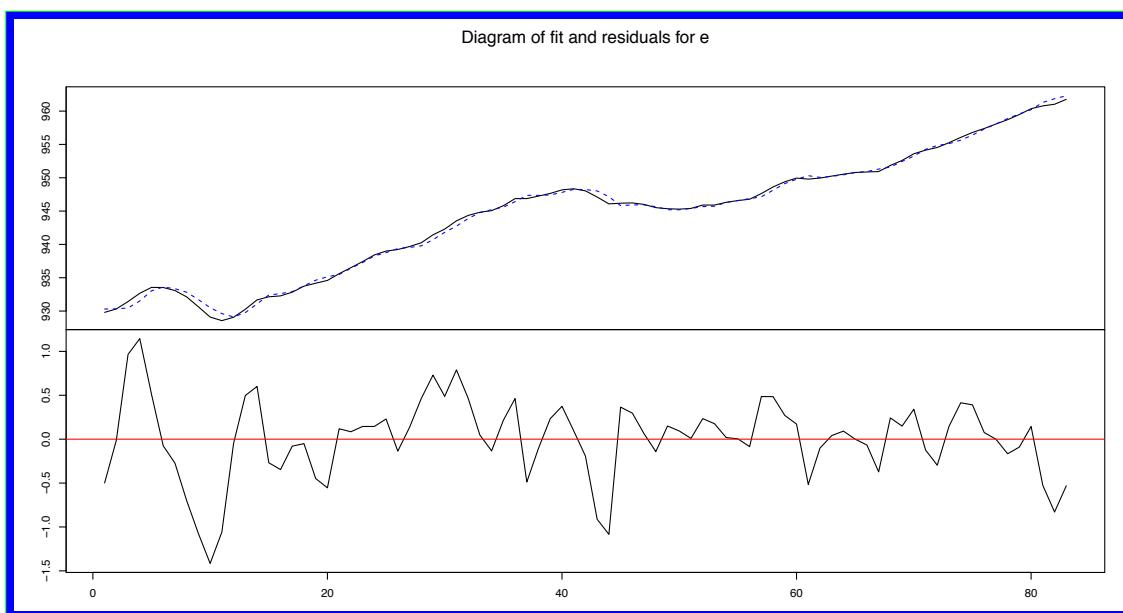
Ejemplo Modelo VAR - instrucción VAR para e

Figure 3.10: Diagrama del ajuste de residuos para ecuación “e”.

Ejemplo Modelo VAR - instrucción VAR para e

```
Residual standard error: 0.4701 on 77 degrees of freedom
Multiple R-Squared: 0.9975,      Adjusted R-squared:
0.9973
F-statistic: 6088 on 5 and 77 DF,  p-value: < 2.2e-16
```

Ejemplo Modelo VAR - instrucción VAR para e

Covariance matrix of residuals:

	prod	e	U	rw
prod	0.469517	0.067667	-0.04128	0.002141
e	0.067667	0.22096	-0.13200	-0.082793
U	-0.041280	-0.13200	0.12161	0.063738
rw	0.002141	-0.08279	0.06374	0.593174

Correlation matrix of residuals:

	prod	e	U	rw
prod	1.000000	0.2101	-0.1728	0.004057
e	0.210085	1.0000	-0.8052	-0.228688
U	-0.172753	-0.8052	1.0000	0.237307
rw	0.004057	-0.2287	0.2373	1.000000

Figure 3.11: Matriz de covarianza y de correlación de modelo VAR

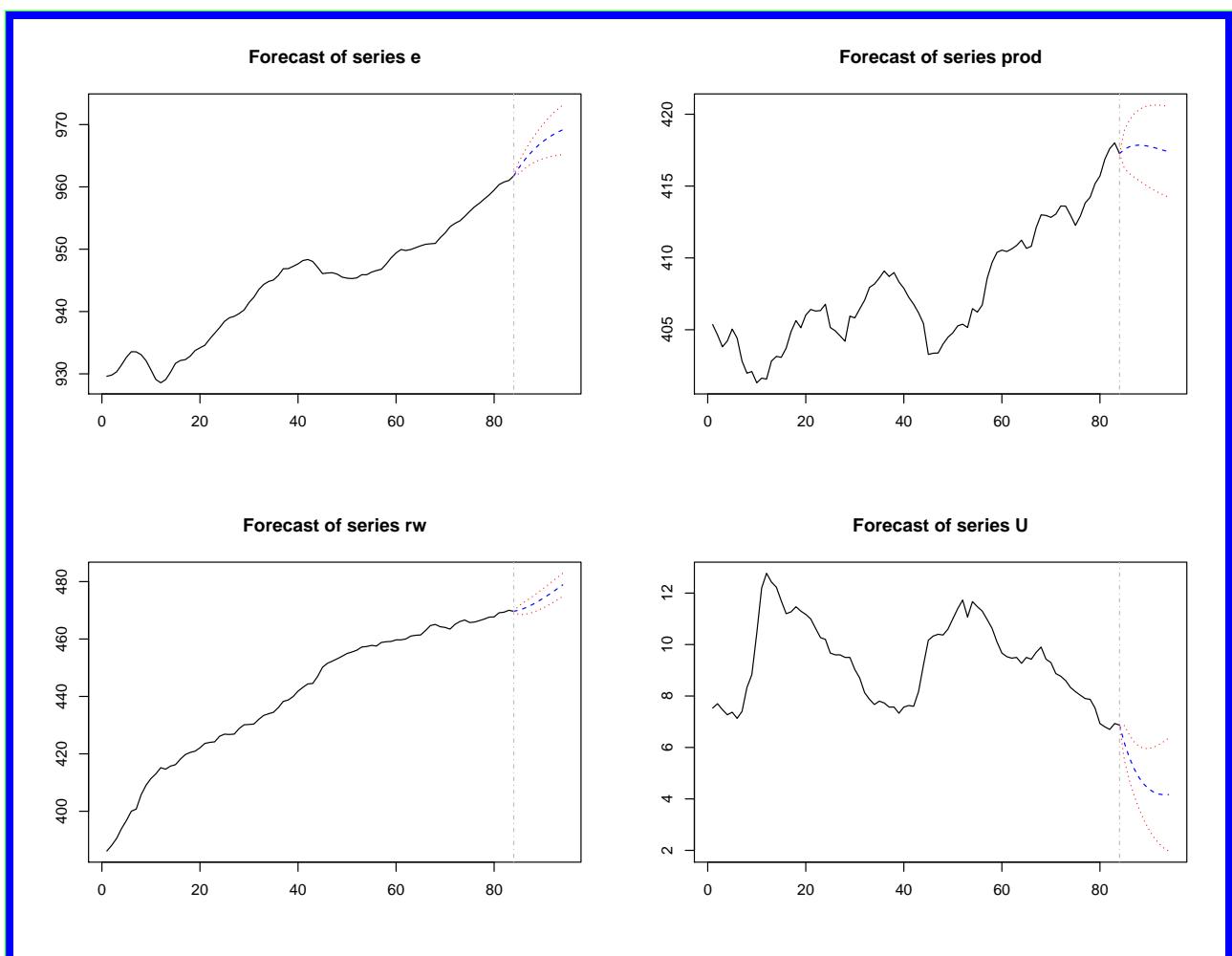
Ejemplo Modelo VAR - instrucción VAR para “e”

Figure 3.12: Pronosticos de modelo VAR.

3.1.5 Forma Primitiva versus Forma Estándar de VARs

¿Incluye el modelo VAR términos contemporáneos? Hasta el momento, hemos asumido que el modelo VAR es de la forma:

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + \mu_1 t$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + \mu_2 t$$

Pero, ¿qué pasa si las ecuaciones tenían un término retroalimentación contemporáneo?

$$y_{1t} = \beta_{10} + \beta_{11}y_{1t-1} + \alpha_{11}y_{2t-1} + \alpha_{12}y_{2t} + \mu_1 t$$

$$y_{2t} = \beta_{20} + \beta_{21}y_{2t-1} + \alpha_{21}y_{1t-1} + \alpha_{22}y_{1t} + \mu_2 t$$

Podemos escribir esto como:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{12} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix}$$

Este VAR está en su forma primitiva....

Forma Primitiva versus Forma Estándar de VARs

Podemos tomar los términos contemporáneos del LHS (lado izquierdo) y escribir:

$$\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{22} & 1 \end{pmatrix} \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \beta_{10} \\ \beta_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \alpha_{21} & \beta_{21} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix}$$

ó

$$By_t = \beta_0 + \beta_1 y_{t-1} + \mu_t$$

Podemos entonces multiplicar ambos lados por B^{-1} y obtener:

$$y_t = B^{-1}\beta_0 + B^{-1}\beta_1 y_{t-1} + B^{-1}\mu_t$$

ó

$$y_t = A_0 + A_1 y_{t-1} + e_t$$

Esto se conoce como un VAR en su Forma Estándar, y se puede estimar por MCO.

3.1.6 Estabilidad de los procesos VAR

Una característica importante de un proceso VAR (p) es su estabilidad. Esto significa que genera series de tiempo estacionarias con medias, varianzas y covarianzas invariantes en el tiempo, dados suficientes valores iniciales. Uno puede verificar esto evaluando el polinomio característico:

$$\det(I_K - A_{1z} - \dots - A_p z^p) = 0. \quad (3.4)$$

Para $|z| \leq 1$.

Si la solución de la ecuación anterior tiene una raíz para $z = 1$, entonces algunas o todas las variables en el proceso VAR (p) son integradas de la orden uno, es decir, I (1).

En la práctica, la estabilidad de un proceso empírico de VAR (p) puede analizarse considerando la forma complementaria y calculando los valores propios de la matriz de coeficientes. Un proceso VAR (p) puede escribirse como un proceso VAR (1):

$$\xi_t = A\xi_{t-1} + v_t$$

$$\xi_t = \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, A = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix}, v_t = \begin{pmatrix} \mu_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Las dimensiones de los vectores ξ_t y v_t son $(Kp \times 1)$ y la dimensión de la matriz A es $(Kp \times Kp)$. Nuevamente, si los módulos de los autovalores de A son menores que uno, entonces el proceso VAR (p) es estable.

Una muestra dada de las variables endógenas y_1, \dots, y_T , y suficientes valores de muestra previa y_{-p+1}, \dots, y_0 , los coeficientes de un proceso VAR (p) pueden estimarse eficientemente mediante mínimos cuadrados aplicados por separado a cada una de las ecuaciones.

Una vez que se ha estimado un modelo VAR (p), la avenida está abierta para su posterior análisis. Un investigador podría / debería estar interesado en las pruebas de diagnóstico, como las pruebas de ausencia de autocorrelación, heterocedasticidad o no normalidad en el término de error. Él podría estar más interesado en la inferencia causal, el pronóstico y / o el diagnóstico del comportamiento dinámico del modelo empírico, es decir, las funciones de respuesta al impulso (en adelante, IRF, por su sigla en inglés **impulse response functions** y pronosticar la descomposición de la varianza del error (en lo sucesivo: FEVD, por su sigla en inglés **forecast error variance decomposition**).

Los dos últimos se basan en la descomposición media móvil de Wold para procesos estables de VAR (p) que se define como::

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \dots$$

con $\Phi_0 = I_K$ y Φ_s se pueden calcular recursivamente según:

$$\Phi_s = \sum_{j=1}^s \Phi_{s-j} A_j$$

para $s = 1, 2, \dots$ donde $A_j = 0$ para $j > p$.

Finalmente, las predicciones para horizontes $h \leq 1$ de un proceso empírico VAR (p) pueden generarse recursivamente según:

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \dots + A_p y_{T+h-p|T}$$

donde $y_{T+j|T} = y_{T+j}$ para $j \leq 0$. La matriz de covarianza de error de pronóstico es dado como:

$$Cov \begin{bmatrix} y_{T+1} - y_{T+1|T} \\ \vdots \\ y_{T+h} - y_{T+h|T} \end{bmatrix} = \begin{bmatrix} I & 0 & \dots & 0 \\ \Phi_1 & I & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \dots & I \end{bmatrix} \left(\sum_u \otimes I_h \right) \begin{bmatrix} I & 0 & \dots & 0 \\ \Phi_1 & I & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ \Phi_{h-1} & \Phi_{h-2} & \dots & I \end{bmatrix}^T$$

y las matrices Φ_i son las matrices de coeficientes empíricos de la representación de media móvil de Wold de un proceso VAR (p) estable como se muestra arriba. El operador \otimes es el producto Kronecker.

3.1.7 Las respuestas de impulso y descomposiciones de la varianza

Ejemplo Impulse Responses - función de respuesta al impulso

Los modelos VAR son a menudo difíciles de interpretar. Soluciones a este problema, como vimos, son la construcción de funciones de respuestas a impulso y las descomposiciones de la varianza.

Las funciones de respuesta al impulso muestran la capacidad de respuesta de las variables dependientes en el VAR a shocks al término de error. Una unidad de shock se aplica a cada variable y sus efectos son almacenados.

Considere, por ejemplo, un VAR(1) bivariante sencillo:

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11} y_{1t-1} + \alpha_{11} y_{2t-1} + \mu_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21} y_{2t-1} + \alpha_{21} y_{1t-1} + \mu_{2t} \end{aligned}$$

Inicialmente, en $t = 1$ suponemos un shock en el término de error μ_{11} de la primera ecuación. Este impacto tiene un efecto directo en y_{11} , de exactamente la misma cantidad. Considerando que y_{21} todavía no se efectúa, y suponiendo que $\mu_{21} = 0$ con $t = 1, \dots, T$. En el segundo período ($t=2$), el impacto original todavía tiene un efecto sobre el valor rezagado de y_{11} . El efecto en y_{12} , es $\beta_{11}\mu_{11}$, y el efecto en y_{22} es $\beta_{21}\mu_{11}$. En el tercer período, el efecto sobre y_{13} , no es solo $\beta_{11}(\beta_{11}\mu_{11})$, sino también $\beta_{12}(\beta_{21}\mu_{11})$. En consecuencia, el efecto sobre y_{23} es $\beta_{21}(\beta_{21}\mu_{11}) + \beta_{22}(\beta_{21}\mu_{11})$.

Por lo tanto, es posible resumir el efecto de un shock no recurrente en una variable, con todas las variables a lo largo del tiempo. Uno podría resumir el resultado anterior en:

$$y_t = \sum_{k=0}^{\infty} C_k \mu_{t-k} \quad (3.5)$$

donde como vimos, $C_0 = I$ (Vector-Moving-Average Process) y donde C_k son el peso de las existencias anteriores.

Ejemplo Orthogonal Impulse Responses

En el modelo anterior de respuesta al impulso, asumimos que los términos de error de la ecuación diferente no están correlacionados. Sin embargo, esta suposición es bastante poco plausible. Un choque hipotético en una sola ecuación no responde a un proceso de ajuste realista. Para controlar la correlación entre los términos de error, tenemos que usar las secuencias de respuesta de impulso ortogonal. La idea es modificar la construcción media móvil original de forma que los residuos no estén correlacionados, es decir, los residuos deben ser ortogonales entre si. Por lo tanto, podemos escribir:

$$y_t = \sum_{k=1}^{\infty} \hat{C}_k v_{t-k} \quad (3.6)$$

donde $\hat{C}_k = C_k G$, donde G es una matriz de transformación con la propiedad GGI (Descomposición de Cholesky)

Los términos del error del sistema modificado son $v_{t-k} = G^{-1} \mu_{t-k}$. La matriz de varianza-covarianza de μ_{t-k} es diagonal, de acuerdo con las propiedades de G . Sin embargo, la G no está claramente definida por la descomposición de Cholesky ($\hat{\Omega} = G^{-1} G'^{-1}$), donde $\hat{\Omega}$ es la matriz original de varianza-covarianza). Además, tenemos que especificar el orden de las variables. El orden elegido supone la relación causal entre las variables. Los resultados de la respuesta al impulso pueden depender mucho del orden de las variables, especialmente cuando están altamente correlacionadas.

Descomposiciones de varianza

Las "Descomposiciones de Varianza" permiten también examinar la dinámica de los modelos VARs. Estas entregan la proporción de los movimientos en las variables dependientes que se deben a sus "propios" shocks, versus a los shock de otras variables.

La descomposición de la varianza da información acerca de la importancia relativa de cada shock en las variables en el VAR.

3.1.8 La función de respuesta al impulso y descomposiciones de la varianza: El orden de las variables

Sin embargo, para el cálculo de las respuestas al impulso y descomposiciones de la varianza, el ordenamiento de las variables es importante.

La razón principal de esto es que, asumimos que los términos de error del modelo VAR eran estadísticamente independientes uno de otro.

Sin embargo, en general, esto no es cierto. Los términos de error estarán siempre correlacionados en cierto grado. El ajuste dinámico de la dependencia recíproca no es inmediatamente considerable. La prueba de respuesta al impulso muestra los efectos de un shock exógeno en todo el proceso a lo largo del tiempo. Por lo tanto, uno puede detectar las relaciones dinámicas a lo largo del tiempo.

Por lo tanto, la noción de examinar el efecto de las innovaciones por separado tiene poco significado, ya que tienen un componente común.

Lo que se hace es "ortogonalizar" las innovaciones.

Inicialmente, observe el ajuste de las variables endógenas a lo largo del tiempo, después de un shock hipotético en t. Este ajuste se compara con el proceso de series temporales sin una descarga, es decir, el proceso real. Las secuencias de respuesta al impulso grafican la diferencia entre estas dos rutas de tiempo.

En el VAR bivariante, este problema se puede abordar mediante la asignación de todo el efecto de la componente común a la primera de las dos variables en el VAR.

En el caso general donde hay más variables, la situación es más compleja, pero la interpretación es la misma.

Ejemplo Modelo VAR:

'Código R: ejemplo Modelo VAR

```
var.irf <- irf(p1ct, response = "U", n.ahead = 10, boot = TRUE)
plot(var.irf)
var.irf1 <- irf(p1ct, impulse = "e", response = "U", n.ahead = 10, boot= TRUE)
plot(var.irf1)
fevd.U <- fevd(p1ct, n.ahead = 48)$U
summary(fevd.U)
```

Ejemplo Modelo VAR - irf

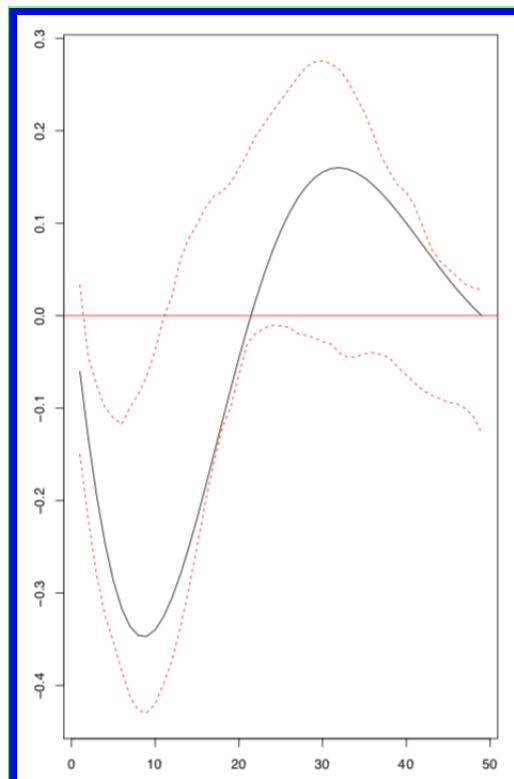


Figure 3.13:

Ejemplo Modelo VAR - fevd

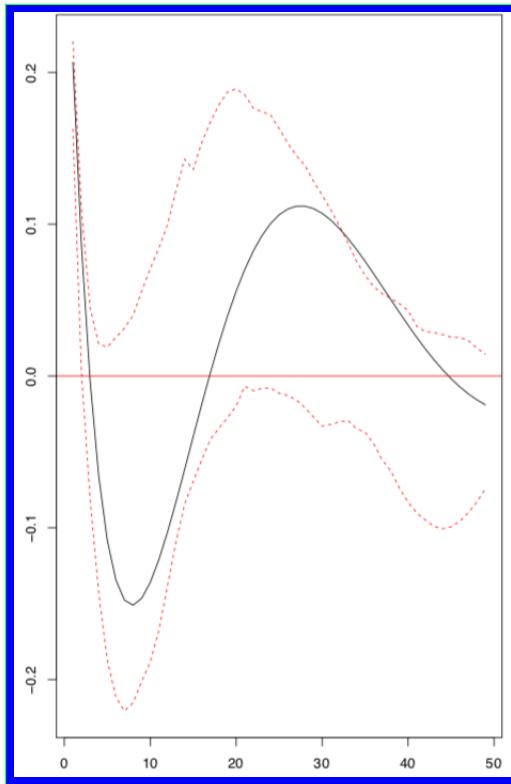


Figure 3.14:

3.1.9 Significancia de un Block y Test de Causalidad

Es probable que, cuando un modelo VAR incluye muchos rezagos de las variables, sea difícil ver qué conjuntos de variables tienen efectos significativos sobre cada variable dependiente y cuáles no. Por ejemplo, considere la siguiente VAR(3) bivariado:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{10} \\ \alpha_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix} + \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} \begin{pmatrix} y_{1t-2} \\ y_{2t-2} \end{pmatrix} + \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{21} & \delta_{22} \end{pmatrix} \begin{pmatrix} y_{1t-3} \\ y_{2t-3} \end{pmatrix} + \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \end{pmatrix}$$

Este VAR puede ser escrito para expresar las ecuaciones individuales como:

$$y_{1t} = \alpha_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + \gamma_{11}y_{1t-2} + \gamma_{12}y_{2t-2} + \delta_{11}y_{1t-3} + \delta_{12}y_{2t-3} + \mu_{1t}$$

$$y_{2t} = \alpha_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + \gamma_{21}y_{1t-2} + \gamma_{22}y_{2t-2} + \delta_{21}y_{1t-3} + \delta_{22}y_{2t-3} + \mu_{2t}$$

Podríamos estar interesados en probar las siguientes hipótesis, y sus restricciones implícitas en las matrices de parámetros:

Hypothesis	Implied Restriction
1. Lags of y_{1t} do not explain current y_{2t}	$\beta_{21} = 0$ and $\gamma_{21} = 0$ and $\delta_{21} = 0$
2. Lags of y_{1t} do not explain current y_{1t}	$\beta_{11} = 0$ and $\gamma_{11} = 0$ and $\delta_{11} = 0$
3. Lags of y_{2t} do not explain current y_{1t}	$\beta_{12} = 0$ and $\gamma_{12} = 0$ and $\delta_{12} = 0$
4. Lags of y_{2t} do not explain current y_{2t}	$\beta_{22} = 0$ and $\gamma_{22} = 0$ and $\delta_{22} = 0$

Cada una de estas cuatro hipótesis conjuntas pueden ser probadas usando un test F, ya que cada conjunto de restricciones sólo contiene los parámetros extraídos de una ecuación.

Estas pruebas también pueden ser llamadas Pruebas de Causalidad de Granger.

Las **pruebas de causalidad de Granger** tratan de responder a preguntas como ¿Los cambios en y_1 causan cambios en y_2 ?

Si y_1 causa y_2 , rezagos de y_1 deberían ser significativos en la ecuación y_2 . Si este es el caso, se dice que y_1 "Granger-causa" y_2 . (y_1 "Granger-causes" y_2)

Si y_2 causa y_1 , retardos de y_2 deben ser significativos en la ecuación de y_1 .

Si ambos conjuntos de rezagos son importantes, existe una "relación de causalidad bidireccional".

'Código R: Ejemplo Modelo VAR - Test de Causalidad de Granger

```
var.cg <- VAR(Canada, p = 2, type = "const")
causality(var.cg, cause = "e")
grangertest(prod~e, order=4, data=Canada)
for (i in 1:4){
  cat("LAG_=", i)
  print(causality(VAR(Canada, p = i, type = "const"), cause ="e")$Granger)
}
```

Test de granger sobre la variable “prod”

```

> causality(var.2c, cause = "e")
$Granger

    Granger causality H0: e do not
    Granger-cause prod U rw

    data: VAR object var.2c
    F-Test = 6.2768, df1 = 6, df2 = 292,
    p-value = 3.206e-06

$Instant

    H0: No instantaneous causality between:
    e and prod U rw

    data: VAR object var.2c
    Chi-squared = 26.068, df = 3, p-value =
    9.228e-06
  
```

Figure 3.15: Resultados de test de causalidad de Granger: Ho: “e” no es Granger-Causa de “prod”, es decir los rezagos de “e” no son significativos en la ecuación de “prod”

3.1.10 Resumen

- Hasta ahora habíamos estudiado modelos univariados.
- Ventajas del modelo VAR: a) es relativamente fácil de especificar y de estimar, b) las variables pueden ser no-estacionarias, c) los errores pueden estar correlacionados contemporáneamente.
- Desventajas del modelo VAR: muchos parámetros.
- Así, en este tópico ampliamos nuestro entendimiento de la relación entre series de tiempo, permitiendo la posibilidad que exista retroalimentación a partir de shocks idiosincráticos.
- Esta dinámica puede ser capturada por medio de un modelo de vector autorregresivo (VAR), que es fundamentalmente una generalización del análisis de procesos autorregresivos, en el cual, en lugar de considerar una sola variable, se considera un vector de variables.

3.2 Tarea 3

Calibrar un modelo del tipo VAR, y desarrollar las siguientes actividades:

1. Encontrar la raíz unitaria de cada variable.
2. Seleccionar los lags óptimos del modelo.
3. Calibrar el modelo e interpretar sus resultados.
4. Analizar los residuos y la estabilidad de su modelo.
5. Realizar un análisis de impulso y descomposición de varianza.
6. Realizar pronóstico extra-muestral usando ventanas móviles.
7. Comparar los errores de su modelo, con el mejor modelo univariado que pueda desarrollar (recuerde tarea 2).

Basarse en paper anexo.

Instructor: Marcelo Villena Chamorro PhD.

Bibliography

- Box, G. E. P. and G. Jenkins (1970). *Time Series Analysis, Forecasting, and Control*. Francisco Holden-Day.
- Dickey, D. A. and W. A. Fuller (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057–1072.
- Granger, C. W. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of econometrics* 2(2), 111–120.
- Lütkepohl, H. and M. Krätsig (2004). *Applied time series econometrics*. Cambridge university press.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, 1–48.